Mapping the Design Space of Teachable Social Media Feed Experiences

K. J. Kevin Feng
University of Washington
Seattle, USA
kjfeng@uw.edu

Xander Koo
Georgia Institute of Technology
Atlanta, USA
xander@gatech.edu

Lawrence Tan
University of Washington
Seattle, USA
lawtan@cs.uw.edu

Amy Bruckman*
Georgia Institute of Technology
Atlanta, USA
asb@cc.gatech.edu

David W. McDonald*
University of Washington
Seattle, USA
dwmc@uw.edu

Amy X. Zhang*
University of Washington
Seattle, USA
axz@cs.uw.edu

ABSTRACT

Social media feeds are deeply personal spaces that reflect individual values and preferences. However, top-down, platform-wide content algorithms can reduce users' sense of agency and fail to account for nuanced experiences and values. Drawing on the paradigm of interactive machine teaching (IMT), an interaction framework for non-expert algorithmic adaptation, we map out a design space for *teachable social media feed experiences* to empower agential, personalized feed curation. To do so, we conducted a think-aloud study ($N = 24$) featuring four social media platforms—Instagram, Mastodon, TikTok, and Twitter—to understand key signals users leveraged to determine the value of a post in their feed. We synthesized users' signals into taxonomies that, when combined with user interviews, inform five design principles that extend IMT into the social media setting. We finally embodied our principles into three feed designs that we present as sensitizing concepts for teachable feed experiences moving forward.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; *Web-based interaction.*

KEYWORDS

social media, interactive machine teaching, feed curation

ACM Reference Format:
K. J. Kevin Feng, Xander Koo, Lawrence Tan, Amy Bruckman, David W. McDonald, and Amy X. Zhang. 2024. Mapping the Design Space of Teachable Social Media Feed Experiences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3613904.3642120

*Equal contribution as senior authors.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642120

1 INTRODUCTION

Social media feeds shape our everyday online interactions with others. Their interface designs and affordances define boundaries on *how* we can engage with people and content, while their algorithms dictate both *what* and *who* we engage with in the first place. Taken together, they shape our collective behaviors and social norms, cementing their role as social architects in the digital age.

The rise of algorithmic content curation and distribution via feeds happened alongside a key shift in the ownership structure of social networks—in the past 20 years, social media platforms transitioned away from being hosted by distributed, independent servers to operating under a small number of private corporations reaching millions of people worldwide [22]. The funneling of capacities for algorithmic curation into the hands of a select few results in what Reviglio et al. call a lack of "algorithmic sovereignty" [91]. It is by this process that the complexity and richness of our social realities have been distilled into a small number of homogenized parameters. In the face of seemingly ubiquitous promises of in-feed personalization, this has brought about a "personalization paradox" [99].

Increasingly centralized curation can have significant negative consequences for users' agency [6, 45, 54, 67, 85, 93]. Platforms today often employ a top-down, one-size-fits-all approach when it comes to platform design and governance. In doing so, they marginalize those who fail to conform to the platform-wide majority. For example, user groups with culturally significant patterns of language use may have their content mislabeled and "downranked" by platforms [45, 93], while neurodiverse users may find many posts too overwhelming to consume [85]. Even users who fit within the majority may get frustrated from being shown irrelevant content with no efficient way to set controls that would eliminate such content from their feeds.

Prior work has documented users' attempts to reclaim their agency by deriving algorithmic folk theories to probe black-box feed curation algorithms [23, 30, 31, 51, 63, 98] and "teaching" these algorithms to better align with their preferences through strategic in-feed interactions [30, 55]. The efficacy of these ad-hoc techniques, however, is often unclear, and using them can even leave users with undesirable feelings of coercion and manipulation [14]. Why might this be? The problem is unlikely to be rooted in the *quantity of feedback* that the user provides to the algorithm—after all, modern recommender systems leverage a wide variety of both implicit (e.g., content dwell time, mouse movements) and explicit (e.g., likes,

blocks) feedback elicitation techniques [57, 73, 78, 104], sometimes learning user preferences to a startling degree of accuracy [99]. Instead, we posit that this is due to users' lack of opportunity to *agentially articulate pertinent feedback* to the algorithm, and have the algorithm respond accordingly based on their feedback.

Given this, we draw upon literature in interactive machine teaching (IMT) [79, 90, 111] to chart out avenues for enabling *teachable feed experiences*[1] on social media. IMT proposes an interaction framework by which a user (the human teacher) without expertise in machine learning can train a model (the algorithmic learner) to accomplish desired tasks with limited amounts of pre-labelled data. Applications of IMT in social media settings, however, has been limited. To extend IMT's framework to social media, we conducted a think-aloud study ($N = 24$) with users from four feed-based social media platforms with diverse cultures and affordances—Instagram, Mastodon, TikTok, and Twitter[2]—to answer the following question:

> **What are prominent signals relied upon by users to judge the value of content in their feeds, and are thus amenable to teaching to an algorithmic learner?**

We define a "signal" as a pair of one feature (a category of information that can be extracted from a post, such as the AUTHOR or INCLUDED HASHTAGS) and one characteristic (a statement describing the significance of the feature, such as *"is part of a recent fashion trend I've been following"* for the feature INCLUDED HASHTAGS)[3]. We define "value" broadly based on the *desired order of consumption*: Post A has a higher value than Post B if the user prefers to see A before B in their feed.

We find from our study that users leveraged a variety of signals to evaluate content from their feeds. These evaluations are nuanced in ways that current preference elicitation methods, such as "likes" or content dwell time, may fail to capture. Many users also expressed desires for feed experiences centered around individuals they cared about rather than content-based recommendations, better management of saved content, and more agency—in particular self-causality—when curating their feeds. Supplementing our findings with prior work on IMT, we offer five IMT-inspired design principles for teachable feeds. Finally, we embody these principles into three proposed feed designs that serve as sensitizing concepts to catalyze future research in this area.

Concretely, our paper offers the following contributions:

(1) Cross-platform taxonomies of prominent signals used to determine the value of posts in social media feeds, enriched by themes extracted from user interviews.
(2) Five principles to guide the design of teachable social media feed experiences.
(3) Three proposed feed designs that illustrate our principles and serve as sensitizing concepts for teachable social media feed experiences going forward.

These contributions pave a path to equipping today's social media systems with novel design patterns to empower agential, personalized feed curation.

2 RELATED WORK

To motivate our work, we review prior literature on agency and algorithms on social media, incorporating human values into content distribution algorithms (most notably recommender systems), and interactive machine teaching.

2.1 Agency and Algorithms on Social Media

In HCI literature, increasing user agency is often discussed as an aspirational ideal [6, 7, 20, 67, 70, 88, 115]. But how exactly do HCI researchers understand agency? Bennett et al. [7] surveyed 161 publications across 30+ years of HCI and identified 4 key aspects of agency: self-causality/identity, material/experiential, interaction time-scales, and tradeoff between independence and interdependence. In our work, we draw mostly from the aspect of self-causality/identity[4], which refers to the level and directness of a user's decision-making and action execution in line with their own values. This concept is similar the concept of agency in cognitive science, which refers to the sense of deliberately controlling one's actions and affecting the world through those actions [20]. From a more normative angle, prior literature has also argued that agency holds intrinsic value as a "fundamental human need" [115], a "basic psychological need" [67], and a "moral right" [91]. We integrate these perspectives into our work.

Reduction of user agency is a common concern on social media [67, 78, 108]. A primary source of this concern is the opacity with which social media feed algorithms operate. Indeed, these algorithms are commonly referred to as "black-boxes" in HCI and social science literature [5, 19, 69, 84, 91]. While prior work in explainable AI for social media has attempted to open the black-box and make the algorithms more understandable to lay users [1, 27, 56, 59, 60, 95], even a fully transparent "glass-box" algorithm may still reduce agency. For one, attempts to explain a complex algorithm may trigger information overload, weakening users' decision-making abilities [87]. Transparency also does not guarantee self-causality—users may watch and understand the inner workings of the algorithm without any opportunity for control [66]. Transparency aside, widespread user behavior such as mindless scrolling [88] and dissociation [6] are telltale signs of users' agency loss when interacting with algorithmically-driven feeds.

Dwindling agency has led to users deriving "algorithmic folk theories" to make sense of their social media experiences [23, 24, 30, 31, 51, 98]. Examples of folk theories include that users will see more content from friends who are more similar to them in their Facebook News Feed [30] or that the TikTok For You Page algorithm prioritizes videos that feature aesthetics associated with wealthier lifestyles (e.g., large houses) [51]. Theories may emerge from both *endogenous* (originating within the platform, such as content patterns, friend count, likes, and comments [8]) as well as *exogenous* (originates outside the platform, such as user location [13]) information [23, 92].

[1] By "experiences," we refer to the combination of in-feed interface affordances and underlying algorithmic behavior orchestrated by, or enabled using, those affordances.
[2] As of July 23, 2023, Twitter has been rebranded to X. Since the platform was still known as Twitter during the study, we will refer to it as so throughout the paper.
[3] For more on features and characteristics, see Section 3.1.2.
[4] Hereafter, we refer to this aspect as simply "self-causality."

Upon formulating their theories, users applied them to better align the algorithm to their preferences through strategic in-feed interactions. For example, Facebook users regularly visited profile pages of those whom they wanted to see more of and sought out more opportunities to tag them in posts, to signal their preferences to the algorithm [30]. To attempt to spread a video to other users' For You pages, TikTok users watched videos multiple times (even when they understood it perfectly the first time through), left longer comments, and hit the like/share buttons repeatedly even though they could only like or share a video once [51]. On Twitter, users systematically created and shared content with certain words omitted or misspelled to evade the surveillance of the algorithm [14]. Users, however, remained uncertain of the efficacy of their techniques and even considered their actions to be manipulative and forced [14, 30].

In our work, we seek to bring forth interactive tuning as a core component of social media algorithms, instead of treating such interactions as under-the-radar workarounds for preference and value alignment. We frame these interactions as teaching to highlight two essential qualities: naturalness and agency. As teaching is an inherent human ability that we both perform and receive throughout our lives, we are already naturally acquainted with its methods and mental models. Additionally, teaching is an agential activity—teachers are in charge of the teaching curriculum, delivering concepts to students, and evaluating student performance. Given this, how might we leverage our well-acquainted mental models of teaching in our online algorithmic interactions? We look to the paradigm of interactive machine teaching [79, 90, 111] for inspiration.

2.2 Human Values in Recommender Systems

Here, we focus on a particular class of content algorithms ubiquitous on social media and other everyday applications—recommender systems. A growing body of work proposes techniques and broader calls to action to embed human values into recommender systems [21, 36, 64, 82, 103, 104], drawing upon the field of value-sensitive design [11, 35]. Borning and Muller define "value" as "what a person or group of people consider important in life" [11]. We borrow this definition of value in our work; we use the term "values[5]" to abstractly refer to a unique set of important considerations of an individual or group, and "preferences" to refer to particular considerations within that set.

Many modern recommender systems *infer* human values by implicitly learning them from user information and interaction history [78, 104]. Commonly tracked attributes for implicitly eliciting values include tracking clicks [117], content dwell time [114], and affinity with other users [43, 50]. Implicit learning can be desirable as it reduces interface-level friction and allows for more effortless and rapid consumption of content [16, 57]. Additionally, advancements in deep reinforcement learning have improved the robustness and sophistication of implicit learning to the point that learned systems are, at times, capable of truly reflecting users' values [16, 62]. For example, users were startled by the accuracy with which TikTok's For You Page algorithm could capture their preferences without any explicit signals [99]. However, an important downside of implicit learning is its inability to facilitate user agency. A 2021 investigation of TikTok's algorithm found that it quickly led users down niche content rabbit holes towards "fringe" content [112]. Others have raised concerns about recommender systems' ability to distort speech [94, 105], exacerbate polarization [2], discriminate users and creators [61], and erode mental health [68, 103].

Given the concerns around implicit learning, many have sought to develop recommender systems with *explicit controls* [25, 36, 42, 57, 65, 77, 80, 86]. Users have reported higher levels of satisfaction [46, 55], trust [80], and engagement [42, 65] when they were given opportunities to exert control over the system, even when the controls had no impact on the output [7, 110]. Examples of explicit controls include thumbs up/down buttons to rate recommended items [37, 116], sliders and toggles for adjusting desired content characteristics [44], drag-and-drop topic specifiers [25], keyword critique [86], and manual selection of the recommender algorithm [9, 28]. More recent works have leveraged the semantic capabilities of large language models to enable the expression of preferences conversationally through a chat interface [36] and through editing a natural language user profile [77]. Explicit controls come with their own set of challenges—users may not know that these controls exist or what they do [39, 44, 100], find them cumbersome to use and keep up-to-date [42], or do not see value in engaging with them [58]. Indeed, prior work showed that most users prefer a hybrid approach that combines implicit and explicit learning [57, 73]. One promising approach in this vein is to design controls that allow for simultaneous expression of direct feedback and less direct social signals; real-world examples include "react" options on Facebook and LinkedIn [104].

A common theme across both implicit and explicit learning is that particular content features are assumed to be more important to the user—features that the system then learns on. However, few works have questioned whether those features are truly ones a user cares about or wants the system to learn when they consume content. A news recommender may accurately capture (implicitly, explicitly, or some combination of both) a user's preferred article length, but will still fail to align with user values if the user does not care much for article length. In our work, we seek to elicit *which content features users truly value* in social media feeds to orient future work in value-sensitive recommender systems.

2.3 Interactive Machine Teaching

Interactive machine teaching (IMT) is an interaction framework by which subject matter experts—who are often not experts in machine learning—draw upon their personal expertise to train machine learning (ML) models that can operate effectively within their domain [32, 48, 79, 90, 111, 119]. While conventional ML is primarily concerned with developing algorithms that automatically learn conceptual representations from training data, IMT argues that learnable representations should directly come from human knowledge [79, 111]. This way, users feel more agency while maintaining a firmer grasp of what the model learns, making models more transparent and debuggable [54, 90]. IMT consists of three main stages that form a "teaching loop" [90]:

(1) **Planning:** the human teacher (the subject matter expert) identifies a task for the algorithmic learner (the machine

[5]Note the distinction between "values" and the "value" of a post. The latter is defined in Section 1.

learning model) to complete, along with a curriculum (set of examples and representations to help teach the learner, typically in the form of a small dataset).
(2) **Explaining:** the teacher shows the learner examples and explicitly identifies concepts the agent should learn.
(3) **Reviewing:** the teacher allows the learner to predict some unseen examples, corrects any erroneous predictions, and updates the teaching strategy/curriculum accordingly.

Central to IMT is the *teaching language*, an interface by which the teacher crafts expressive representations that communicate desired concepts and are learnable by an algorithmic agent [90]. In order to derive and use a teaching language, users must engage in a practice known as *knowledge decomposition*. Ng et al. [79] define knowledge decomposition as "a process of identifying and expressing useful knowledge by breaking it down into its constituent parts or relationships."

IMT shares similar goals with the paradigm of reinforcement learning and, more specifically, reinforcement learning with human feedback (RLHF) [81]. Both IMT and RLHF strive to interactively and iteratively embed specialized human knowledge into a machine-learned system. In RLHF, an algorithmic agent strives to find a *policy* (a map of the agent's states to actions) that maximizes some long-run measure of reinforcement as defined by a *reward function* in a dynamic environment [18, 49, 106], after which human feedback is be used by the agent to fine-tune the model to act in accordance with the user's intentions [81, 102]. The key difference between IMT and RLHF, however, is that in IMT, the human teacher has agency—defining the curriculum, explaining concepts, and evaluating performance—whereas in RL(HF), the human is merely an oracle that the agent queries to guide its decisions [18]. While we see potential in both IMT and RLHF to better align a feed curation agent with user values, we seek to center user agency throughout the feed curation experience. As such, we focus on IMT and defer exploration of RLHF approaches to future work.

How can IMT inform end-user empowerment in social media feed curation? Applications of IMT in social media contexts so far have been limited. Jahanbakhsh et al. [41] employed an iterative loop bearing some resemblance to IMT to train a personalized AI capable of learning and predicting a user's misinformation assessments, but identifying misinformation is just one of many salient dimensions for users when consuming content in social media feeds. Additionally, higher-level tasks such as misinformation assessment may be composed of lower-level observations such as low-quality images or suspicious links that can be informative characteristics for curating content in other scenarios. Our paper seeks to draw out such observations through knowledge decomposition to articulate the design space for teachable social media feed experiences.

3 METHOD

In this work, we map out the design space for teachable social media feed experiences through the lens of IMT. Specifically, we sought to understand *what* or *how* users would teach a personalized feed curation agent. To answer our research questions, we first designed a study inspired by techniques in knowledge decomposition [79] to obtain taxonomies of salient signals for assessing the value[6] of the posts in a social media feed. We supplemented our taxonomies with qualitative interview data from the study.

3.1 Signal Elicitation Study

3.1.1 Participants. Our study was conducted with 24 regular social media users between the ages of 18–65. We recruited some participants through the platforms of interest in the study (Instagram, Mastodon, TikTok, and Twitter; see Section 3.1.3 for more details) and others through our institution's Slack channels and word-of-mouth. Prospective participants indicated which of the four platforms of interest they used, their frequency of use for each platform, and basic demographic information such as age range and gender, in our screening form. We started by accepting participants who used their platform(s) at least a few times a month, on a first come, first-serve basis. Selected participants then chose one of the four platforms of interest to use for the study. Later on, because we wanted similar numbers of participants for each platform, we only sent invites to participants who had a higher likelihood of choosing a platform for which we were seeking more participation, based on their indicated frequency of use for that platform. We stopped recruiting when we reached data saturation across all platforms.

In the end, we had an even distribution of participants across our 4 platforms, such that there were 6 participants per platform. The far majority of participants (20) used their platform of choice daily, while 3 used it a few times a week and 1 used it a few times a month. Half of our participants (12) were aged 18–24, 6 were 25–34, 3 were 35–44, 3 were 45–54, and 1 was 55–64. 14 identified as women, 8 as men, and 2 identified as non-binary. All had at least 3 years of experience on social media, with all but one having at least 5 years and just over half (14) having at least 10 years.

We note that our participant pool may not be a representative sample of social media users, neither on a per-platform basis nor in aggregate. However, our study's intention is to map out a design space illustrating some interactive possibilities for teachable feed experiences, rather than to make generalizable assertions about the relative frequency or importance of observed interactions. We also recognize that our resulting taxonomy is not exhaustive and new signals may emerge from specialized use cases and communities.

3.1.2 Study setup and procedure. To formalize the knowledge decomposition process [79], we defined a *signal* as an information unit consisting of two primary components: a **feature** and a **characteristic** (see Fig. 1). A **feature** is a class of information that can be extracted from a post. Example features include the post's author, the textual content, image(s), the number of likes, and the post's topic. A **characteristic** is a subjective statement that describes a feature. It is subjective in the sense that its significance may vary between participants. For example, the feature "the post's author." is described by the characteristic "is someone I know from in-person interactions." A third, optional component, an **action**, is something the participant can perform to the post in response to a feature-characteristic combination. For example, if the post author is someone the participant knows through in-person interactions, they may choose to take an action of "trigger a notification" for

[6] In this work, we define value broadly based on what participants want to see before others. Posts that participants want to see first first in their feeds has the highest value.

posts from that author. Actions were an optional component of a signal and could also be triggered by any combination of feature-characteristic pairs.

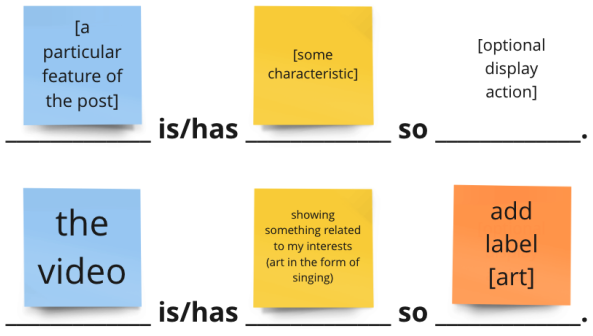

Figure 1: Examples of signals from our study. On top is a default signal template that we provide to users, and on the bottom is one completed by a participant.

Participants were asked to compose signals (Fig. 1) and arrange them on a Miro board.[7] Our study was conducted virtually, 1:1 through Miro and Zoom. We invited each participant onto a copy of the board and interactively co-authored signals with them in Miro while communicating through Zoom. Before the start of the study, participants submitted 10 screenshots of content on their own social media feeds from a platform of their choice out of Instagram, Mastodon, TikTok, or Twitter. We specifically asked for the first 10 posts they saw (and were comfortable sharing with us) in their "home"[8] feed. Participants' posts were privately confined to their board and were not shared with anyone beyond the research team.

The study board was separated into 5 main areas; Fig. 2 shows the board in its initial state. We populated the 10 screenshots participants submitted to us to area (3) in a random order before the start of the study. Area (1) contained editable signal templates while area (2) contained optional elements such as actions and boxes for participants to visually group posts they want to keep together. Area (4) was a space for participants to construct an "ideal" feed by organizing their posts into an upper, middle, and lower feed. While a linear feed may not have this exact distinction, we used it to roughly separate participants' perceived value of posts into 3 categories of descending value. Finally, area (5) was where participants could place any content that they would like to remove from the "ideal" feed they were assembling. An example of a Miro board after the activity has been completed can be seen in Fig. 3.

Using this board, we first asked participants to drag their content from area (3) into one of the 3 feed sections in areas (4) or area (5). We then asked them to further elaborate on their organization (why they value or do not value a particular post) by writing at least one signal using the templates in area (2) for all 10 posts. We co-authored these signals with participants by collaboratively editing the text in the signal template until participants were satisfied that the signal accurately represented their perspectives. Because the signals conformed to given templates, we relied on participants' think-aloud dialogue, which was recorded and transcribed, to capture their more nuanced reasoning. Many participants wrote more than one signal per post as there were multiple features they took into consideration. After all 10 posts were associated with at least one signal, we allowed participants to optionally assign actions to the signals.

Upon conclusion of this interactive activity, we conducted a brief interview where participants reflected on their most frequently mentioned features and characteristics, as well as general experiences and desiderata with social media feeds. In total, the study took around 45 minutes to complete.

Our study was reviewed and approved by our institution's IRB under Study #00016757. All participants received a $20 USD gift card after completing the study. All studies were conducted virtually on Zoom between January and April 2023, audio recorded, and transcribed.

3.1.3 Platforms. Our participants submitted content from their choice of one of 4 platforms: Instagram, Mastodon, TikTok, and Twitter. We were interested in these platforms as they were already popular or were gaining popularity at the time of the study, and also varied along two dimensions of interest: content format and engagement optimization. *Content format* refers to the primary format with which content is displayed and consumed on the platform. For Mastodon, this was text, while Twitter uses a combination of text and visual media (images and videos). Instagram is mostly image-based, while TikTok is mostly video-based. *Engagement optimization* refers to the extent to which a platform draws from its broader content pool based on a user's prior engagement with a piece of content, rather than relying solely on those the user follows. Mastodon is strictly reverse chronological and has zero engagement optimization, while Twitter provides users with a choice of viewing an engagement-optimized "For you" feed and a more lightly optimized "Following" feed. Instagram offers an engagement-optimized home feed by default. TikTok is well-known for its engagement optimization on its For You feed [112].

We recognize that users may choose to engage with these platforms for different reasons—for example, one may log onto Instagram to share pictures from their personal life and catch up with friends, while only logging onto Mastodon for professional networking. We see this as a potentially rich source of insight in our study. Feed design and affordances are heavily influential in shaping perceptions of what a platform is best used for [52], and hearing participants' varied interaction strategies across different use cases can help us better envision the design principles and affordances users may seek given a particular use case or goal.

3.2 Data Analysis

Our study generated data in the form of 411 signals (with some repetition within and between participants) and 24 interview transcripts. We analyzed the two data sources separately while noting complementary and contrasting themes between the two.

[7] Miro is an interactive and collaborative virtual whiteboarding tool [75].
[8] On Instagram and Mastodon, this refers to the main feed users see when they log in. For TikTok, we equated the home feed with the For You Page. For Twitter, we equated it with the Following feed, as Twitter was just starting to roll out its For You feed as we were conducting our study.

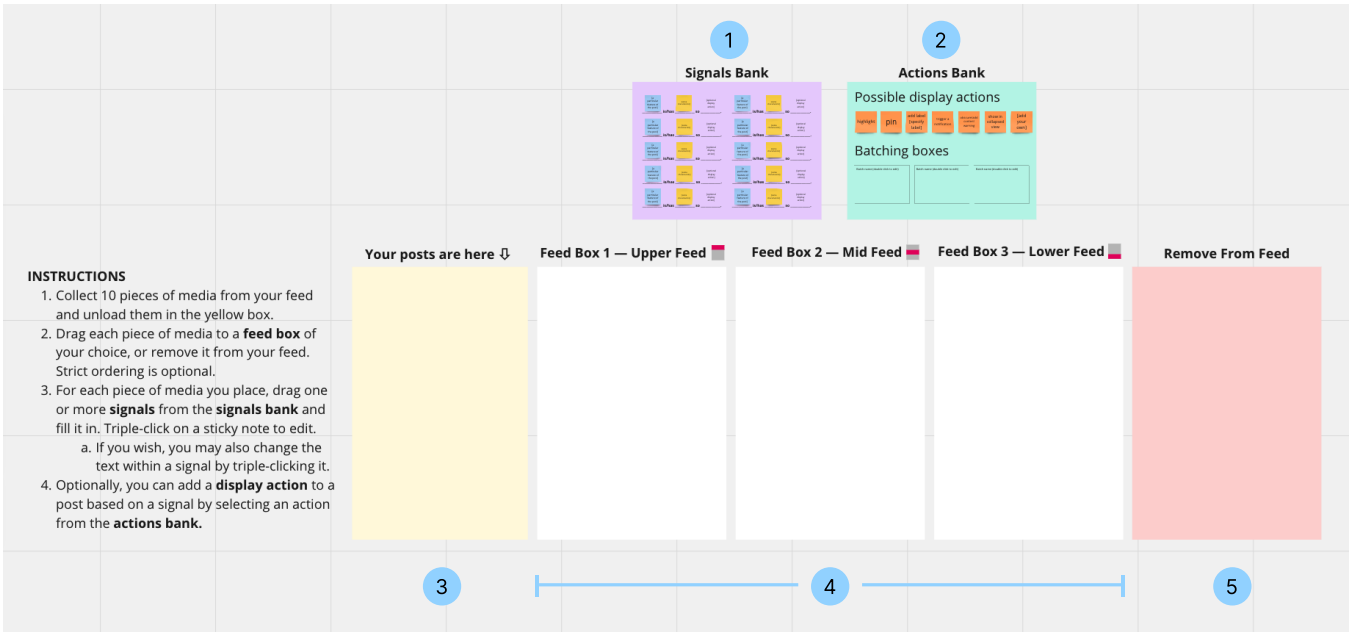

Figure 2: Areas of the Miro board. (1): signals bank with editable signal templates. (2): actions bank with sample actions and boxes for grouping content. (3): area where we uploaded participants' posts prior to the study. (4): the upper, middle, and lower feed boxes. (5): area for placing content that the participant would like to be removed from their feed.

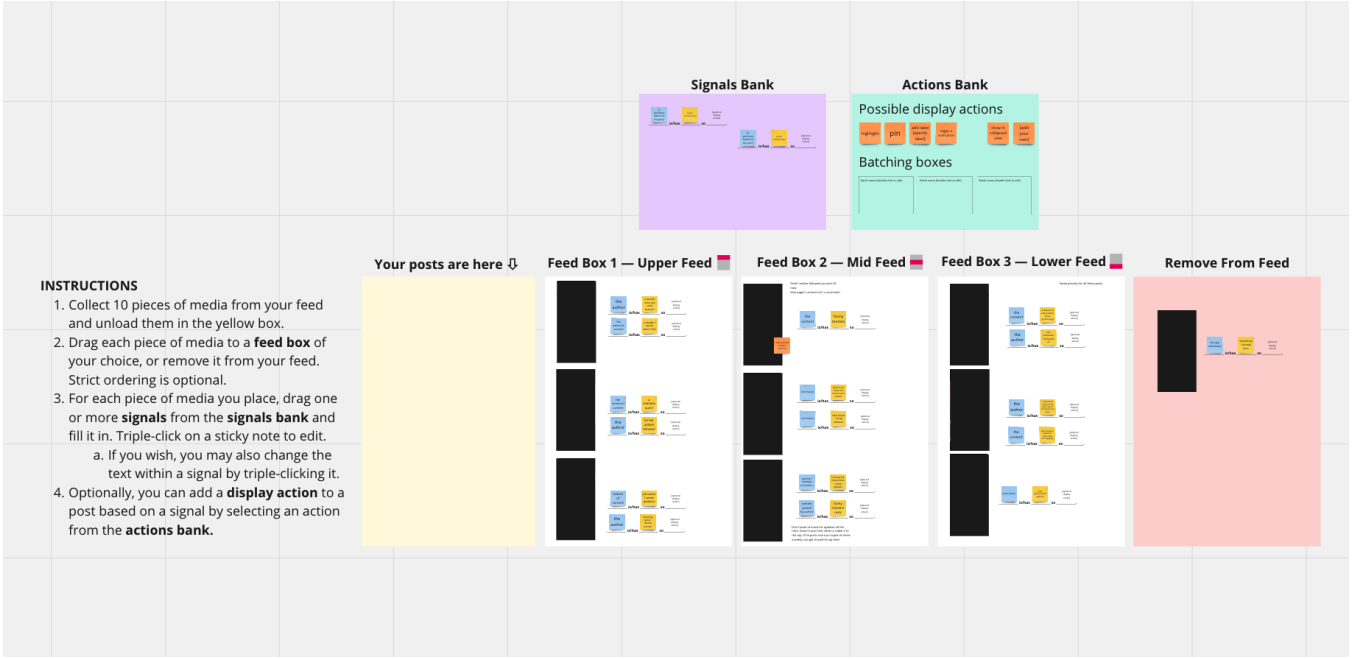

Figure 3: An example of a Miro board after a participant has completed the signal elicitation study. The posts themselves have been obscured by the research team to preserve the participant's privacy.

3.2.1 Signal data. After all studies were completed, the first author aggregated signals across all participants and processed them into a spreadsheet to preserve pairing between features and characteristics. Within each group, the first author separated features or characteristics referring to an *account* from which a post was made

from those referring to a post's *content*,[9] looking up the pairing in the spreadsheet as necessary.

We used a hybrid approach of inductive and deductive coding [33] to transform our data into taxonomies. The first author inductively sorted features and characteristics into broad themes through affinity diagramming; this process was independently performed on features and characteristics. Results from our inductive analysis were discussed and iterated upon with the broader research team at weekly meetings. For features specifically, the first author referenced common information available on interfaces of social media posts (e.g., username, handle, like/reshare/comment buttons) to further code the features in a deductive manner. Initially, the coding was not informed by the need to accommodate multiple platforms and some platform-specific features, such as lists on Twitter and Mastodon, were included. However, it became clear as coding proceeded that the vast majority of features were shared across all platforms in our study. As a result, we incorporated platform-specific codes into platform-agnostic ones—for example, the "list" code was split up and merged into multiple account-related features depending on how participants discussed their use of lists. We instead relied on our interview data to capture discussions of platform-specific experiences.

Our initial round of coding produced 7 and 16 account-related features and characteristics, respectively, and 20 and 31 content-related features and characteristics, respectively. The results seemed excessively granular for a taxonomy, so we iteratively grouped our codes into higher-level labels, discussing with the rest of the team as necessary. Labels that fell outside of this paper's scope were also eliminated. Our final set of labels contain 4 and 10 account-related features and characteristics, respectively, and 9 and 18 content-related features and characteristics, respectively. Features and characteristics defined the vertical and horizontal axes of our taxonomies, respectively. We did not include an analysis of the "action" component of signals as participants used them sparingly in the study. The definitions for all labels in our taxonomies can be found in our Supplementary Materials.

The intersection of each feature-characteristic theme was tallied and marked with either "+" to indicate that this combination had positive value (increased a post's perceived value), "−" to indicate negative value (decreased perceived value), or "○" to indicate that there was no consensus among participants. Although we tallied exact counts, we expressed them as broader value ranges using saturation maps in the taxonomies to embrace the qualitative nature of our data and avoid invoking notions of generalization or numerical comparisons based on our counts.

After our aggregate taxonomy with data from all platforms was finalized, the first three authors then completed the same taxonomy using disaggregated, platform-specific data. We separated the signals by platform and deductively coded the features and characteristics using the themes we developed from the aggregate data, resolving any disagreements in weekly data analysis meetings.

3.2.2 Interview data. We took an inductive approach to performing thematic analysis on our interview data. Interview transcripts from the studies were initially auto-generated from Zoom recordings, after which the first author manually reviewed them while referencing the audio recordings to correct any incorrectly transcribed text. The first author took a first pass of open coding over the data, identifying insightful regions of the transcript and possible themes for further analysis. The first three authors then developed a codebook containing themes from open coded data that the team agreed were relevant and interesting. This codebook was used to collaboratively code the transcripts, with disagreements resolved through discussions among coders. Coders also wrote summary memos from the themes and discussed them with the broader team at weekly meetings, iterating on the themes and memos as necessary. Initially, the codebook consisted of 7 high-level themes synthesized from over 20 sub-themes. After some deliberation among the team, some high-level themes were combined and sub-themes were remixed to form the 5 high-level themes in the final codebook. Our final codebook is shown in Table 1.

[9]This step was necessary for a coherent analysis as the two were fundamentally distinct (see Section 4.1)

4 FINDINGS

We begin this section by presenting our taxonomies for account- and content-based features and characteristics. We then showcase findings from our transcribed data, some of which complement our taxonomy data while others reveal what our taxonomies could not capture by themselves. Throughout this section, we use a two-letter abbreviation (IG for Instagram, MA for Mastodon, TT for TikTok, and TW for Twitter) to indicate a participant's chosen platform in the study.

4.1 Taxonomies

Our analysis yielded two distinct sets of taxonomies: account-based (with features and characteristics relating to accounts that post and engage with content) and content-based (with features and characteristics relating to content that accounts post). We made this distinction as the two sets were fundamentally different in nature—characteristics from one were incompatible with features from the other, and vice versa. Definitions for all labels used in our taxonomies can be found in our Supplementary Materials.

Fig 4 shows our account-based taxonomy with data aggregated across all four platforms. We see participants rely heavily upon the ORIGINAL POSTER as a feature when evaluating a post. This feature is described with a wide range of characteristics—that is, there are diverse reasons why participants may pay attention to the original poster. Some characteristics, such as *"knows personally"* depend on personal context and even offline interactions the user may have had with the account holder. Other prominent characteristics were concerned with users' evaluation of the account holder's personal traits (e.g., if the account holder had a trait that the participant admired) or some trend in the content they released (e.g., alignment with the participants' topical interests). We note that recognizing an account from network effects is the only characteristic that garnered mixed valuations, dependent on whether the participant's network portrayed the account in a positive light. In our disaggregated, per-platform view of the same taxonomy (Supplementary Materials), we observe that TikTok's taxonomy is noticeably sparser than the others, indicating there is less interest in account-related features, a possible result of platform design. We dive into a deeper investigation of this with our transcript data (see Section 4.2.2).

Theme	Description
Passive and active use of feeds	Participants' engagement in passive browsing, active consumption, or a combination of the two patterns.
People- vs. content-centric feeds	Participants' (often platform-specific) preferences for people-centric or content-centric feed experiences.
Unactualized value of saved content	Participants' strategies and pain points in using existing archival features to to revisit content on platforms.
Nuanced evaluation of posts	Participants' consideration of positive and negative facets of a post, as well as factors outside the post context, when judging a post's value.
Self-causality and lack thereof	Participants' experiences with (the lack of) algorithmic control and transparency while curating content.

Table 1: Our 5 themes based on analysis of the signal elicitation study participants' responses

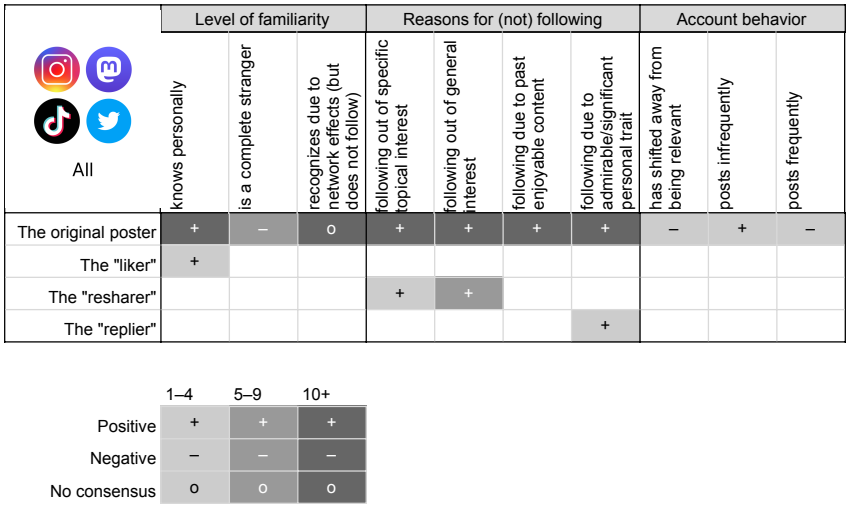

Figure 4: Our account-based taxonomy, aggregated across Instagram, Mastodon, TikTok, and Twitter.

Our aggregate content-based taxonomy (Fig. 5) shows that participants' evaluation of content is distributed over a much greater range of features when compared to our account-based taxonomy. Unsurprisingly, the most prominent characteristic associated with a negative evaluation is *"related to ads and consumerism."* Topic was a prominent feature contributing to positive evaluation, especially when the topic aligned with users' existing personal interests. However, topic differs from many of the other features as it is often *interpreted* by the user rather than *explicitly defined* in a post. Participants were also generally more drawn towards personal interests rather than professional ones in their feeds (except for Mastodon), although both were considered across all platforms. *"High consumption effort"* is a noteworthy characteristic due to the mixture of positive and negative evaluations in the same column. While some participants enjoyed shorter content due to their ease of consumption, P4[TT] saw longer-form content as more likely to contain material of interest. In our disaggregated taxonomies, we observe that TikTok was the sole contributor of the characteristic *"belongs to a notable trend/genre."* This leads us to believe that participants are more attuned to the trendiness of content on TikTok than on other platforms.

Ultimately, our taxonomies show that participants' evaluations of posts from their feed were diverse and multi-faceted. Such preferences are unlikely to be captured in their full richness with a simple "like" or even a "react" on today's platforms, nor a one-size-fits-all approach to feed curation. We thus see potential in leveraging techniques from IMT to design new affordances that allow users to teach more nuanced preferences to the algorithm for more agential and expressive curation.

4.2 Interview Findings

In addition to writing signals, we asked participants to think aloud throughout the activity and reflect on their broader experiences with social media feeds in a post-activity interview. Here, we discuss salient themes that emerged from our transcript data.

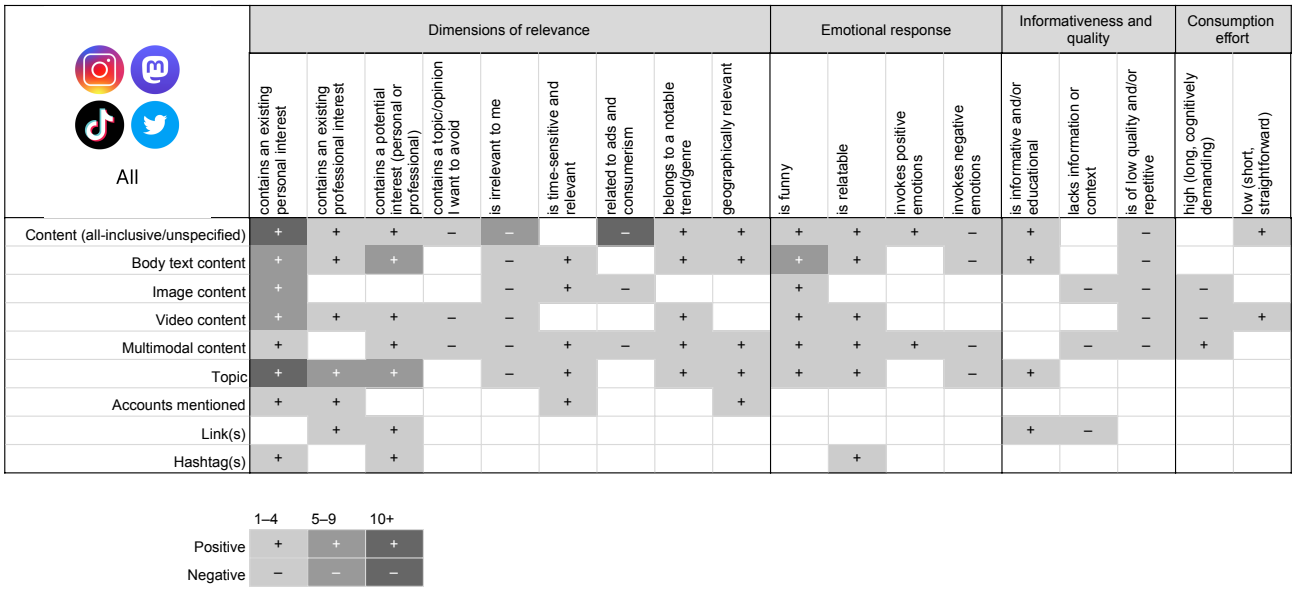

Figure 5: Our content-based taxonomy, aggregated across Instagram, Mastodon, TikTok, and Twitter.

4.2.1 Participants engaged in both undirected and directed use of their feeds. Many participants described using their feeds to consume information in an undirected manner. We characterize *undirected consumption* here as consumption without a specific goal in mind. Common scenarios in which participants engaged in undirected consumption include: (1) browsing the feed to be entertained (2) casually catching up on interesting conversations, and (3) instinctively opening the feed out of habit. This entertainment-induced passive consumption, which was especially common among our TikTok participants, can often lead participants to perceive content as more disposable. For example, P11[TT] noted that *"if [a post] doesn't raise a question or doesn't seem like something I'm interested in, then I'll just scroll away because there's an infinite number of videos I can watch."* Besides entertainment, participants also engaged in undirected consumption when casually browsing for interesting updates and catching up with relevant conversations. P2[TW] likened this to *"hallway chats [...] or water cooler conversations."* Finally, some admitted that their undirected browsing tendencies were simply driven by habit and instinct. P7[IG] shared that *"if I'm waiting for something, my hand just automatically goes [to my phone]."* P23[TT] felt similarly for TikTok: *"I feel like it's more so a habit of [...] opening the app, and just hoping I find something interesting."*

On the other hand, participants also mentioned many instances of using their feeds in a directed manner. We characterize *directed consumption* as information seeking with a specific goal. Active consumption was more frequently linked to the activity of *searching* rather than *browsing*. P4[TT] discussed their use of TikTok as *"kind of like a Google for me. I want to learn about, for example, these new car features, or 'how do I bake a cake?' And what are some good restaurants in [my area]?"* P24[MA] even kept a dedicated search column in their multi-column Mastodon feed to search for particular hashtags.

Given these characterizations, participants did not limit themselves to solely undirected or directed consumption. They would often use their feeds for both, switching between the two based on their goals and context. P15[TW] summarized this aptly:

> *"If I am in a low focus parsing mode, then I will read everything in the timeline. If I want to focus and get things that are a little bit more important [...] or I'm trying to answer a question, I'll say, 'I care about this topic, or I care about [hearing from] these experts.' Let me dig in and see what's going on here." (P15[TW])*

P13[TT] mentioned that they will engage in undirected browsing *"if I have time to fill"* but otherwise will actively seek or follow up on information, such as restaurant recommendations. P6[TW] used Twitter *"very intentionally for professional [networking]"* but is otherwise just *"scrolling and looking around."*

Our observation of participants' consumption behaviors not only highlights the need for feeds to accommodate diverse modes of browsing, but also enable users to fluidly move between them.

4.2.2 Tension existed between desires for people- and content-centric feed experiences. In recent years, social media platforms have shifted towards content-based recommendations in an attempt to more effectively drive user engagement, particularly for younger demographics [38]. While it is not clear from our taxonomies (Figures 4 and 5) alone how participants prioritized account-based or content-based signals, our interviews shed more light on this. Many expressed a desire for more *people-centric* experiences in their feeds—curating and viewing content based on individuals they care about rather than the contents of the post itself. To P17[MA], the post

author can serve as a reliable indicator of relevance: *"There are some people that I know 95% of the things they say are relevant to what I'm doing."* Indeed, when asked about the features they prioritized for post evaluation, all but 7 participants indicated an explicit preference for account-based features over content-based ones.

This people-centric mental model thus came into tension with platforms' increasingly content-centric recommendations. The most striking example of this tension was Instagram. Of the 6 Instagram users in our study, 4 saw no content from close friends or acquaintances in their first 10 posts of their home feed. P7[IG] felt that their feed hindered their ability to stay updated with those they cared about: *"None of these 10 posts were posts about people I know [...] And I'd like to know what's going on in their lives."* P12[IG] echoed this and the desire for prioritizing posts from specific people: *"I would love if I could see what my friends post at the top."* P8[IG] leveraged the 'Close Friends' feature [40] to combat the algorithm's content-based recommendations so that they could still *"be connected to [my friends]."*

Interestingly, 5 of the 7 users who did not express a preference for account-related features used TikTok for this study. For TikTok users such as P10[TT], their home feed (For You Page) is mostly filled with content from *"people I don't follow, or I've never heard of."* Unlike P17[MA], who trusted certain authors to post relevant content, P23[TT] was the opposite: *"even if it wasn't [this creator], I would stay tuned to the specific topic she's talking about."* A couple reasons (or combinations thereof) may prompt deeper engagement with content on TikTok. First, the primary use of TikTok was for entertainment (see Section 4.2.1). As is the case with YouTube, TikTok users' focus on entertainment may lead them to be more concerned with content consumption rather than fostering social interactions with people [53]. Second, the design of the TikTok interface may serve to prioritize content, more so than other platforms. As prior design analyses of social media UIs [52] point out, TikTok offers an immersive content-viewing interface with minimal whitespace—unlike other platforms, all other features within a post are overlaid on top of a TikTok video.

Our takeaway here is that both people- and content-centric feed experiences can be desirable. On some platforms, users prefer content-centric experiences; on others, they prefer people-centric ones. The tension between the two comes into play when, on a particular platform, users desire one but their feed forces them towards the other. Thus, when formulating designs for teachable feed experiences, we aim to equip users with necessary tools to agentially express their desired type of experience, and tweak that experience as they see fit.

4.2.3 Perceived value of saved content was often not actualized. Many participants used archival features (e.g., "save", "bookmark") on their platforms for a variety of reasons. For some, the times when they checked social media were not when they wanted to take action on a piece of content. For example, P2[TW] shared that they *"usually check social media in the mornings and then I have to work,"* so they saved posts related to science fiction (a topic of their interest) to read after work. P7[IG] saved encouraging quotes for *"when I'm sad, or when I really need a pep talk."* Over several months, P7[IG] also saved career-related posts in preparation for the upcoming recruiting season.

It is important to note the distinction between simply *allowing a user to save content* and allowing them to *curate their content archive* such that they can easily retrieve desired content. While all platforms in our study offered the ability to save content, there is much less support for managing saved content so participants could actualize their value. To circumvent the issue of not being able to retrieve saved content, P8[IG] said they would sometimes send content to another person instead of saving it on the platform so that they could rely on that person's memory to help retrieve the content later on: *"I would send to my husband a lot of the things I like [...] That's also a strategy for 'watch later.'"*. It also does not help that saved content often has low discoverability in platforms' interfaces. As of the time of writing, saved content in Instagram is 4 taps away from the home feed and requires traversing a user's profile menu, an area reserved for options such as privacy settings and payment information.

In many ways, an archive of saved content resembles a feed, but is free of algorithms—all curation happens manually. P24[MA] considers the lack of algorithms to be a double-edged sword: *"inside of [archives], I don't get the good of the algorithm, but at least I wasn't getting the bad stuff either."* While a sophisticated curation algorithm may be unnecessary, we recognize that an organized content archive can act as a collection of examples representative of user preferences, acting precisely like a curriculum in IMT. We further explore this idea in Section 6.2.

4.2.4 Participants' evaluations of posts were nuanced. Many current social media platforms have explicit and implicit means by which users can provide feedback to content algorithms [78, 104]. However, this feedback can fail to capture users' consideration for complex factors that influence their overall assessment. Participants often gave nuanced evaluations of posts which took into account multiple feature-characteristic pairs contributing to both positive and negative evaluations of a post, or factors outside the context of the post.

More complex evaluations were frequently due to conflicting assessments of different aspects of the content, such as a post that is about an undesirable topic but provides value to the participant in another way. For example, P10[TT] said that they are not interested in the Grammy Awards, but still wanted to keep related posts in their feed *"just to stay updated on it."* Similarly, P16[MA] described how an undesirable post could still enrich their social media experience: *"I wish to some degree that it wasn't posted but at the same time, it does have value to me. It gives me information about the person and it makes it feel like a real conversation."*

Sometimes, users may give mixed assessments of the post content and the post author. A negative assessment of the content of a post may outweigh a positive assessment of the author (or vice versa) in evaluating the desirability of a post in their feed. For example, in response to an offensive Instagram post, P22[IG] noted that *"there's no way for me to tell Instagram, 'don't show me pictures like this anymore,'"* although they still wanted to see posts from the author and *"don't really want to mute the person."* A negative evaluation of the content of a post may even impact a positive/neutral evaluation of the author, or vice versa. P1[MA] explained how the distasteful content of a toot on Mastodon led them to change their evaluation of the author: *I might even unfollow this person for tooting this thing*

out [...] [the post was] in some ways useful in that, like 'wow, I can't believe the person is buying into this garbage, I don't know if I can trust other stuff that they put in my feed.' This suggests that our two taxonomies, which capture both account- and content-based evaluations of posts, respectively, are not strictly independent and should be considered in concert.

In other cases, however, a participant's evaluation of a post was influenced by context not local to the post itself. For example, mood can affect a participant's social media preferences. P8[IG] preferred "short and sweet" posts when stressed, to avoid spending too much time on their phone, and P12[IG] only wanted to see *"corny Instagram captions"* when they were in a good mood. Such temporary preference shifts might extend beyond a user's mood on a given day and instead apply across a larger span of time. P17[MA] noted that while they typically did not mind seeing professional content on their Mastodon feed, they had *"been kind of overwhelmed with mid and senior career people recently."* Echoing this need, P19[MA] suggested a potential feed customization feature where users *"could just mute a category for a day, or for some period of time."*

Users' evaluations of posts involve multi-dimensional and often conflicting assessments of features within a post, as well as outside factors, including mood. We see a need for teachable experiences that accept both structured (e.g., via UI buttons) and unstructured feedback (e.g., via natural language) over longer timescales. This way, users have more flexibility to indicate temporary or context-specific preferences for feed curation.

4.2.5 Participants experienced a lack of self-causality in feed curation. Many participants did not trust algorithms' ability to curate feeds per their preferences. While describing why they were not interested in a post from an author they usually appreciate, P2[TW] said that they *"don't trust the algorithm to know"* that distinction. P9[IG] explicitly said that they *"don't want algorithms to try to sort the best content for me"* and would rather do it themselves.

P9[IG]'s statement demonstrated a desire for *self-causality*, a key aspect of agency defined by the ability to make personal choices and see these decisions reflected in an outcome in line with their values and goals [7]. Participants even prioritized self-causality above content enjoyment: P8[IG] said that while they sometimes enjoyed the algorithm's suggested posts, they *"would put it towards the end [of the feed], because it's not something I've decided I want to see."* This desire for greater control over feed content also led participants to take actions to experience self-causality, for example, by forming rudimentary feed training behaviors. P10[TT] shared that they *"will like videos that I liked and dislike on the ones I don't want to see anymore"*, because they are *"pretty sure the algorithm will learn what you like"*.

Participants' efforts to experience self-causality were hampered by an inability to determine why the algorithm fed them specific content. Participants often evaluated this based on different dimensions of relevance of the content in question, as represented in our content-based taxonomy. When describing how platforms know what kinds of feeds users are interested in, P23[TT] said that they *"think [the platforms] kind of guess"*, as they had a STEM-related feed that they said was *"cool, but I don't know why [the platform] would pick that feed specifically for me."* P10[TT] described how they were *"somehow on autism TikTok, even though I'm not diagnosed with it or anything,"* which *"feels weird"* as they *"don't know how I got there."* This uncertainty made it difficult for P10[TT] to use like/dislike signals to curate their feed, as they described the process as *"not foolproof... I will still get fully random [posts]."* Although participants could see that the algorithm was not accurately learning their preferences, they did not know why it was learning in this way or how to teach it more effectively, limiting their self-causality.

While participants expressed some negative sentiment towards algorithmic feed curation, we see that this sentiment may not be caused by the presence of algorithms themselves, but rather the lack of self-causality experienced in current platforms. IMT was motivated by similar concerns, and has the potential to increase algorithmic teachability and user agency if applied to feed settings.

We now proceed by synthesizing our study's findings through the lens of prior literature on IMT to inform design principles for teachable feed experiences.

5 DESIGN PRINCIPLES FOR TEACHABLE FEED EXPERIENCES

A primary objective of IMT is to leverage an end user's subject-matter expertise to train a learnable agent that can effectively aid the user in fulfilling their desired goals [79, 90, 119]. However, teaching languages[10] proposed in prior work, such as PICL [90] and Pearl [48], may not translate smoothly to social media settings—they demand high-effort, meticulous interactions from users to provide meaningful labels and descriptions from data examples. Given that social media is designed to support fast-paced consumption of information, actively introducing excessive friction via conventional IMT interfaces may worsen the user experience rather than improve it.

Based on our taxonomy and participant interviews, along with prior literature in IMT, we outline five design principles to set the stage for teachable social media feed experiences. We do so to extend the core principles of IMT to the modern social media landscape. In some cases, this extension directly builds off of classic IMT principles; in others, those principles are re-examined and reconfigured. Our principles may be used independently or (ideally) in combination.

D1: Situate the teaching language within the feed. In IMT, teaching has conventionally been performed in isolation from the environment in which the teaching materials originated. For example, when creating a recipe classifier in PICL [90], the user first imports a set of documents containing recipes before they can start teaching. Note that PICL is a separate environment from where users may encounter recipes in-the-wild, such as on websites. A separate teaching environment can enable more feature-rich and expressive teaching languages, but it can also disengage users from the act of teaching in social media settings. For one, the additional effort required to move posts and organize them outside of the feed is already seen as burdensome by our participants (see Section 4.2.3). Additionally, it may not be in platforms' best competitive interest to support exporting a post and any informative metadata off the platform. We therefore situate the teaching language *within the feed*

[10]Recall that a *teaching language* is an interface by which the human teacher can expressively communicate desired concepts for an algorithmic agent to learn.

itself. This way, we can directly leverage platforms' existing representations and users' existing mental models, while lowering the effort required to perform teaching. A key question then arises: what exactly do users' mental models of current platform representations look like in the context of feed curation? Our taxonomies shed valuable light on this; we can leverage salient features identified in our taxonomies to inform the design of our in-feed teaching language.

D2: Be available, but not intrusive. We heard from many participants that social media can act as a much-needed break or distraction in the middle of the day, a means of relaxation, and a casual time-killer when small pockets of idle time arise. That is, participants were satisfied simply by letting the algorithm entertain them. In these scenarios (described by P15[TW] as *"low focus parsing mode"*), persistently demanding extra attention and effort via IMT may in fact worsen the user experience and force users to expend more energy than they desire. We saw in practice that many participants switched between this low-focus mode and a more attentive, information-seeking mode (see Section 4.2.1), and that this switching was largely dependent on difficult-to-predict factors such as mood and context. In light of this, we ensure that our teaching language is available, but unobtrusive. That is, the interface is easily accessible to users across the entire feed experience, but can also be easily dismissed or hidden when not needed.

D3: Embrace a multiplicity of feeds. Some of our participants questioned why so many feed experiences had to be constrained to a singular feed and expressed a desire for multiple feeds to better organize their content. P12[IG] touched on the oddness of having diverse content mix in their feed: *"it's weird to go past someone's bikini photo and then, '5 people killed at a refugee camp.'"* Indeed, the range of characteristics used to describe features of posts as captured in our taxonomy corroborates this diversity. As a less extreme example, it may be jarring for users to see content that "contains an existing professional interest" if they expect to scroll through posts whose content "is funny." P16[MA] mentioned that they would like to create different feeds for the various Mastodon instances they were on. P8[IG] likened their feed to a newspaper and suggested different sections: *"news, updates, events, pop culture, etc."* In fact, platforms have also started to explore multi-feed experiences and broadening of algorithmic choice. Many, including Twitter, now offer an engagement-based "For You" feed alongside a "Following" feed featuring content from followed accounts. One smaller social platform called Bluesky has introduced Custom Feeds, where developers can create feed algorithms to which users can browse and subscribe [9]. TikTok has also introduced topic feeds based on inferred user interests [71]. We consider this shift towards feed multiplicity a promising direction, especially when providing users with a means of organizing their IMT curriculum.

D4: Seek structured and unstructured feedback. IMT workflows have traditionally been scaffolded with affordances that elicit structured feedback (e.g., data highlighting and labelling). Our findings in Section 4.2.4 and the range of both positive and negative feature-characteristic pairs in the taxonomy suggest the need for teachable algorithms that accept both structured and unstructured feedback in order to capture users' nuanced evaluation of social media posts, which corroborates existing research in IMT. For example, Jörke et al. [48] provided a form-like interface for specifying key teaching concepts in which some fields allow the selection of existing data tags while others allow the user to write freeform natural language. Likewise, Zhou et al. [119] noted the importance of relying on users' unstructured object show-and-tell gestures in addition to a structured UI. This combination of structured and unstructured feedback is desirable in real-world applications as users may make nuanced judgements that cannot be captured with structured feedback alone, but also require some guidance to express concepts in a format parsable by an algorithmic learner. Parsing unstructured input, however, is now significantly less challenging due to technological advancements in large language models. We take this into consideration when balancing structured and unstructured feedback.

D5: Enable teaching and evaluation at varying timescales. Bennett et al. [7] identified timescales of interaction—ranging from *micro-interactions* (a few seconds or less) to *episodes* (seconds to hours) to *life* (days to years)—to be a key aspect of human agency and autonomy. When asked to articulate their preferences with limited access to their feeds during the study, many participants had difficulty doing so because their preferences *evolved over time* based on the content they saw. As such, they expressed a desire to agentially refine their feeds' behavior over an extended time period. This also happened to be less cognitively demanding, as P6[TW] pointed out: *"I don't have to intentionally be like, okay, I'm gonna sit down and coordinate everything."* Our participants also described how their preferences shifted temporarily based on factors like mood (see Section 4.2.4). Indeed, we can see the pitfalls of assuming static preferences and designing only for interactions at short timescales via the inefficacy of set-and-forget personal content moderation tools [44]. That said, classic debates in HCI over direct manipulation interfaces (UIs providing immediate control feedback through elements such as buttons and sliders) versus interface agents (systems that perform actions on behalf of users, often after learning user preferences over time) reveal that aspects of both are vital in information-dense environments [96].We thus explore designs that enable teaching and evaluation at varying timescales. Specifically, we aim to leverage teaching interactions afforded through direct manipulation as a familiar interaction pattern to ground evaluations of algorithm performance over longer timescales. In doing so, we situate the teaching language as not only a tool for the teacher to articulate key concepts, but also one that aids evaluation of the algorithmic learner.

6 PROPOSED FEED DESIGNS

We now propose three feed designs for teachable social media feed experiences that embody our findings and design principles. Our goal is not to constrain the design space of teachable feeds with these designs, but rather present them as sensitizing concepts—emergent ideas that help direct attention to promising topics or phenomena [10, 120]—to illuminate salient paths for future research.

Note that the mockups illustrating our feed designs feature a generic microblogging platform, similar to Twitter or Mastodon. Our reasons for making this decision were twofold. First, we wanted to show how our designs can operate on platforms with a diverse range of feature and characteristic preferences, as opposed to ones like TikTok where a particular media modality (and therefore certain sets of features and characteristics) is significantly prioritized

over others. Second, we tap into the increasing design and development interest in microblogging alternatives since Twitter's change in ownership [47, 97]. Many platforms in this space, including Mastodon and Bluesky, are also open source, making experimentation with novel ideas more accessible than on closed platforms.

6.1 Exploded UI Views

In 3D diagramming, an *exploded view* is one where the individual components of an object are shown slightly separated from each other as if a small explosion occurred at the center of the object. Exploded diagrams depict inter-component relationships and are thus commonly found in instructional manuals. We employ a similar concept in a post's in-feed UI to serve as a teaching language for the elicitation of more granular preferences. On current platforms, "liking" a post can signal that the user enjoys *some feature(s)* in the post, but it does not clearly communicate *what specific feature(s)* they found like-worthy. An exploded UI view aims to recover this information by allowing the user to specify features within the post that they find (un)appealing.

Once a user expresses a generic signal (e.g., like, reshare) on a post, the UI explodes out into individual features, such as the author, text descriptions, attached media, and hashtags (all features in our taxonomies—see Section 4.1). These features may be directly available in the post or algorithmically extracted—indeed, TOPIC was a popular feature in our content-based taxonomy that needed to be inferred from posts (Fig. 5). In Fig. 6, we distinguish inferred features from extracted ones by rounding out their UI. The user can then engage in teaching the feed algorithm their preferences in this exploded view by selecting the features that they consider (ir)relevant. While we could further elicit detailed characteristics associated with those features, just like we did in our study's Miro board, we reduced this down to simple plus and minus buttons (indicating a positive or negative characteristic, respectively) to avoid overburdening the user.

We ensure that "explosion" happens within the feed so that teaching can be done without leaving the feed, satisfying **D1**. Additionally, since the exploded view can occupy more space than the normal post UI, especially when there are numerous features available to teach with, we provide a convenient option for the user to collapse back the exploded view (see Fig. 6 (B)). The user can also simply scroll away. This simple dismissal of the exploded UI aligns with **D2**. Finally, even as direct manipulation interfaces affording immediate interaction, exploded UIs allow users to gradually accumulate preferences informed by what they view, satisfying **D5**.

6.2 Multi-Feed Curriculum Organization and Seeding

A teaching language alone cannot close the teaching loop. Here, we propose a design in which assembling the teaching curriculum and evaluating the learner's performance is closely integrated with the teaching language via feed multiplicity. Key to this design is the use of curriculum organization to curate a multi-feed experience.

As a user expresses preferences using a teaching language, those preferences are saved into curriculum "folders." Due to the differing nature of account- and content-based preferences, we separate the two in Fig. 7. A folder can then spawn a new feed that aims to provide more focused content that adheres to the folder's theme. The post on which the user first expressed preferences becomes a "seed" for the new feed to guide the recommendation of related content. This multi-feed experience serves as a way for users to evaluate the algorithmic learner's performance—a relevant and well-curated feed is a sign that the learner is effectively acting on taught preferences. Otherwise, the user can provide feedback to the learner through the in-feed teaching language, closing the teaching loop.

If a user does not want a folder to form a feed, they can toggle it off in the curriculum; folders created from negative feedback are toggled off by default. Furthermore, as participants pointed out in Section 4.2.3, lightweight algorithmic interventions, such as suggesting folders for unorganized areas of the curriculum (see the "Unsorted" folder's Sort For Me option in Fig. 7) can also aid with curriculum organization.

Given that user trust hinges on observed learner performance [76, 111], additional support may be added to further facilitate learner evaluation. One possible approach is to reuse already-familiar interaction patterns in the teaching language scaffold evaluation. Fig. 8 offers an "Explain" option for posts recommended by the learner in a feed formed from a folder. Selecting that option would expand the post into an exploded UI with pre-selected preferences that the learner infers from existing ones. The titles of features then become explanations that link back to folders and other curriculum material used by the learner to infer that preference. With limited work in explanation and evaluation techniques in IMT [107], employing the teaching language as an aid for evaluation suggests one approach to mend this gap.

This design's use of feed multiplicity embodies **D3**. Additionally, unlike the set-and-forget approach to creating custom feeds on Bluesky [9], this design enables users to iteratively build, refine, and evaluate their feeds in the spirit of **D5**. This design can also be easily dismissed (**D2**): users can toggle feeds off from within the curriculum and the curriculum itself is located in another tab separate from regular feed activities. However, if the user does choose to engage with this design, the proximity to and reliance on an in-feed teaching language satisfies **D1**.

6.3 Purposefully Finite Feeds with Natural Language Feedback

The infinite scroll is a dominant design pattern in contemporary social media. Implementation-wise, this effect is achieved by loading posts in batches such that another batch of content is quickly available once the user reaches the end of the previous batch. What if we can repurpose this transition between batches into an opportunity for preference elicitation and reflection?

In this feed design, we explore what it means for feeds to be *purposefully finite*. We split a feed into individual "stacks" of content; users can set the size (number of posts) of each stack. When the user reaches the end of a stack, they are presented with a teaching language in the form of a text input area in which they can specify, in natural language, any preferences to incorporate into future stacks. The combination of finite stacks and natural language feedback addresses two insights participants raised in our study. First, preferences may not be well-formed before consuming content—by allowing users to first view content and then reflect on what they

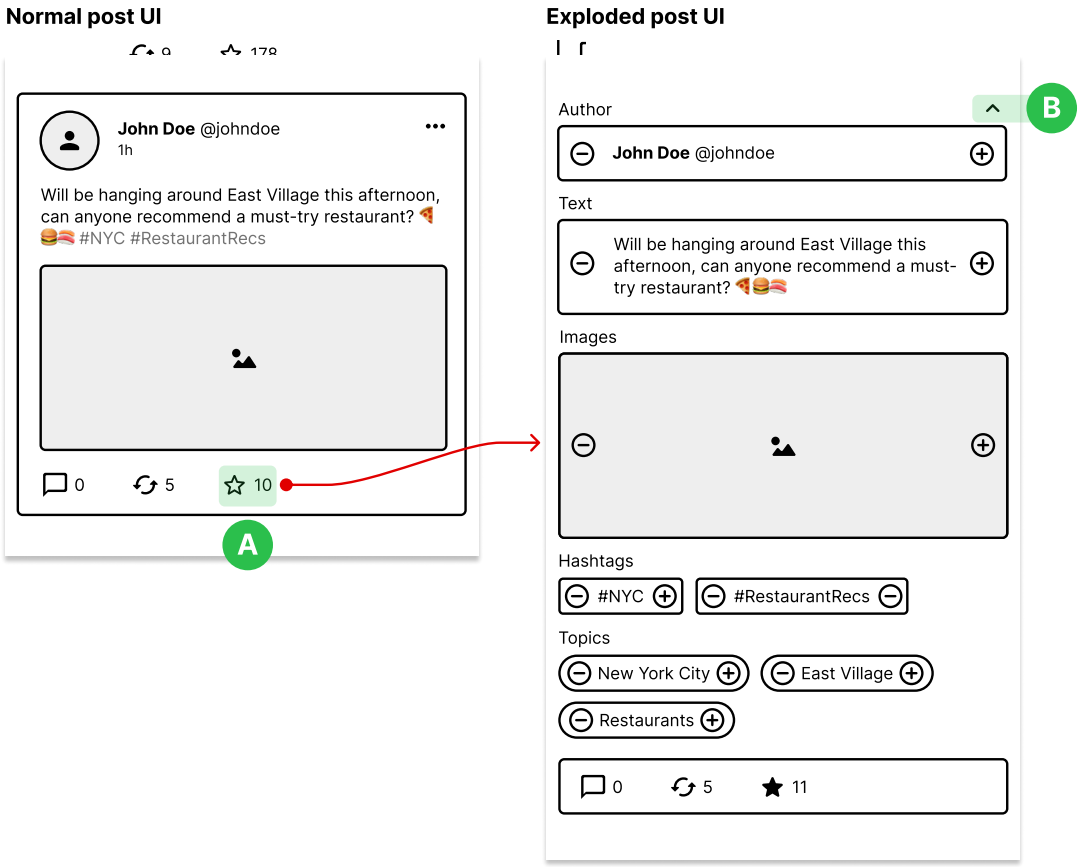

Figure 6: An example of the exploded view feed design with 5 features: author, text content, images, hashtags, and topics. (A): the exploded view is triggered as the user provides a generic signal—in this case, a "like." (B): the exploded UI can be collapsed back into the normal UI using the caret.

viewed, users may be able to articulate their preferences with more precision. Second, the feed was used to both browse and search for content. An adjustable stack size allows a user to smoothly transition between the two consumption modes. With a low stack size and frequent feedback input, the feedback starts to resemble intent-driven search queries, while a high stack size brings the experience closer to that of a conventional infinitely scrolling feed. On top of all this, unstructured natural language allows users to specify nuances that may be difficult to communicate through structured means, such as UI buttons.

The accumulation of natural language feedback also presents novel interaction opportunities. Users may use consecutive pieces of feedback to assemble a chain of preferences that can help them discover content with specific features and characteristics. Users can then exit from these more focused views by deleting preferences from their chain, or removing the entire chain to start afresh. Users' preferences can also be automatically summarized by the algorithmic learner into "Observations" (see Fig. 9) are shown to the user for additional reflection. These observations can be edited by the user to guide future recommendations, similar to editable natural language user profiles in modern conversational recommender systems [36].

This natural language feedback, when used in combination with the more structured feedback presented in Section 6.1, results in an expressive set of teaching languages (**D4**). Like exploded UI views, it also operates in-feed, per **D1**. If the user does not wish to engage in this design, they may simply proceed to the next stack without providing feedback, or revert to an infinitely scrolling feed by setting the stack size to infinity, satisfying **D2**. Finally, by first accumulating feedback over time as they view content, users can refine system behavior over longer timescales, while still being afforded the ability to directly and immediately update the system's learned knowledge as preferences evolve (**D5**).

7 DISCUSSION

7.1 Ethical Implications

Our goal of enabling teachable social media feed experiences is to empower users in reclaiming agency and enriching self-expression. While we hope that such a goal would bring positive change to

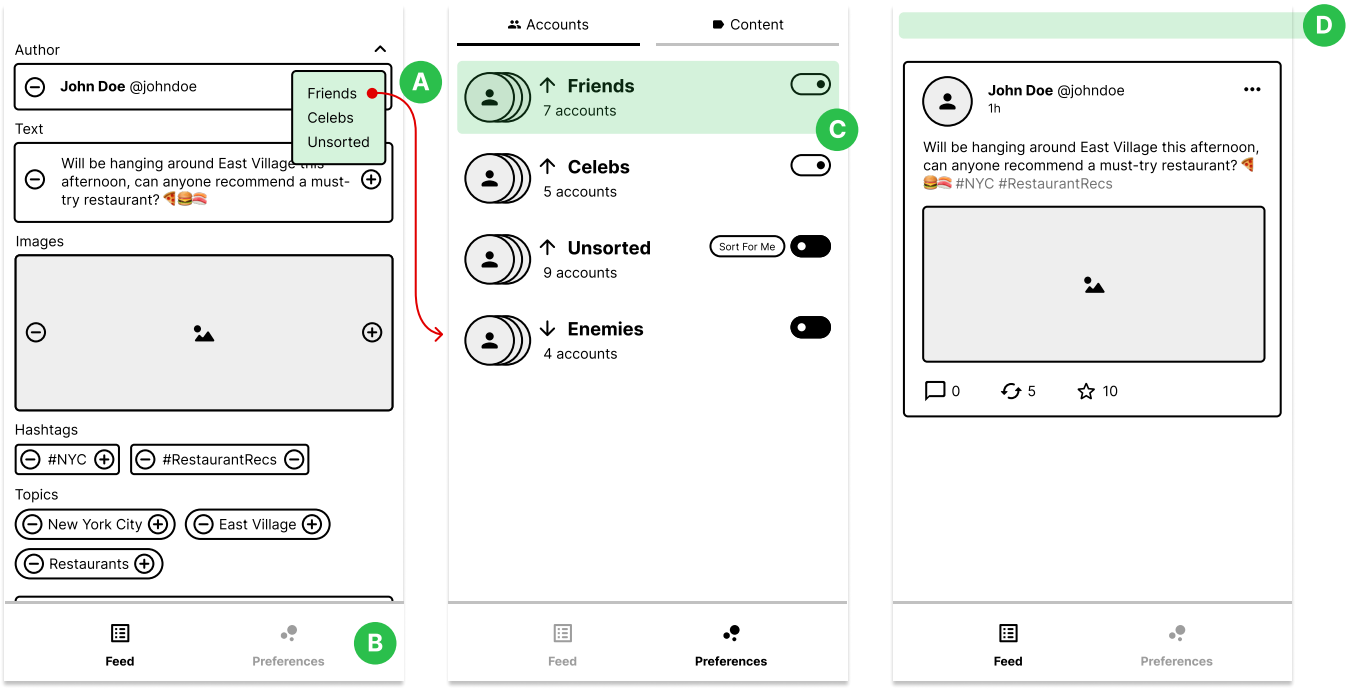

Figure 7: The teaching curriculum, depicted on the interface as "Preferences," as a multi-feed experience. (A): the positive account-based feedback provided by the user in the exploded UI view prompts further organization from the user before the account is deposited into the curriculum. (B): tab to easily switch between viewing the feed and the curriculum. (C): the user's categorization of the account as "Friend" creates a new entry in the corresponding curriculum folder. Folder titles have arrows indicating whether the folder was created from positive or negative feedback. (D): feeds are formed from folders that users have toggled on. The posts on which users first provided feedback act as seeds for the folder feed.

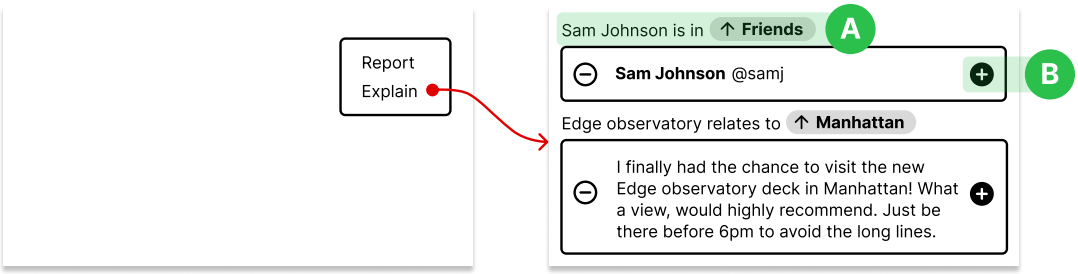

Figure 8: The teaching language (exploded UIs in this example) is invoked to scaffold learner evaluation once the "Explain" option is selected. (A): explanatory feature titles indicate relation to users' previously expressed preferences and links to an existing curriculum folder. (B): the learner pre-selects option(s) in the UI to show its preference prediction, which the user can correct by un-selecting if desired.

social media systems, certain factors, if not carefully considered, may lead to the opposite outcome.

7.1.1 Issues around (hyper)personalization. Granular preferences captured through teachable feeds can lead to hyper-personalization—the use of deeply personal traits such as one's personality, motivations, and goals [113, 118] to personalize experiences. Hyper-personalization may exacerbate privacy concerns many already have about social media platforms. In particular, users' lay knowledge of behavioral profiling for targeted advertisements has led to negative perceptions and distrust of platforms to responsibly store personal information [4, 109]. Prior work has also warned about social pressures that can lead to over-sharing of hyper-personalized information on social media [113].

More broadly, John Cheney-Lippold coined the term "dividuals" [17] to highlight an inherent paradox of personalization on large-scale, data-driven platforms: one's "personalized" experiences online are never truly local to an individual, but are instead constantly shifting collections of data points—dividuals—aggregated across

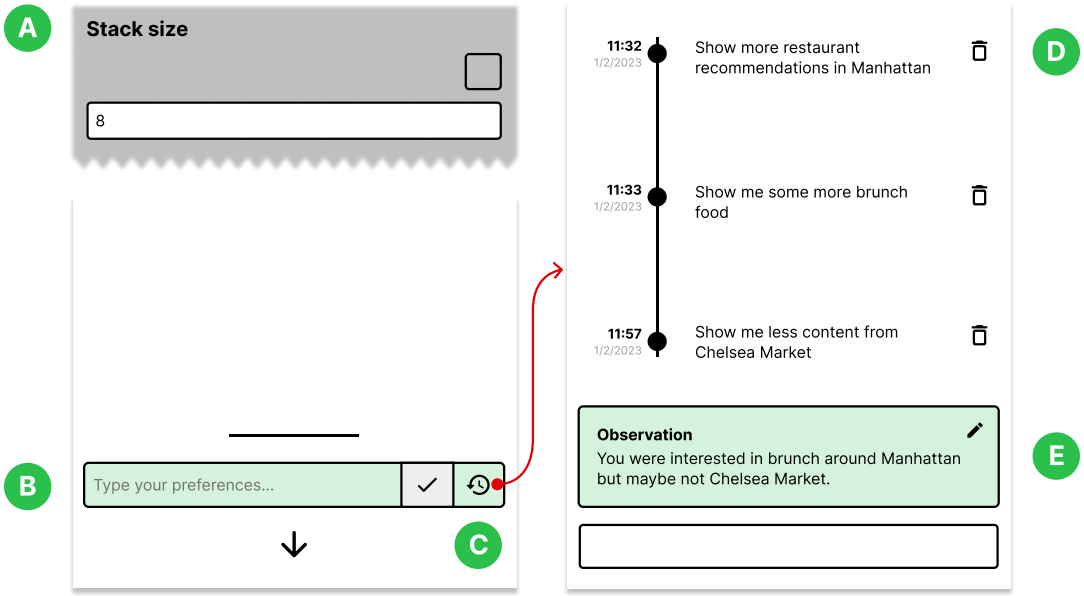

Figure 9: A stack of feed content followed by an opportunity for natural language feedback. (A): the user can set the size of the stack, or select the infinity option for an infinitely scrolling feed. (B): text box for unstructured natural language preference expression. (C): option to view the history of submitted preferences. (D): submitted preferences are cumulative and appear in a timeline. Users can remove individual preferences from the timeline or clear all of them using the option at the bottom of the page. (E): to aid user reflection, a synthesis of their preferences is generated and can be edited to guide future recommendations.

many users. That is, one is never an "individual" on these platforms, but is instead an algorithmically assembled profile of dividuals. Simpson et al. [99] claim that this shifting assemblage results in an "identity spiral" in which users' online identities are always being revised and connected to other identities in uncontrolled and sometimes disagreeable ways. Enhanced personalization, including via teachable feeds, may perpetuate this spiral and dividualization as a whole.

7.1.2 Potential solutions. At the same time, teachable feeds have potential to alleviate some of these concerns. Many concerns around user privacy and targeting stem from the opacity with which algorithms operate—users are unaware of what kind of data is collected about them and where the data is transferred to after collection. By eliciting explicitly taught preferences, users can more clearly grasp what the algorithm has learned about them. From a platform's perspective, the improved granularity of user feedback may reduce the need for implicitly inferring certain preferences. Taken together, the reclamation of user agency may in fact aid platforms in reaching transparency ideals. Furthermore, by accumulating their own set of taught preferences, users can be their own architects of their individualized online identity rather than relying on an algorithm to assemble it from dividuals. While some forms of automated profiling may still persist, the nuanced experiences of the individuated user can now remain uniquely local, not simply dissolved into the dividual crowd.

We also hope that our proposed design patterns can pave the path for new forms of content moderation, such as detecting curriculum folders containing high concentrations of malicious content and natural language feedback that indicates intent to disseminate such content. This will hopefully disincentivize adversarial actors from using teachable feeds to spread harm (e.g., collaboratively plot disinformation campaigns [101]). Similar methods can also be used to identify when users are exposed to increasingly homogeneous perspectives[11] from their content, and when to introduce nudges to increase content diversity. Users can also use these moderation mechanisms to better reflect on their own content browsing habits, treating their role as a teacher *as a learning experience*. Prior work in IMT showed that reflecting upon workplace calendar data throughout the teaching process surfaced new insights from personal data and informed future digital wellbeing goals [48]. We see room to further explore this in a social media context.

7.2 Limitations and Future Work

As with any sensitizing concept, our design patterns are meant to be iterated upon and refined. We showed that these patterns can embody the design principles derived from our study data, but we did not perform empirical user evaluations of our designs. Future work may integrate these designs into existing or new social media platforms and run user studies to gather feedback. Our design patterns were also applied to a microblogging-style platform similar to Mastodon or Twitter. Future work may explore translating them into a TikTok-style platform, as 1) the medium of video may cast additional considerations onto our proposed patterns, and 2) the user can only see one post at any given time, unlike the other

[11]We use "increasingly homogeneous perspectives" in this work but recognize the rich, yet inconclusive, body of empirical work investigating similar phenomena termed "echo chambers" and "filter bubbles" [3, 15, 26, 34, 83, 89].

feeds. Social media aside, other feed-based systems, such as news applications, may also benefit from our design patterns. Further research is needed to determine if or how they should be applied in domains outside of social media.

Teachable feeds also present exciting opportunities for shared, collaborative IMT experiences. For example, a user who has a curriculum folder for celebrities they are interested in seeing more of in their feed may share that folder with a friend with similar celebrity interests. Users living in the same town or city may share curricula with local events or news. More broadly, users who share their established curricula assembled over time can help those with less experience as a teacher in IMT jumpstart the IMT process. Collaborative IMT can thus be of particular interest to future work aiming to expand existing strategies to support novice teachers, strategies that are currently limited to showing push notifications of expert teachers' best practices using Wizard-of-Oz methods [111].

Additionally, the platforms used in our study were more different than they were similar—users opened their feeds for different purposes, followed different accounts, and were exposed to different feed affordances. While this enabled us to create a broader mapping of our design space and witness shifts in taxonomies across the platforms, it did not allow us to isolate any particular variables. The launch of Threads [74] in July 2023 presents a rare opportunity to address this. Because Instagram and Threads accounts are automatically linked, one may follow the same accounts on both platforms but find themselves in two completely distinct user experiences. Future work can use this as a case study for how variations in feed affordances can impact content perception and preferences, thus shaping the "culture" of a platform.

Finally, even though we proposed teachable feeds with users' best interests in mind, we acknowledge that social media platforms may not always act in those interests en route to achieving their financial goals [12]. We do, however, see potential avenues for incentive alignment and adoption of our proposed designs. Currently, the dominant financial model for platforms revolves around maximizing ad revenue by engagement optimizing techniques to prolong users' ad exposure. We encourage future work to investigate how to *increase ad quality* using feedback willingly provided by users in the teaching loop to reduce reliance on *ad quantity* and *ad exposure time*. This way, ad targeting can be made more transparent, as users have a record of what they taught to their feed. Irrelevant ads, which clutter many of today's platforms, may also be reduced upon platforms receiving higher-quality user feedback. Platforms may, and have already started to, explore new financial models. For example, Mastodon has adopted a non-profit financial model, choosing "not [to] implement any monetization strategies in the software" [72] and instead funding ongoing development with grants and crowd-sourced funds [29]. This lack of commercial involvement means that Mastodon can better align with user interests but constrains development resources to build experiences that are polished and compelling enough for mass-market adoption [12]. Public funding may be crucial in the success of these non-profit models. As such, policymakers may see more opportunities to engage with this space to advocate for users while fostering financially sustainable platforms.

8 CONCLUSION

An architect may design a home, but it is ultimately up to the home's occupants to decorate the space to reflect their unique tastes and preferences. Today, social media feeds act as digital architects of personalized social spaces, but users often lack such decorative agency.

In this work, we explored the idea of teachable social media feed experiences for agential, personalized feed curation. To do so, we conducted a study with 24 social media users across Instagram, Mastodon, TikTok, and Twitter to elicit key signals they used to determine the value of posts in their feeds. We found that users evaluated content in multi-faceted and nuanced ways that cannot be fully captured by affordances on current platforms. To enable users to better "teach" an algorithm their preferences, we offered five IMT-inspired five design principles for teachable feed experiences, informed by findings from our study. We then embodied these principles in three feed designs to inspire future efforts on integrating teachable feeds into real-world social media systems.

Altogether, our contributions lay the groundwork for continued exploration of teachable feed experiences. We hope this exploration can empower users and platforms to craft more comforting and expressive digital homes.

REFERENCES

[1] Sverrir Arnórsson, Florian Abeillon, Ibrahim Al-Hazwani, Jürgen Bernard, Hanna Hauptmann, and Mennatallah El-Assady. 2023. Why am I reading this? Explaining Personalized News Recommender Systems. (2023).

[2] Nejla Asimovic, Jonathan Nagler, Richard Bonneau, and Joshua A Tucker. 2021. Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences* 118, 25 (2021), e2022819118.

[3] Brooke E Auxier and Jessica Vitak. 2019. Factors motivating customization and echo chamber creation within digital news environments. *Social media+ society* 5, 2 (2019), 2056305119847506.

[4] Natã M Barbosa, Gang Wang, Blase Ur, and Yang Wang. 2021. Who Am I? A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–32.

[5] Nathan Bartley, Andres Abeliuk, Emilio Ferrara, and Kristina Lerman. 2021. Auditing Algorithmic Bias on Twitter. In *13th ACM Web Science Conference 2021* (Virtual Event, United Kingdom) *(WebSci '21)*. Association for Computing Machinery, New York, NY, USA, 65–73. https://doi.org/10.1145/3447535.3462491

[6] Amanda Baughan, Mingrui Ray Zhang, Raveena Rao, Kai Lukoff, Anastasia Schaadhardt, Lisa D. Butler, and Alexis Hiniker. 2022. "I Don't Even Remember What I Read": How Design Influences Dissociation on Social Media. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 13 pages. https://doi.org/10.1145/3491102.3501899

[7] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa Mekler. 2023. How does HCI Understand Human Autonomy and Agency? *arXiv preprint arXiv:2301.12490* (2023).

[8] Michael S. Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the Invisible Audience in Social Networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/2470654.2470658

[9] Bluesky. 2023. Algorithmic Choice with Custom Feeds. https://blueskyweb.org/blog/7-27-2023-custom-feeds.

[10] Herbert Blumer. 2017. What is wrong with social theory? In *Sociological methods*. Routledge, 84–96.

[11] Alan Borning and Michael Muller. 2012. Next Steps for Value Sensitive Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1125–1134. https://doi.org/10.1145/2207676.2208560

[12] Amy S Bruckman. 2022. *Should you believe Wikipedia?: online communities and the construction of knowledge.* Cambridge University Press.

[13] Laura Burbach, Johannes Nakayama, Nils Plettenberg, Martina Ziefle, and André Calero Valdez. 2018. User Preferences in Recommendation Algorithms:

The Influence of User Diversity, Trust, and Product Category on Privacy Perceptions in Recommender Algorithms. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 306–310. https://doi.org/10.1145/3240323.3240393

[14] Jenna Burrell, Zoe Kahn, Anne Jonas, and Daniel Griffin. 2019. When Users Control the Algorithms: Values Expressed in Practices on Twitter. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 138 (nov 2019), 20 pages. https://doi.org/10.1145/3359240

[15] L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. 2019. Controlling polarization in personalization: An algorithmic framework. In *Proceedings of the conference on fairness, accountability, and transparency*. 160–169.

[16] Minmin Chen, Yuyan Wang, Can Xu, Ya Le, Mohit Sharma, Lee Richardson, Su-Lin Wu, and Ed H. Chi. 2021. Values of User Exploration in Recommender Systems. *Proceedings of the 15th ACM Conference on Recommender Systems* (2021).

[17] John Cheney-Lippold. 2017. We are data. In *We Are Data*. New York University Press.

[18] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep Reinforcement Learning from Human Preferences. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf

[19] Angèle Christin. 2020. The ethnographer and the algorithm: beyond the black box. *Theory and Society* 49, 5-6 (2020), 897–918.

[20] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own Actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. https://doi.org/10.1145/2207676.2208350

[21] Sarah Dean and Jamie Morgenstern. 2022. Preference Dynamics Under Personalized Recommendations. *Proceedings of the 23rd ACM Conference on Economics and Computation* (2022).

[22] Laura DeNardis and Andrea M Hackl. 2015. Internet governance by social media platforms. *Telecommunications Policy* 39, 9 (2015), 761–770.

[23] Michael A. DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173694

[24] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3163–3174. https://doi.org/10.1145/3025453.3025659

[25] Simon Dooms, Toon De Pessemier, and Luc Martens. 2014. Improving IMDb Movie Recommendations with Interactive Settings and Filters. In *RecSys Posters*.

[26] Elizabeth Dubois and Grant Blank. 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, communication & society* 21, 5 (2018), 729–745.

[27] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. https://doi.org/10.1145/3411764.3445188

[28] Michael D. Ekstrand, Daniel Kluver, F. Maxwell Harper, and Joseph A. Konstan. 2015. Letting Users Choose Recommender Algorithms: An Experimental Study. *Proceedings of the 9th ACM Conference on Recommender Systems* (2015).

[29] Elysse Bell. 2023. How Mastodon Makes Money. https://www.investopedia.com/how-mastodon-makes-money-7482865.

[30] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" It, Then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2371–2382. https://doi.org/10.1145/2858036.2858494

[31] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/2702123.2702556

[32] K. J. Kevin Feng and David W. Mcdonald. 2023. Addressing UX Practitioners' Challenges in Designing ML Applications: An Interactive Machine Learning Approach. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 337–352. https://doi.org/10.1145/3581641.3584064

[33] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[34] Richard Fletcher, Craig T Robertson, and Rasmus Kleis Nielsen. 2021. How many people live in politically partisan online news echo chambers in different countries? *Journal of Quantitative Description: Digital Media* 1 (2021).

[35] Batya Friedman. 1996. Value-sensitive design. *interactions* 3, 6 (1996), 16–23.

[36] Luke Friedman, Sameer Ahuja, David Allen, Zhenning Tan, Hakim Sidahmed, Changbo Long, Jun Xie, Gabriel Schubiner, Ajay Patel, Harsh Lara, Brian Chu, Zexiang Chen, and Manoj Tiwari. 2023. Leveraging Large Language Models in Conversational Recommender Systems. *ArXiv* abs/2305.07961 (2023).

[37] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Syst. Appl.* 56 (2016), 9–27.

[38] Alex Heath. 2021. Facebook's lost generation. https://www.theverge.com/22743744/facebook-teen-usage-decline-frances-haugen-leaks.

[39] Silas Hsu, Kristen Vaccaro, Yin Yue, Aimee Rickman, and Karrie Karahalios. 2020. Awareness, Navigation, and Use of Feed Control Settings Online. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376583

[40] Instagram Help Center. 2023. Create a Close Friends list on Instagram. https://help.instagram.com/476003390920140/?cms_platform=android-app&helpref=platform_switcher.

[41] Farnaz Jahanbakhsh, Yannis Katsis, Dakuo Wang, Lucian Popa, and Michael Muller. 2023. Exploring the Use of Personalized AI for Identifying Misinformation on Social Media. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–27.

[42] D. Jannach, Sidra Naveed, and Michael Jugovac. 2016. User Control in Recommender Systems: Overview and Interaction Challenges. In *International Conference on Electronic Commerce and Web Technologies*.

[43] Jeff Widman. 2015. EdgeRank. http://edgerank.net/.

[44] Shagun Jhaver, Alice Qian Zhang, Quanze Chen, Nikhila Natarajan, Ruotong Wang, and Amy Zhang. 2023. Personalizing Content Moderation on Social Media: User Perspectives on Moderation Choices, Interface Design, and Labor. *Proceedings of the ACM on Human-Computer Interaction* (2023).

[45] Chenyan Jia, Alexander Boltz, Angie Zhang, Anqing Chen, and Min Kyung Lee. 2022. Understanding Effects of Algorithmic vs. Community Label on Perceived Accuracy of Hyper-Partisan Misinformation. 6, CSCW2, Article 371 (nov 2022), 27 pages. https://doi.org/10.1145/3555096

[46] Yucheng Jin, Bruno Cardoso, and Katrien Verbert. 2017. How do different levels of user control affect cognitive load and acceptance of recommendations?. In *Jin, Y., Cardoso, B. and Verbert, K., 2017, August. How do different levels of user control affect cognitive load and acceptance of recommendations?. In Proceedings of the 4th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2017)*, Vol. 1884. CEUR Workshop Proceedings, 35–42.

[47] Jennifer Jolly. 2022. Considering joining the Twitter migration? Check out these platform alternatives. https://www.usatoday.com/story/tech/2022/11/06/twitter-alternatives-mastodon-bluesky-countersocial-more/8287038001/.

[48] Matthew Jörke, Yasaman S. Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. 2023. Pearl: A Technology Probe for Machine-Assisted Reflection on Personal Data. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 902–918. https://doi.org/10.1145/3581641.3584054

[49] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4 (1996), 237–285.

[50] Hyeonsu B Kang, Rafal Kocielnik, Andrew Head, Jiangjiang Yang, Matt Latzke, Aniket Kittur, Daniel S. Weld, Doug Downey, and Jonathan Bragg. 2022. From Who You Know to What You Read: Augmenting Scientific Recommendations with Implicit Social Networks. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022).

[51] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 305 (oct 2021), 44 pages. https://doi.org/10.1145/3476046

[52] Kay Kender and Christopher Frauenberger. 2022. The Shape of Social Media: Towards Addressing (Aesthetic) Design Power. *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (2022). https://api.semanticscholar.org/CorpusID:249579166

[53] M Laeeq Khan. 2017. Social media engagement: What motivates user participation and consumption on YouTube? *Computers in human behavior* 66 (2017), 236–247.

[54] Hankyung Kim and Youn kyung Lim. 2021. Teaching-Learning Interaction: A New Concept for Interaction Design to Support Reflective User Agency in Intelligent Systems. *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (2021). https://api.semanticscholar.org/CorpusID:235662991

[55] Hankyung Kim and Youn-Kyung Lim. 2023. Investigating How Users Design Everyday Intelligent Systems in Use. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (2023). https://api.semanticscholar.org/CorpusID:259376338

[56] Jan Kirchner and Christian Reuter. 2020. Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 140 (oct 2020), 27 pages. https://doi.org/10.1145/3415211

[57] Bart P. Knijnenburg, Niels J. M. Reijmer, and Martijn C. Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *ACM Conference on Recommender Systems*.

[58] Pascal D. König. 2022. Challenges in enabling user control over algorithm-based services. *AI & SOCIETY* (2022).

[59] Ziyi Kou, Lanyu Shang, Yang Zhang, and Dong Wang. 2022. HC-COVID: A Hierarchical Crowdsource Knowledge Graph Approach to Explainable COVID-19 Misinformation Detection. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 36 (jan 2022), 25 pages. https://doi.org/10.1145/3492855

[60] Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q. Vera Liao, Yunfeng Zhang, and Chenhao Tan. 2022. Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 54, 18 pages. https://doi.org/10.1145/3491102.3501999

[61] Anja Lambrecht and Catherine Tucker. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management science* 65, 7 (2019), 2966–2981.

[62] Dennis Lawo, Thomas Neifer, Margarita Esau, and Gunnar Stevens. 2021. Buying the 'Right' Thing: Designing Food Recommender Systems with Critical Consumers. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021).

[63] Angela Y. Lee, Hannah Mieczkowski, Nicole B. Ellison, and Jeffrey T. Hancock. 2022. The Algorithmic Crystal: Conceptualizing the Self through Algorithmic Personalization on TikTok. 6, CSCW2, Article 543 (nov 2022), 22 pages. https://doi.org/10.1145/3555601

[64] Liu Leqi, Giulio Zhou, Fatma Kilincc-Karzan, Zachary Chase Lipton, and A. Montgomery. 2023. A Field Test of Bandit Algorithms for Recommendations: Understanding the Validity of Assumptions on Human Preferences in Multi-armed Bandits. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023).

[65] Yu Liang, Aditya Ponnada, Paul Lamere, and Nediyana Daskalova. 2023. Enabling Goal-Focused Exploration of Podcasts in Interactive Recommender Systems. *Proceedings of the 28th International Conference on Intelligent User Interfaces* (2023).

[66] Zachary C Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[67] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[68] Katerina Lup, Leora Trub, and Lisa Rosenthal. 2015. Instagram# instasad?: exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking* 18, 5 (2015), 247–252.

[69] Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic Authority: The Ethics, Politics, and Economics of Algorithms That Interpret, Decide, and Manage. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1057–1062. https://doi.org/10.1145/2851581.2886426

[70] Michael Madary. 2022. The illusion of agency in human–computer interaction. *Neuroethics* 15, 1 (2022), 16.

[71] Aisha Malik. 2023. TikTok adds dedicated video feeds for sports, fashion, gaming and food. https://techcrunch.com/2023/02/16/tiktok-adds-dedicated-video-feeds-for-sports-fashion-gaming-and-food/.

[72] Mastodon. 2023. What is Mastodon? https://docs.joinmastodon.org.

[73] Christine Mendez, Vlatko Lukarov, Christoph Greven, André Calero Valdez, Felix Dietze, Ulrik Schroeder, and Martina Ziefle. 2017. User Groups and Different Levels of Control in Recommender Systems. In *Interacción*.

[74] Meta. 2023. Introducing Threads: A New Way to Share With Text. https://about.fb.com/news/2023/07/introducing-threads-new-app-text-sharing/.

[75] Miro. 2023. Miro: An Online Whiteboard & Visual Collaboration Platform. https://www.miro.com/.

[76] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. 2023. Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review* 56, 4 (2023), 3005–3054.

[77] Sheshera Mysore, Mahmood Jasim, Andrew McCallum, and Hamed Zamani. 2023. Editable User Profiles for Controllable Text Recommendation. *arXiv preprint arXiv:2304.04250* (2023).

[78] Arvind Narayanan. 2023. Understanding Social Media Recommendation Algorithms. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

[79] Felicia Ng, Jina Suh, and Gonzalo Ramos. 2020. Understanding and Supporting Knowledge Decomposition for Machine Teaching. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1183–1194. https://doi.org/10.1145/3357236.3395454

[80] Jeroen Ooge, L Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. *Proceedings of the 28th International Conference on Intelligent User Interfaces* (2023).

[81] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 27730–27744. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf

[82] Aviv Ovadya and Luke Thorburn. 2023. Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance. https://doi.org/10.48550/ARXIV.2301.09976

[83] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

[84] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[85] Nikolay Pavlov. 2014. User interface for people with autism spectrum disorders. *Journal of Software Engineering and Applications* 2014 (2014).

[86] Diana Petrescu, Diego Antognini, and Boi Faltings. 2021. Multi-Step Critiquing User Interface for Recommender Systems. *Proceedings of the 15th ACM Conference on Recommender Systems* (2021).

[87] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and Measuring Model Interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 237, 52 pages. https://doi.org/10.1145/3411764.3445315

[88] Aditya Kumar Purohit, Kristoffer Bergram, Louis Barclay, Valéry Bezençon, and Adrian Holzer. 2023. Starving the Newsfeed for Social Media Detox: Effects of Strict and Self-Regulated Facebook Newsfeed Diets. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 196, 16 pages. https://doi.org/10.1145/3544548.3581187

[89] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. 2016. Echo chambers on Facebook. *Available at SSRN 2795110* (2016).

[90] Gonzalo Ramos, Christopher Meek, Patrice Simard, Jina Suh, and Soroush Ghorashi. 2020. Interactive machine teaching: a human-centered approach to building machine-learned models. *Human–Computer Interaction* 35, 5-6 (2020), 413–451.

[91] Urbano Reviglio and Claudio Agosti. 2020. Thinking outside the black-box: The case for "algorithmic sovereignty" in social media. *Social Media+ Society* 6, 2 (2020), 2056305120915613.

[92] Leonid Rozenblit and Frank Keil. 2002. The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science* 26, 5 (2002), 521–562.

[93] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics*. 1668–1678.

[94] Martin Saveski, Brandon Roy, and Deb Roy. 2021. The structure of toxic conversations on Twitter. In *Proceedings of the Web Conference 2021*. 1086–1097.

[95] Ruoxi Shang, K. J. Kevin Feng, and Chirag Shah. 2022. Why Am I Not Seeing It? Understanding Users' Needs for Counterfactual Explanations in Everyday Recommendations. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1330–1340.

https://doi.org/10.1145/3531146.3533189
[96] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *interactions* 4, 6 (1997), 42–61.
[97] Amanda Silberling. 2022. A beginner's guide to Mastodon, the open source Twitter alternative. https://techcrunch.com/2022/11/08/what-is-mastodon/.
[98] Ignacio Siles, Andrés Segura-Castillo, Ricardo Solís, and Mónica Sancho. 2020. Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society* 7, 1 (2020), 2053951720923377.
[99] Ellen Simpson, Andrew Hamann, and Bryan Semaan. 2022. How to Tame "Your" Algorithm: LGBTQ+ Users' Domestication of TikTok. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 22 (jan 2022), 27 pages. https://doi.org/10.1145/3492841
[100] Spandana Singh. 2020. Charting a Path Forward. https://www.newamerica.org/oti/reports/charting-path-forward/.
[101] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
[102] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems* 33 (2020), 3008–3021.
[103] Jonathan Stray. 2020. Aligning AI Optimization to Community Well-Being. *International Journal of Community Well-Being* 3 (2020), 443 – 463.
[104] Jonathan Stray, Alon Y. Halevy, Parisa Assar, Dylan Hadfield-Menell, Craig Boutilier, Amar Ashar, Lex Beattie, Michael D. Ekstrand, Claire Leibowicz, Connie Moon Sehat, Sara Johansen, Lianne Kerlin, David Vickrey, Spandana Singh, Sanne Vrijenhoek, Amy X. Zhang, McKane Andrus, Natali Helberger, Polina Proutskova, Tanushree Mitra, and Nina Vasan. 2022. Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ArXiv* abs/2207.10192 (2022).
[105] Jonathan Stray, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell. 2021. What are you optimizing for? Aligning Recommender Systems with Human Values. *ArXiv* abs/2107.10939 (2021).
[106] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
[107] Stefano Teso, Öznur Alkan, Wolfgang Stammer, and Elizabeth Daly. 2023. Leveraging explanations in interactive machine learning: An overview. *Frontiers in Artificial Intelligence* 6 (2023), 1066049.
[108] Tristan Harris. 2016. How Technology is Hijacking Your Mind — from a Magician and Google Design Ethicist. https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3.
[109] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *proceedings of the eighth symposium on usable privacy and security*. 1–15.
[110] Kristen Vaccaro, Dylan Huang, Motahhare Eslami, Christian Sandvig, Kevin Hamilton, and Karrie Karahalios. 2018. The Illusion of Control: Placebo Effects of Control Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173590
[111] Emily Wall, Soroush Ghorashi, and Gonzalo Ramos. 2019. Using expert patterns in assisted interactive machine learning: A study in machine teaching. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part III 17*. Springer, 578–599.
[112] Wall Street Journal. 2021. Investigation: How TikTok's Algorithm Figures Out Your Deepest Desires. https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612E3796.
[113] Jeffrey Warshaw, Tara Matthews, Steve Whittaker, Chris Kau, Mateo Bengualid, and Barton A Smith. 2015. Can an Algorithm Know the" Real You"? Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 797–806.
[114] Xing Yi, Liangjie Hong, Erheng Zhong, Nathan Nan Liu, and Suju Rajan. 2014. Beyond clicks: dwell time for personalization. In *ACM Conference on Recommender Systems*.
[115] Mingrui Ray Zhang, Kai Lukoff, Raveena Rao, Amanda Baughan, and Alexis Hiniker. 2022. Monitoring Screen Time or Redesigning It? Two Approaches to Supporting Intentional Social Media Use *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 60, 19 pages. https://doi.org/10.1145/3491102.3517722
[116] Xiaoxue Zhao, Weinan Zhang, and Jun Wang. 2013. Interactive collaborative filtering. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 1411–1420.
[117] Zhe Zhao, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, and Ed Chi. 2019. Recommending What Video to Watch next: A Multitask Ranking System. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 43–51. https://doi.org/10.1145/3298689.3346997
[118] Michelle X Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 2-3 (2019), 1–36.
[119] Zhongyi Zhou and Koji Yatani. 2022. Gesture-Aware Interactive Machine Teaching with In-Situ Object Annotations. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 27, 14 pages. https://doi.org/10.1145/3526113.3545648
[120] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. 2010. An analysis and critique of Research through Design: towards a formalization of a research approach. In *proceedings of the 8th ACM conference on designing interactive systems*. 310–319.