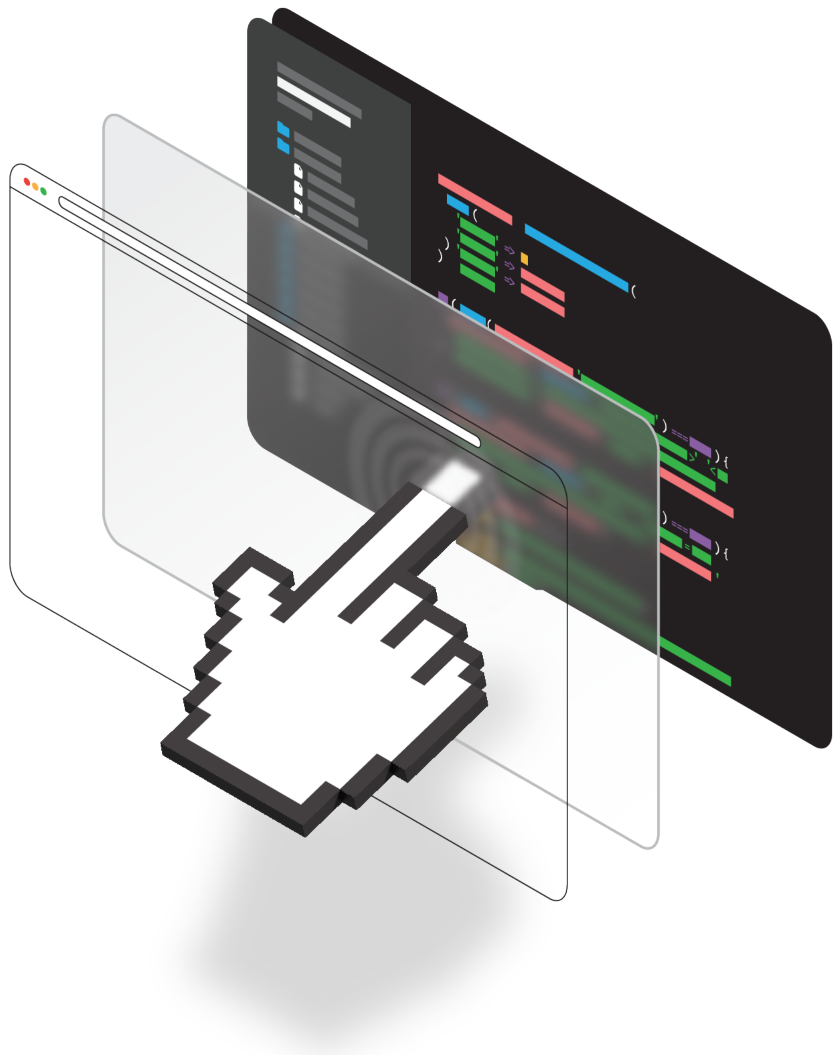
A report from

cdt | Research



"This is Transparency to Me"

User Insights into Recommendation Algorithm Reporting

Michal Luria

October 2022

CENTER FOR
DEMOCRACY
& TECHNOLOGY

The Center for Democracy & Technology (CDT) is a 27-year-old 501(c)3 nonpartisan nonprofit organization that fights to put democracy and human rights at the center of the digital revolution. It works to promote democratic values by shaping technology policy and architecture, with a focus on equity and justice. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

MICHAL LURIA

Research Fellow at the Center for Democracy & Technology.

A report from

cdt | Research

“This is Transparency to Me”

User Insights into Recommendation Algorithm Reporting

Michal Luria

WITH CONTRIBUTIONS BY

Dhanaraj Thakur, Caroline Sinders, Gabriel Nicholas, Emma Llansó, Caitlin Vogus, Aliya Bhatia, Michael Yang, and Samir Jain. Ari Goldberg copy-edited the draft of the report, and Tim Hoagland helped coordinate the layout and design led by Objectively P.B.C.

ACKNOWLEDGEMENTS

We would like to thank Caroline Sinders who worked with us to generate prototypes for Study 2, as well as Schessa Garbutt for their contribution to the graphical designs in the study. We would also like to thank Michael Madaio and Bertram Lee for their feedback on an earlier draft of this report. Lastly, we would like to acknowledge and thank the participants in this study for their time and insights into the topic. All views in this report are those of CDT.

This work is made possible through a grant from the John S. and James L. Knight Foundation.

Suggested Citation: Luria, M. (2022). “This is Transparency to Me”: User Insights into Recommendation Algorithm Reporting. Center for Democracy & Technology. https://cdt.org/insights/this-is-transparency-to-me-user-insights-into-recommendation-algorithm-reporting/

References in this report include original links as well as links archived and shortened by the Perma.cc service. The Perma.cc links also contain information on the date of retrieval and archive.

Contents

Executive Summary

Recommendation algorithms by and large determine what people see on social media. Users know little about how these algorithms work or what information they use to make their recommendations. But what exactly should platforms share with users about recommendation algorithms that would be meaningful to them? Prior research efforts have looked into frameworks for explainability of algorithms as well as design features across social media platforms that can contribute to their transparency and accountability. We build on these prior efforts to explore what a recommendation algorithm transparency report may include and how it could present information to users.

Transparency reports have been on the rise among technology and social media companies, with 88 companies publishing such reports as of July 2021 ("Transparency Reporting Index - Access Now's Global Database," n.d.). These reports are the result of much advocacy from civil society and activist groups and primarily include information about content moderation practices and government requests for data (Vogus & Llansó, 2021). They rarely include information about service providers' content recommendation algorithms, although these algorithms deeply impact the experiences of people who use platforms, as well as advertisers, public figures, and businesses.

Transparency reports are also usually generated periodically, with general information about the company and their practices. As the goal is to suggest a way for people to better understand how recommendation algorithms impact their personal experience, we suggest a more engaging data-driven, interactive, and personalized approach to recommendation algorithm reports. While some information about recommendation algorithms and their use of data can be found in companies' Privacy Policies and Settings pages, we argue that platforms should publish a stand alone report in which everyday users can learn about how personalized content and recommendation algorithms work and affect their online experiences.

In order to understand how to do so, and what such a report might include to best support everyday users' needs, we conducted this two-part, human-centered co-design research project. Co-design is a method that involves end-users in multiple and early stages of the design process and builds on their insights to create tools that would be most meaningful for them. We conducted two sets of individual sessions with a diverse group of casual social media users (n=30) to understand what information they would like companies to share about recommendation algorithms and how that information should be shared.

In Study 1, participants were invited to participate in several design activities aimed at creating a reflective process about their needs and desires. The goal was to form a foundation of what everyday users are interested in and care about—the outcome of Study 1 was a set of guidelines for a recommendation algorithm transparency report as well as insights about features that can be incorporated in prototypes of future reports.

In the preparation for Study 2, we created sketches of screens (“prototypes”) of what a recommendation algorithm report could look like based on findings from Study 1. These prototypes were created as provocations—primarily intended to generate a second conversation with participants about their needs and values (as opposed to suggesting that this is how a report should look). In Study 2, the same participants were invited to examine the manifestations of their and other users’ own ideas, and to reflect on the strengths and drawbacks these prototypes suggest.

Based on the interviews from both studies, we develop guidance about recommendation algorithm reports: what they should include, what aspects they should emphasize, and how they should be communicated.

Research Findings

We present findings in two parts: (1) guidelines about what information should be included in recommendation algorithm transparency reports, and how it should be presented; and (2) initial suggestions, in the form of prototypes, about more engaging and interactive ways of presenting such information to users, with an evaluation of their strengths and weaknesses.

Participants primarily wanted a recommendation algorithm transparency report to include:

- Information about what they do see, as opposed to what is being filtered out;
- What data is collected and inferred about them to be used in recommendation algorithms, and whether and how that data is shared with external partners;
- What data is obtained from other sources to be used in recommendation algorithms and from whom; and
- Whether and how they can (or cannot) make changes to an algorithm and the data it uses.

They wanted the information presented to them to be:

- **Specific**—to clearly explain the choices made by platforms when creating recommendation algorithms and how it would impact them, and to avoid general phrasing such as "to improve user experiences";
- **Direct**—to include data-driven, to-the-point information that does not attempt to frame platform recommendation systems or data collection practices in an overly positive light; and
- **Demonstrated**—to include specific examples of how a given recommendation was made (e.g., "you are seeing this because you follow the cosmetic brand x"), attached to a more general description of how the recommendation algorithm works.

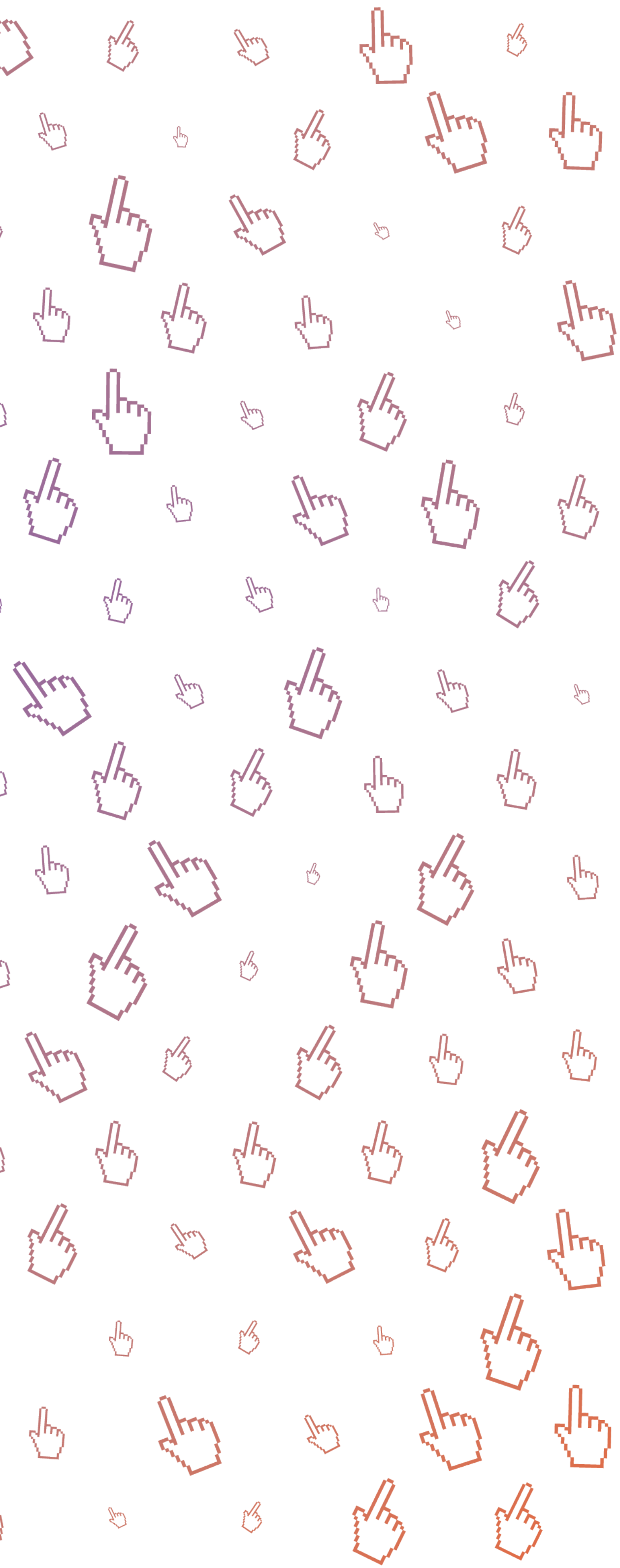
Study 2 introduced several ways to present recommendation algorithm transparency reports that may differ from what one might initially expect. We found that participants were particularly positive about transparency reports that were:

- **Visual**—Information that was presented graphically allowed participants to gain and process lots of information about the topic by quickly skimming the report;
- **Interactive**—Participants were excited about interactive designs that allowed them to "play around" with aspects of the algorithm to better understand its impacts;
- **Personalized**—While not necessary in every report, participants viewed the personalized nature of prototypes as more engaging and more inviting than general information about how a recommendation algorithm works; and
- **Controllable**—Once participants learned about how a recommendation algorithm worked, they appreciated when they were also able to exert control over it, especially around what personal data it incorporated.

In summary, this work provides the perspectives of everyday social media users and identifies their needs and values that future recommendation algorithm transparency reports should support. There is no single right way to provide an algorithmic transparency report, and, therefore, we are not offering a template or recommendation for how every social media platform should provide recommendation algorithm transparency.

Rather, these perspectives are critical for social media platforms to consider, as users have a right to choose what content they consume online. Implementing meaningful forms of recommendation algorithm transparency can increase people's trust in platforms and contribute to a safer ecosystem of transparent and accountable social media platforms.

Introduction



When people talk about algorithms on social media, they are often referring to recommendation systems and algorithms. Recommendation algorithms are automated systems used by social media platforms to recommend content to users, such as by displaying content in a feed or inviting users to view additional content. These algorithms have a significant impact on people's experiences, on- and offline. They can help expose people to new people and ideas (Chen et al., 2009), enable individuals to learn about topics they care about, keep up with news and trends (Boczkowski et al., 2018) and even draw attention towards social movements or change (Poell et al., 2015).

Still, they can bring challenges and concerns regarding how people are exposed to different kinds of content—content can be limited (Bechmann & Nielbo, 2018), negatively shape individuals' body images (Elsesser, 2021), and impact critical offline behaviors such as voting (Hsu, 2018). The complex nature of recommendation algorithms and their impact begs for more transparency and access to information (Vogus & Llansó, 2021) about how these algorithms work.

Recommendation algorithms are also significantly informed by data collected on social media platforms, such as personal data and online behaviors, as well as information obtained from third parties, such as data purchased from brokers (Leetaru, 2018; Shenkman et al., 2021). These practices make it even more critical for platforms to share information about recommendation algorithms and how they work with the broader public. Yet social media companies are typically not very transparent about these issues, leaving people in the dark when it comes to critical questions like: What factors influence content recommendation algorithms' decision-making? How are the algorithms trained? What errors might they introduce? What personal data is being collected to inform them?

Companies do make some information about recommendation algorithms and their use of data available, primarily in their Privacy Policies. To some extent, this may be due to requirements from privacy laws, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Other valuable information can be found on various pages on social media platforms (for example, Meta's effort to share information about their practices of demoting harmful content (Types of Content We Demote | Transparency Center, n.d.) or Twitter's suggestion of accounts to follow (About Twitter's

Account Suggestions, n.d.)). However, much of the information that does exist is not always easy to find and understand, especially the information presented in Privacy Policies. In many ways, current information sharing practices are likely to discourage users from attempting to better understand how platforms' recommendation algorithms work.

Social media platforms have expressed several reasons for not fully disclosing explanations about their recommendation algorithms. One is the fear that being transparent about recommendation algorithms may encourage scammers, spammers, and trolls to "play" the algorithm for higher reach (Newton, 2017); this is a familiar challenge for a variety of online services, such as the field of search engine optimization (Patil Swati et al., 2013). Another reason is the competitive advantage of keeping the mechanisms of an algorithm secret, as well as the necessity of protecting the company's trade secrets (Davis & Aggarwal, 2020). Further, social media companies themselves have sometimes admitted they do not fully understand the potential biases in their algorithms, or in some cases even how they work. For example, Twitter had previously reached out to the broader research and hacker communities for help in identifying algorithmic harms (Chowdhury & Williams, 2021). While there is likely merit to these concerns, there is also a wide gulf between the current lack of information and potentially harmful disclosures.

But even if platforms are interested in sharing more about their algorithms, it is not always easy. Explaining how algorithms work is objectively challenging, and for many years academic researchers and civil society activists have sought to answer the question of how best to provide algorithmic transparency and explainability (Lipton, 2017; Szymielewicz et al., 2020; Yang, 2021). Among these efforts are the creation of structured frameworks for what transparency and explainability can and should include (Diakopoulos, 2016; Sokol & Flach, 2020), a survey of current recommendation algorithm explanations (Tintarev & Masthoff, 2007) and a range of suggested data and machine learning model documentation approaches that would provide transparency (such as "datasheets" (Gebru et al., 2021), "model cards" (Mitchell et al., 2019) and "fact sheets" (Richards et al., 2020)). Recent efforts have taken a more human-centered approach to system and algorithms explainability and transparency, such as "Social Transparency" (Ehsan et al., 2021) and Meta's "System Cards" (Green et al., 2021).

One way in which platforms currently share information with the public is through transparency reports—following much advocacy from digital rights groups around the world (Llanso, 2021; Llansó & Morgan, 2014; Singh & Bankston, 2018; Woolery et al., 2016), it has now become common practice for technology companies to produce semi-regular reports about a range of topics. Primarily, they include information about platforms' responses to government demands for user data and content removal and, more recently, about their content moderation practices (Vogus & Llansó, 2021). As of July 2021, 88 tech companies have published such reports publicly ("Transparency Reporting Index - Access Now's Global Database," n.d.). This is important progress that can help the public and lawmakers to understand how platforms operate and to hold them accountable when needed. Yet transparency reports include little to no information about how algorithms are used to recommend content to users, despite their importance and the consistent interest expressed by the media (Stern, 2022; Wall Street Journal, 2021), the public (Hughes, 2022), and governments around the world (Allan, 2022).

In this report, we take a design research approach to provide insight into the aspects of social media content recommendation algorithms that platforms can and should share with their users. Through a methodology of co-design—interviews and ideation activities with diverse users of social media platforms—we identify users' needs in this space, and provide a concrete set of examples of what user-facing algorithmic transparency about social media recommendation algorithms may look like.

The results of this research are reported in two parts: (1) We share guidelines for how to make current information-sharing practices more effective and desired, based on social media users' perspectives; and (2) we present prototypes of interactive ways platforms may share information with people about how algorithms impact them.

Our human-centered design research approach, in which everyday users of social media platforms are co-designers of future "algorithmic transparency reports," provides insights into what *everyday users* may need and desire. Recommendation algorithm transparency reports have many potential audiences, including advocates, researchers, journalists, regulators, and more, and future work is needed to understand how reports could best serve all of their needs. However, because these algorithms most directly affect users of platforms, understanding their perspectives is critical to ensuring that transparency reports are worthwhile and effective, and to provide a sound basis for further public discussion and policymaking around the role of social media recommendation algorithm reports.

Research Approach and Methodology

Using human-centered design research is ideal in a situation of unknown unknowns—it is useful for framing a set of problems that need to be addressed and can reveal hidden and surprising aspects that have not yet been discussed. It also allows for the broader community to participate in designing technology that is intended for them. In this study, we made use of several design research methods: Co-Design, a process of collaborating with people who may use a technology to design it (Steen, 2013; Valencia et al., 2021); Card Sorting, a method that makes use of cards to form a tangible ideation and reflection process (Golembewski & Selby, 2010; Wood & Wood, 2008); and Experience Prototyping, which allows participants to examine an interaction through active engagement with prototypes, say, a mock-up of a screen that they might see online on a social media platform (Buchenau & Suri, 2000; Luria & Candy, 2022).

Design research does not strive to be scientific in the sense that it does not attempt to confirm or refute hypotheses, nor to generate theories about the current state of things. Rather, it sets out to identify what is worthwhile to design for, what people care about, and how technology can be made better. Using design methodology, research can reveal hidden and surprising aspects that have been overlooked and identify some (but not all) tangible opportunities to design technology in a way that would be meaningful for people. The design research approach in this work sets out to understand the current state of algorithmic transparency perceptions and to suggest ways to shift that current state into a preferred state by considering alternative paths to interacting with algorithmic transparency reports.

Participants in the study took part in two sessions which were based on remote one-on-one interviews with the lead researcher. Procedures for both Study 1 and Study 2 were approved by an external Institutional Review Board (IRB), Advarra. The interviews were semi-structured, i.e, the list of questions for participants was pre-defined, but also open to unstructured discussions and follow up questions based on participants' answers. Each session was between 45-75 minutes and included several predefined activities which were executed using a "digital whiteboard" called Mural.

In the first session, participants were asked to organize a list of cards with topics that social media platforms could share about their recommendation algorithms, on a scale between "most important for me to know about" and "least important for me to know about" (see Figure 1).

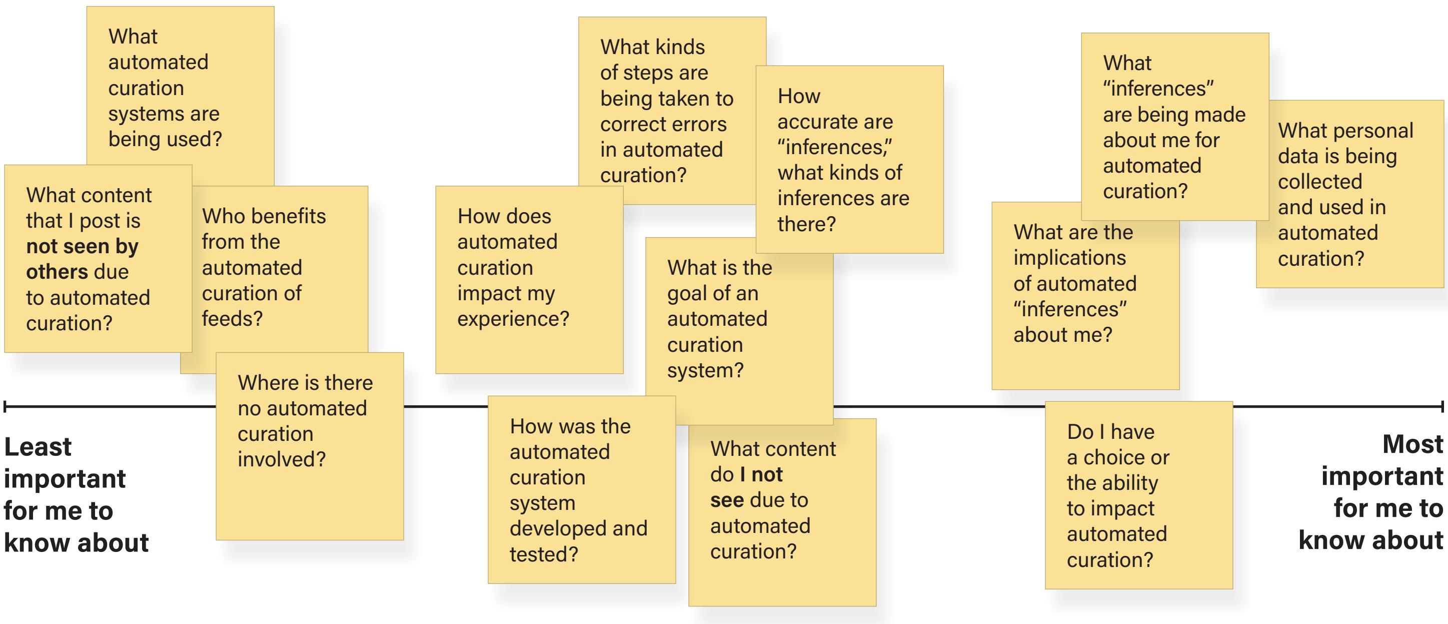


Figure 1. Aspects of Automated Curation on Social Media. The above diagram illustrates how a participant organized recommendation algorithm-related topics on the scale between "Least important for me to know about" and "Most important for me to know about."

Source - CDT

As participants were organizing the cards, they were asked to "think out loud" and reflect on their choices. In the same session, they were also asked to graphically sketch what a transparency report about social media recommendation algorithms may look like (using shapes and text boxes on the digital whiteboard), as well as to read content that social media platforms currently share about their algorithms and discuss what works in these descriptions and what information is missing and desired.

From these activities, the lead researcher, in collaboration with a second design researcher, extracted "weak signals." "Weak signals" is a term used across business, futures, and design communities to indicate an early sign of interest that currently has little or no impact, but has the potential to have significant impact or change in the future (Hiltunen, 2006). The "weak signals" that were identified in the first study were then used to create prototypes of novel interfaces through which platforms could share information about their recommendation algorithms (read more about this process in the Appendix: Methodology).

These prototypes were used as discussion probes in the second co-design session with social media users, where participants were again invited to remote one-on-one interviews. They were presented with prototypes that manifested some of the discussions and ideas from Study 1, and were asked to read through these simulated interfaces and reflect on their strengths and weaknesses.

Thirty (30) participants took part in Study 1, and of those, 16 participated in the follow-up Study 2. We invited participants that used a range of social media platforms, including large platforms such as Facebook and Twitter, but also users of Reddit, Discord, TikTok, Pinterest, Linkedin and others. We made sure to include a diverse sample of social media platform users as participants: 15 participants identified as female, 14 identified as male, and one as non-binary. A total of 7 participants (about 23%) were people of color. Seven (7) participants reported being politically right-leaning, and 20 were left-leaning. Three (3) participants, a total of 10%, reported having a disability. Participants were frequent computer users (in part due to their recruitment via an online research platform), and varied in their self-reported technical orientation (M = 5.6 of 7) and privacy orientation (M = 4.4 of 7). While a few of our participants were business owners who made use of social media for marketing purposes, we asked them to focus on their experience as consumers, as the need for algorithmic transparency for business owners and “influencers” is outside of scope for this report.

More details about our approach, methodology and detailed procedure can be found in the Appendix: Methodology.

Findings

We present findings in two parts: (1) guidelines for content, phrasing, formatting and style for recommendation algorithm transparency reports; these could be applied to both straightforward reports (for example, similar to current Privacy Policy pages) or to more interactive formats of presenting information; and (2) initial suggestions, in the form of prototypes, for ways people could interact with information about social media recommendation algorithms.

We present prototype suggestions, along with their strengths and weaknesses based on user responses to them. The prototypes we present here are not intended to provide templates for what platforms should do, nor do they attempt to present all design possibilities. Instead, we take a step towards understanding the value of such reports, and provide initial directions for representation of information that could support users' needs. We encourage companies to implement similar co-design methods to identify the right approaches and designs for their platforms and user bases.

Study 1 Findings: Content and Language for Algorithmic Transparency Reports

Overall, most participants felt positively about the idea of social media recommendation algorithms (hereafter also referred to simply as "algorithms") and understood that their use is necessary to create a personalized experience on social media platforms. As P24[1] put it, algorithms are "*going to be involved no matter what...otherwise how are [social media platforms] going to be able to show me content that is relevant to me?*" That said, participants were intrigued by the possibility of confronting the assumption that content should always be recommended algorithmically, and to explore what an "algorithm-free" space might look like, or a space where they hold significantly more control over the content they engage with, as we will further discuss in our findings.

1 Each participant is identified via a consistent number that represents their chronological participation in Study 1. Participants who participated in both Study 1 and Study 2 are identified via the same number for both studies.

"[Algorithms are] going to be involved no matter what…otherwise how are [social media platforms] going to be able to show me content that is relevant to me?"

While algorithms themselves were met with fairly positive attitudes, current reporting about how algorithms work did not receive the same enthusiasm. When we asked what people think a transparency report that explains recommendation algorithms will look like, participants had very low expectations: "*something that's not really saying anything*" (P10), "generic" (P15), "vague" (P4), or as P7 called it, "*gobbledygook industry language.*" Overall, they had a hard time conceptualizing a transparency report as something that contains valuable information for them.

Nevertheless, throughout the study, participants were interested in learning more about many of the topics related to recommendation algorithms. For the rest of this section, we highlight what these topics were, what should be excluded, and how they should be communicated.

WHAT INFORMATION SHOULD ALGORITHMIC TRANSPARENCY REPORTS INCLUDE?

Participants acknowledged that algorithmic decision making on a social media platform might cause content to be *hidden* from them (if, for example, they rarely interact with an individual, if some content contains opinions very different to theirs, or if some content is shadowbanned (Nicholas, 2022)). Yet it was not a significant concern for participants. Rather, they wanted to know more about what it is that they do see, and why. They also wanted more agency and control over content that they interact with on a daily basis.

Users pointed to three topics that, in their views, should be covered in algorithmic transparency reports: (1) how personal data is collected and used in recommendation algorithms; (2) what control users have over the data and the algorithm; and (3) what data is being collected and what data is being shared outside the platform.

1. How Personal Data is Collected and Used in Recommendation Algorithms

Of all the various subjects that might be included in an algorithmic transparency report, participants cared most about how their personal data was being collected and used to serve them content.

While participants described data collection as an expected, acceptable and even desired practice by social media platforms, their current experience is that "*[social media platforms] are trying to find out as much as they can about you*" (P1) and that they are "*gathering and sharing more information than they should*" (P27). Participants expressed a desire to have a report that would state the reasons for data collection and the defined boundaries of what data will and will not be collected.

"[I don't see] how that would be helpful based on what the [platform] does as a service…it just doesn't make sense."

Participants requested more reasons for collecting data primarily because, for many types of collected data, it was difficult for participants to see *"how that would be helpful based on what the [platform] does as a service...it just doesn't make sense"* (P25). This was especially true for inferences a platform makes about a user, in contrast to data that users explicitly give to platforms: "*If I just solicit it without reasoning [give the platform something voluntarily], then that's different, I gave them authorization and I'm ok with it. But if they are [making an inference] on their own, I want to know about it and I want to know why*" (P22).

2. What Control Users Have Over the Data and the Algorithm

Most participants wanted to know if they "*have any input or say on how [the algorithm] works, or if it is all just up to the company itself?*" (P15). They agreed that they should "*be able to have a choice of what to see and what to not see*" (P24), and, in particular, they asked for "*[platforms] to say whether or not [the user] can turn [something] off*" (P15). It seems that clear communication about the extent of user control would be an inherent part of an ideal algorithmic transparency report. In other words, it is not enough to share what is being done as part of an algorithmic experience; rather, platforms should also explain users' options for customization within a particular service and the impact of each choice.

During the co-design sessions, participants were presented with statements from platforms' Privacy Policies on a range of topics, including the topic of user control. This was a way to learn about what current information does or does not resonate with individuals. They mostly agreed the current wording about user control is not straightforward. Some suggested that platforms are intentionally trying to discourage their involvement—P23 explained what seemed to them to be the platforms' thought process: "*they are telling you [what you can control] because they have to [in Privacy Policies]...it's like: 'well if we don't explain it as much, they probably won't do it [change things], so we're just going to kind of gloss over it.'*"

Instead, participants suggested that a full picture of both what they can change and adapt when it comes to recommendation algorithms, as well as what they cannot change or turn off when using a particular service, would be ideal—understanding the boundaries of their control and clear statements about what cannot be done as a user would alleviate their need to read between the lines.

3. Data Collection and Data Sharing Outside a Platform

Finally, participants agreed that algorithmic transparency reports should include information about how information collected within a platform for content recommendation purposes is also shared with third parties, as well as how platforms obtain information from other sources to support recommendation algorithms. Participants were mostly concerned with the fact that they essentially have little to no insight into how their personal information moves around from one platform to another, and are left with many questions on this topic, some of which include: "*Is [the platform] collecting data on the site itself? Or is it also pulling data from other sources?*" (P10), "*are [platforms] selling data to third parties?*" (P24), "*Is [data] just being used for marketing or can it be used in some sort of [police] investigation?*" (P25).

"What I'm doing [on the platform] I don't mind them knowing…but I don't want them tracking me when I'm not actually logged in and using their platform."

For most participants, the default was that "*as long as [platforms] are not pulling information from outside sources, that's ok*" (P27). P27 elaborated: "*What I'm doing [on the platform] I don't mind them knowing...but I don't want them tracking me when I'm not actually logged in and using their platform.*" To address this concern, platforms should include detailed, specific and understandable information in algorithmic transparency reports about data sharing and collaboration with partners and third parties to inform recommendation algorithms.

WHAT INFORMATION SHOULD ALGORITHMIC TRANSPARENCY REPORTS MINIMIZE?

Along with topics that should be included in recommendation algorithm transparency reports, participants were also in agreement on certain aspects that are not of interest. Namely, the goals of a recommendation system, who benefits from it, and technical details. These topics were primarily met with skepticism and were therefore not perceived as valuable to users. These findings fall along the lines of prior work that suggests that as much information as possible is not always helpful, but that transparency requires more nuanced and intentional choices for sharing information (Ananny & Crawford, 2018). We elaborate below on the topics that participants were less interested in, and suggest some aspects of these topics that could be of interest, based on our conversations with users.

1. Recommendation Algorithm Goals

Most participants were not interested in receiving information about the goals of a particular algorithm. This is because they believed that they already "*have a good understanding of what the goal is for the system*" (P14): "*[It is to] feed people content to engage with the social media platform*" (P15), "*make your feed as personalized as possible*" (P16) and "*show you things that you would be interested in buying*" (P23) to "*sell more things*" (P18). On the more pragmatic side, participants argued that "*the [platforms'] goal is to make money. Anything [users] get out of it is a happy accident.*"

Users' skepticism and perceptions of what they think the goals of recommendation systems are present an opportunity for social media platforms to improve how those goals are communicated in more relevant and succinct ways. When providers explain the goals of their recommendation algorithms they should take their audience's skepticism into consideration, and thus (a) provide a concise explanation of the multiple goals of the system (including the platform's incentives, like helping advertisers target interested users), and (b) follow any general explanation with concrete information about how this works in practice. In other words, in order for people to find the information interesting and believable, providers should give short and straightforward assessments of their incentives, along with some more practical information that backs up their claims.

Since participants were confident about knowing what the platforms' goals are and expressed little motivation in learning more about the topic, a platform with a different or more nuanced goal for its recommender system will need to take great care to explain that clearly and in a way that users would be willing to interact with.

2. Who Benefits from the Algorithm

Similar to statements about the algorithm goals, participants expressed skepticism when we asked about the inclusion of information about who benefits from an algorithm design in transparency reports: "*who benefits from it? I already know that. It's the advertisers and [platform], so I don't care*" (P28). Like with the algorithm goal, we recommend minimizing information about who the platform thinks benefits from the system, especially if the answer they would like to propose is "the user." Users view such claims as insincere, even if there is some truth to it.

3. Technical Details

Participants shied away from explanations about how algorithms work, simply because they "*do not really care about that*" (P10), but also because they assumed they "*would not know what anything meant*" (P15). P7 mentioned that "*organizations often include that kind of information, and I skip it*."

Thus, we recommend that platforms share technical information more intentionally, and more tailored to audiences. While people with technical backgrounds, along with researchers, auditors and other groups may be more interested in technical details about how systems work, everyday users in our study were less concerned about that, and rather wanted to know more about how the choices in the algorithm design *impact* them. We recommend making these aspects the focus, with additional technical details shared separately with its target audiences, to reduce information overload.

WHAT STYLE AND LANGUAGE SHOULD ALGORITHMIC TRANSPARENCY REPORTS USE?

Straightforward, data-driven, easy to understand language is best for recommendation algorithm transparency reports. When we presented participants with excerpts of platforms' Privacy Policies, any "marketing" language ("*we use your information to provide and improve your experience*") or vague statements ("*we may receive and process information about your location*") were not well received. We elaborate on three prominent style and language characteristics that were desired based on our interviews with participants.

1. Specific

"*We have to tell you this, so we're going to, but we're not going to offer extra information*" (P23). This quote illustrates a key reaction to current information about recommendation algorithms that is shared through Privacy Policies: participants experienced most existing descriptions as fulfilling an obligation, lacking a sincere attempt to be transparent. They were mostly bothered by the fact that the Privacy Policy excerpts we shared "*do not give any specific information, [which] raises a few red flags*" (P11), but rather general information about what they "might" do (e.g. the example above: "*We may receive and process information about your location*"). While being specific may result in longer descriptions of the conditions under which something will or will not be done, participants agreed that specificity would be key to including statements that would be understandable and meaningful to them.

[W]e're committed to showing you content that's relevant, interesting and personal to you.

"[I read this as platforms saying:] 'Look how lucky you are that we are gathering your data.' This sounds disingenuous to me."

2. Direct

Participants were highly interested in reading more about some recommendation algorithm practices, but again, when we shared excerpts from real Privacy Policies, they were disappointed to see descriptions that predominantly attempt to portray the platform positively, instead of being direct about platform practices. P17 detailed: "'*we're committed to showing you content that's relevant, interesting and personal to you.' Look how lucky you are that we are gathering your data. This sounds disingenuous to me.*"

Sometimes, the text included information about fundamental features of a platform, for example: "*We collect and process, which includes scanning and analyzing, information you provide when you compose, send, or receive messages [...] Please be aware that messages sent to other users of [the platform] will be accessible by those users [...] We use your information to improve, support and administer the Platform, to allow you to use its functionalities, and to fulfill and enforce our Terms of Service.*" Participants did not appreciate explanations about how users should be careful that were interleaved with information about how data is collected and used *by platforms* for recommendation algorithms (as in the example above).

At best, it was perceived as "*condescending*" (P30). At worst, it was interpreted as a way of diverting the focus from what platforms were doing and contributed to a sense of distrust in the platform: "*It comes off as a little bit disingenuous...like they are trying to cover up the fact that they have [the data] by saying: 'Oh also your friends could do this [hold onto data] too, it's not just us'*" (P17). Similarly, P4 suggested this kind of formatting communicates: "*Hey, we're taking your information. By the way – here's a good tip.*"

Participants were more receptive towards language that seemed to be honest about the platform's incentives and needs, the involvement of advertisers, and so forth, as opposed to excerpts that made it seem that only the users' best interest was in mind. Thus, we recommend that platforms have a standalone report that shares information about platform recommendation algorithms practices and their use of personal data, separate from social media interpersonal aspects, such as how other stakeholders (like message receivers) might use the same personal information.

3. *Demonstrated*

Examples about how data and algorithms work on a particular platform helped participants follow and comprehend the information that was being shared: "*[The excerpt] is telling you, 'we're collecting this information to show you stuff about [something], because you searched [something].' That makes sense, you can put two and two together*" (P23). As noted by P11, "*the value of specific examples is just inestimable [...] 'If you search for mountain bikes, you may see an ad for sports equipment when you're browsing a site that shows ads served by us.' I really like that. It speaks to me in layman's terms, and it gives me examples*" (P11).

"[T]he value of specific examples is just inestimable [...] '*If you search for mountain bikes, you may see an ad for sports equipment when you're browsing a site that shows ads served by us.*' I really like that. It speaks to me in layman's terms, and it gives me examples."

Study 2 Findings: Alternative Forms of Algorithmic Transparency Reporting

In addition to formulating the guidelines above, we used the findings from Study 1 to initiate a design process to create examples of ways of sharing information about recommendation algorithms with users. The design process resulted in a final set of four prototypes of possible user interfaces that exemplify some aspects of future algorithmic transparency reports based on users' expressed needs.

The four prototypes present an imagined social media platform, intentionally designed to not resemble any specific platform in order to maintain as much neutrality as possible for users, and to potentially be applicable to a range of platforms, big and small. The prototypes, as a whole, take a personalized approach to presenting information to users. Thus, in contrast to Privacy Policy excerpts that included general descriptions, the report prototypes we presented were designed as personalized, single-user reports. This is because the topic of "how things apply to me" was frequently raised in Study 1, suggesting that a personalized approach would prompt more discussion about what is or is not valuable for users to know about, while providing a novel way of engaging with information that directly impacts them. Nevertheless, the prototypes bring forward a range of other considerations for algorithmic transparency reports, regardless of whether they are personalized or described in general terms. More about the design process that led to these prototypes can be found in the Appendix: Methodology.

In Study 2, we invited all participants from Study 1 to continue their co-design collaboration with us, and to reflect on the prototypes that we created based on their responses. Previous research on co-design processes suggested that it is worthwhile to include participants in multiple stages of the design process in order to impact the design process itself, as opposed to solely giving feedback at a single point in time (Westerlund et al., 2003). By presenting the interface prototypes to the same group of users interviewees, we set out to understand how these expressions of algorithm transparency may work (or not work), and to identify the strengths and weaknesses of each approach.

In this section, we describe the design of each prototype, the concepts that it brought forward in discussion with participants, the responses to it and conclusions for future designs. The prototypes that we present here, and that we presented to users in Study 2 are not intended to serve as templates for algorithmic transparency reports or to suggest what a report *should* be, but rather to raise alternative design directions for further research and exploration. Each prototype presents several ideas about what a transparency report may include, but the four prototypes do not exhaust all design options in this space. Instead, they set out to inspire platforms to explore these and other design opportunities based on our research findings.

PROTOTYPE 1: YOUR QUANTIFIED SELF

Design Motivation

In Study 1, many participants indicated that knowing what information is collected about them to be used in recommendation algorithms is a top priority. Thus, this first prototype set out to give an overview of the data that a platform collects about an individual. It is presented in an approachable and data-driven way, using large numbers, straightforward labels and a bubble graph. The information presented is in the spirit of "The Quantified Self", a concept that quantifies user behavior, to then be delivered as content to the user (Lupton, 2016), such as step counts or "screen time" applications. Most recently, the "Quantified Self" approach was implemented in Spotify's Wrapped (Steele, 2021), which was specifically mentioned by participants as a legible, interesting and engaging report.

Implemented Concepts

In this prototype we focused on the concept of a personalized transparency report format in the spirit of "The Quantified Self," in an attempt to reflect some of the data that is presumably collected by social media platforms to use for algorithmic recommendations. The prototype is also primarily represented graphically (see the full prototype as Figure 2). Both personalization and visualization were drawn from participants' feedback in Study 1.

Responses

> **"Even though [the report] is telling you stuff about [using] your data, because they are being transparent about it, I feel safer using this platform than I would any other platform that doesn't release a report like this to me." (P1)**

Because this was the first prototype we presented to participants, it was their first encounter with the notion of personalized transparency reports, as well as with the overarching data-driven and graphical presentation approach that we took. Participants' responses to these two aspects were overwhelmingly positive. The graphical presentation was perceived as "*digestible*" (P1), "*easy to take in*" (P14), "*open, honest, and trustworthy*" (P24), and generally "*super transparent and super easy to navigate*" (P8). They agreed that having a visual representation of the information "*feels very forward and trusting as opposed to just seeing it in a body of text*" (P1).

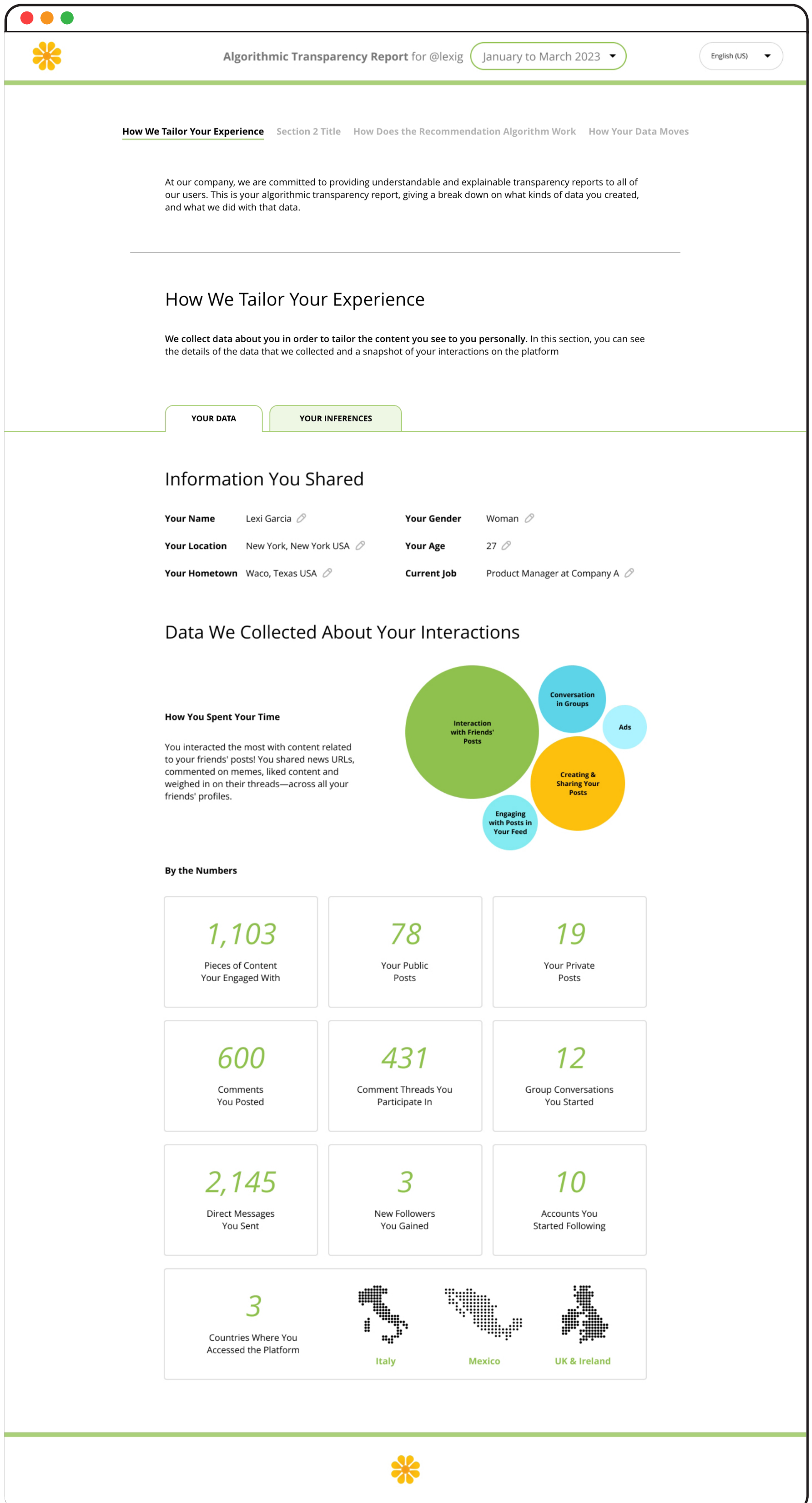
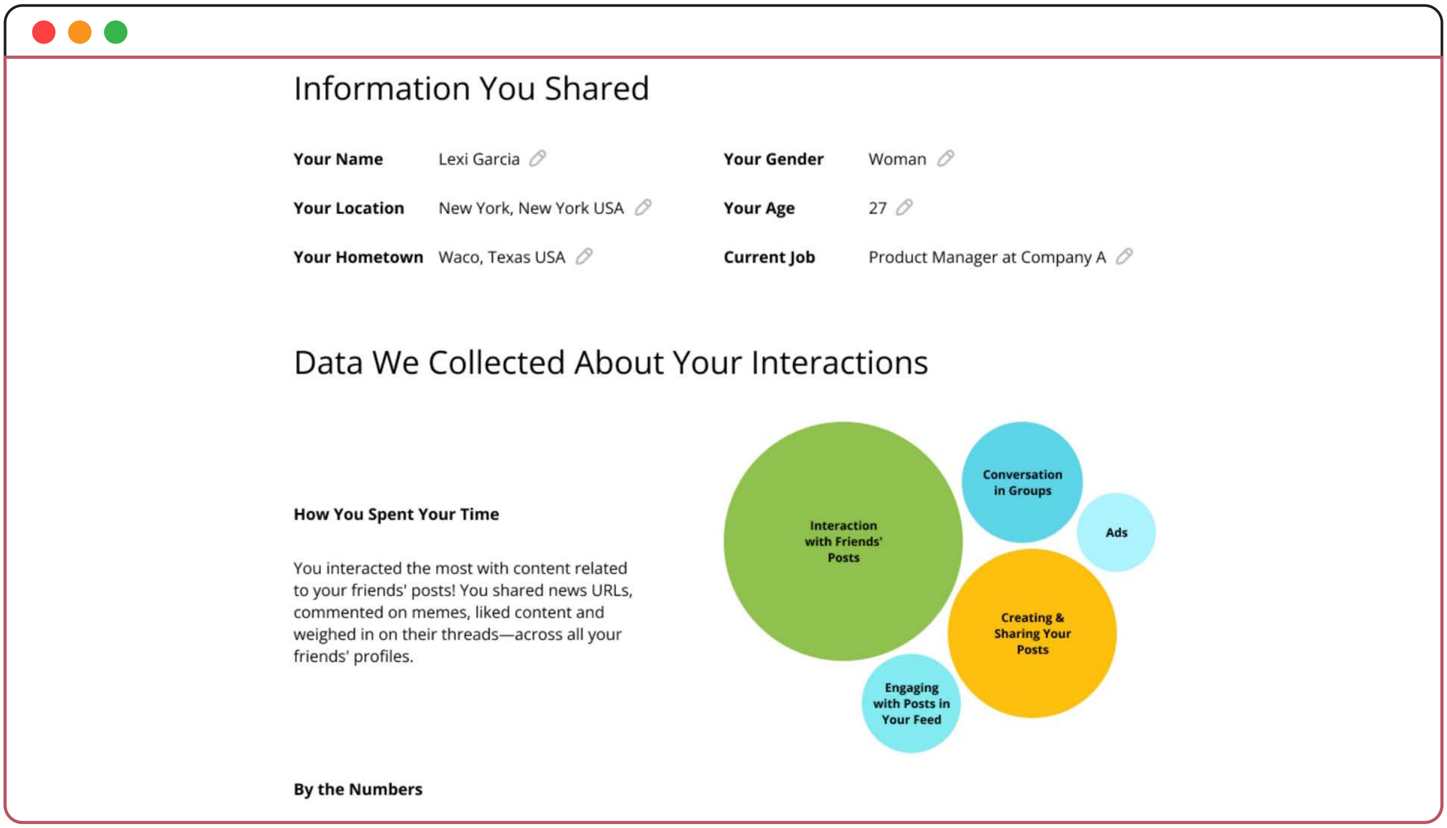


Figure 2. Prototype 1: Your Quantified Self. The first prototype that was presented to users in Study 2. The prototype presents basic information shared by the hypothetical user, how they spend time on the platform, and some statistics about their behavior and engagement with the platform.

Source - CDT



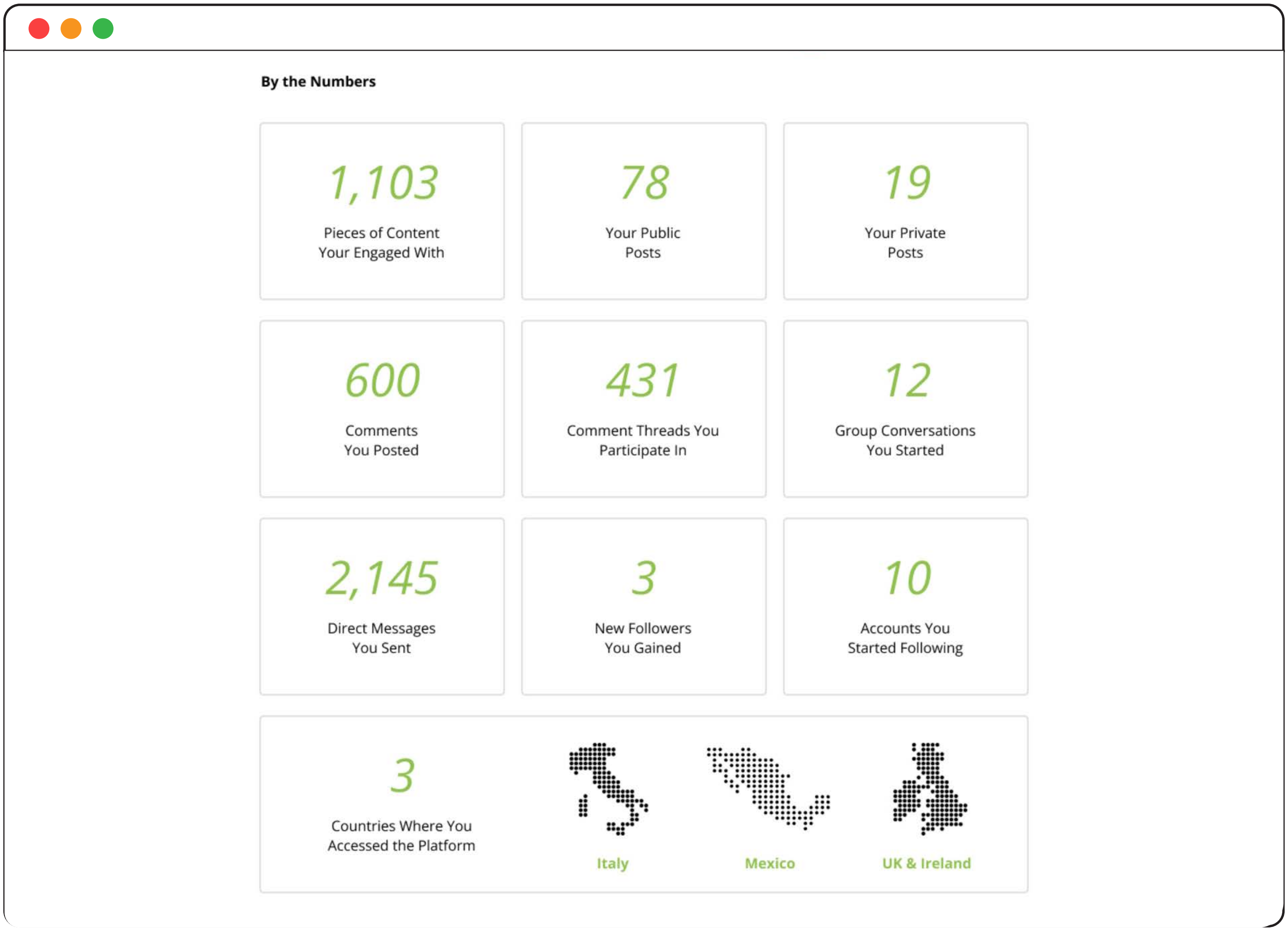
▲ **Figure 3: Data We Collected.** Part of prototype 1 focused on providing users with a visualization of how they spent their time on the platform.

Source - CDT

For the personalization, participants were initially surprised by this concept of the report, as they mostly "*have not seen anything like this before*" (P4). While participants agreed that not every algorithmic transparency report has to be personalized, as this requires significant effort, they thought that "*having [the report be] personalized makes it a lot more relevant and more interesting when taking in the information*" (P14).

Lastly, this prototype included information about the user's engagement with the platform, e.g., "The Quantified Self"—how much time they spent on different features (Figure 3) and how many interactions they had (Figure 4). Most participants found that this information "*wouldn't be any revelation, but [that] it would be interesting to see.*" (P11), in other words—not that important, but perhaps nice to have. Some participants were more skeptical about the usefulness of this information, like P17: "*I don't know what I would do with information that says I've posted X amount of times.*" Similarly, P7 said: "*I don't care how many private posts I made or how many conversations I started, that just makes me feel like they're monitoring me [...] it's not what I want to know*" (P7). Instead, participants noted, platforms should share information about the data that is collected about users by the platform and used in their recommendation algorithm system.

"I don't care how many private posts I made or how many conversations I started, that just makes me feel like they're monitoring me…it's not what I want to know."



▲ **Figure 4. By the Numbers.** The bottom part of prototype 1 presents information about how the user hypothetically interacted and engaged with the platform.

Source - CDT

PROTOTYPE 2: INFERENCES

Design Motivation

In this prototype, we focus on inferences made by social media platforms. In Study 1, participants indicated that inferences were a very intriguing topic for them to learn more about. They agreed that if platforms were making inferences about them, these should then be shared with users. They were also curious to find out what these inferences would be. We therefore created a prototype that specifically focused on the topic of inference-making for recommendation algorithms in an attempt to explain what inferences a platform might make about users to support recommendation algorithms.

Implemented Concepts

In this prototype, we introduce the concept of an interactive tool to walk through the inferences that have been made about an individual, and potential means of control over these inferences. This prototype also includes personalization and straightforward language to communicate collected data. For the full prototype, see Figure 5.

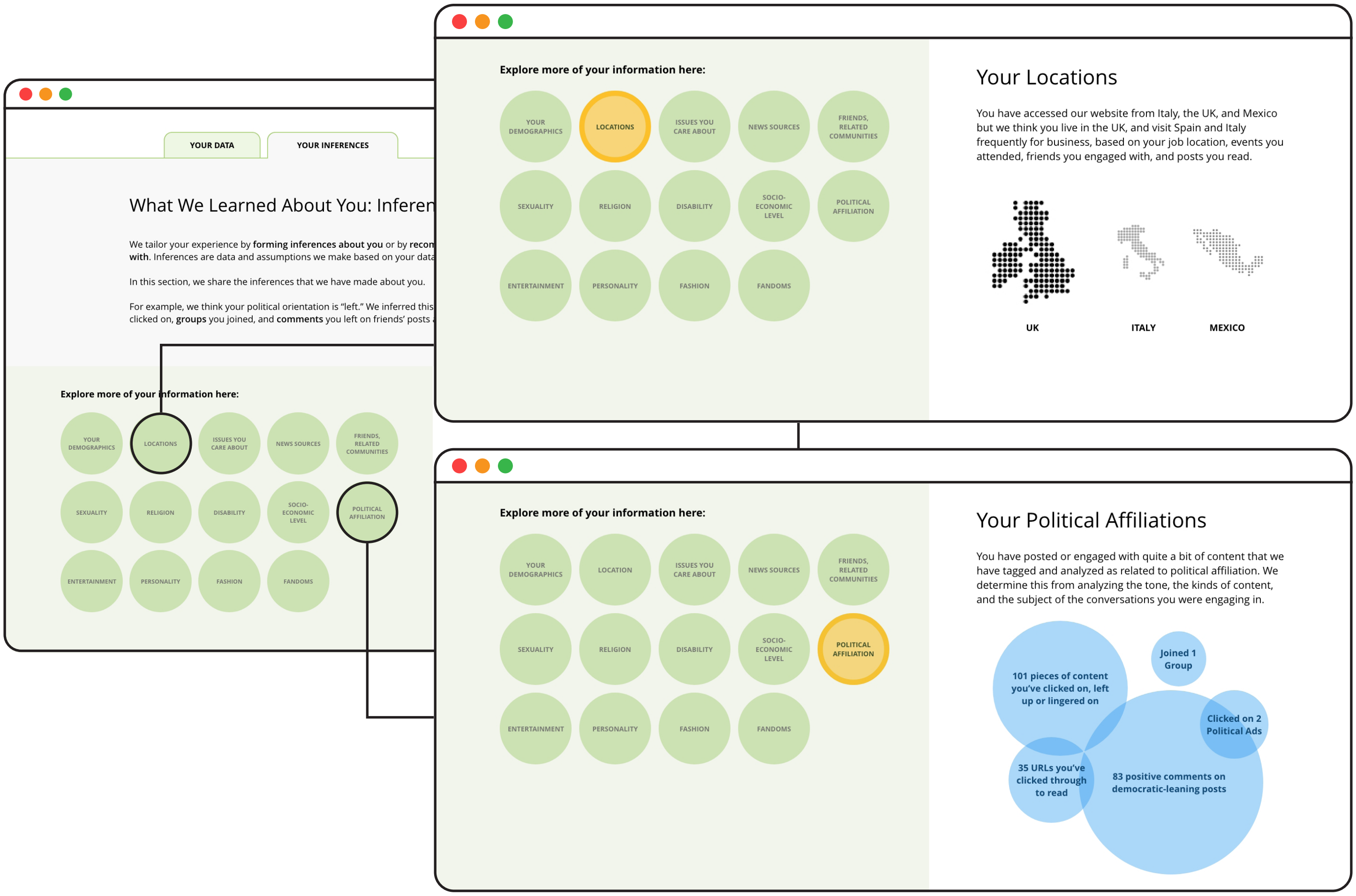
Figure 5. Prototype 2: Inferences. The second prototype that was presented to users in Study 2, which focused on inferences made for the use of recommendation algorithms.

Source - CDT

Responses

> **"It's such a new thing reporting what they do with your private information. It's still kind of surprising when they tell you 'we've looked at your posts and the news articles and stuff you like and we think you're a Democrat. We think you live here and you work here and your friends live here and your parents live over here.' They start making all these inferences and I think at some point it gets kind of spooky when they start being more right. You become less of an individual person and more of just a conglomerate of data. [...] [But] if they have that information I want to know it." (P23)**

In this prototype, we presented participants with information that a platform might share about the different types of inferences that the platform made about them, and the information on which those inferences were presumably based (see Figure 6). Participants frequently had a hard time commenting on



▲ **Figure 6. Explore More About Inferences.** This part of the prototype presents an interactive view that a user might have about inferences made for recommendation algorithms. The prototype suggests that users could click on a topic of interest (left) and learn about the inferences that were made by a platform related to that topic (right).

Source - CDT

the report format itself, as they were mostly distracted and interested in the content details. However, these would likely vary between different platforms, based on the types of information and inferences that each platform makes.

Nevertheless, participants “*appreciate[d] that [platforms] are being forward about [what inferences were made]*” (P1). The presentation of information in the prototype defied their “*expectation that it would be a wall of text that discourages […] reading*” (P3), and instead presented information about inferences in a way that is “*very easy to approach*” (P3). Many participants found it to be “*the most useful page [they have] seen so far*” even though they “*did not like that [platforms are] collecting this information*” (P7).

PROTOTYPE 3: HOW THE RECOMMENDATION ALGORITHM WORKS

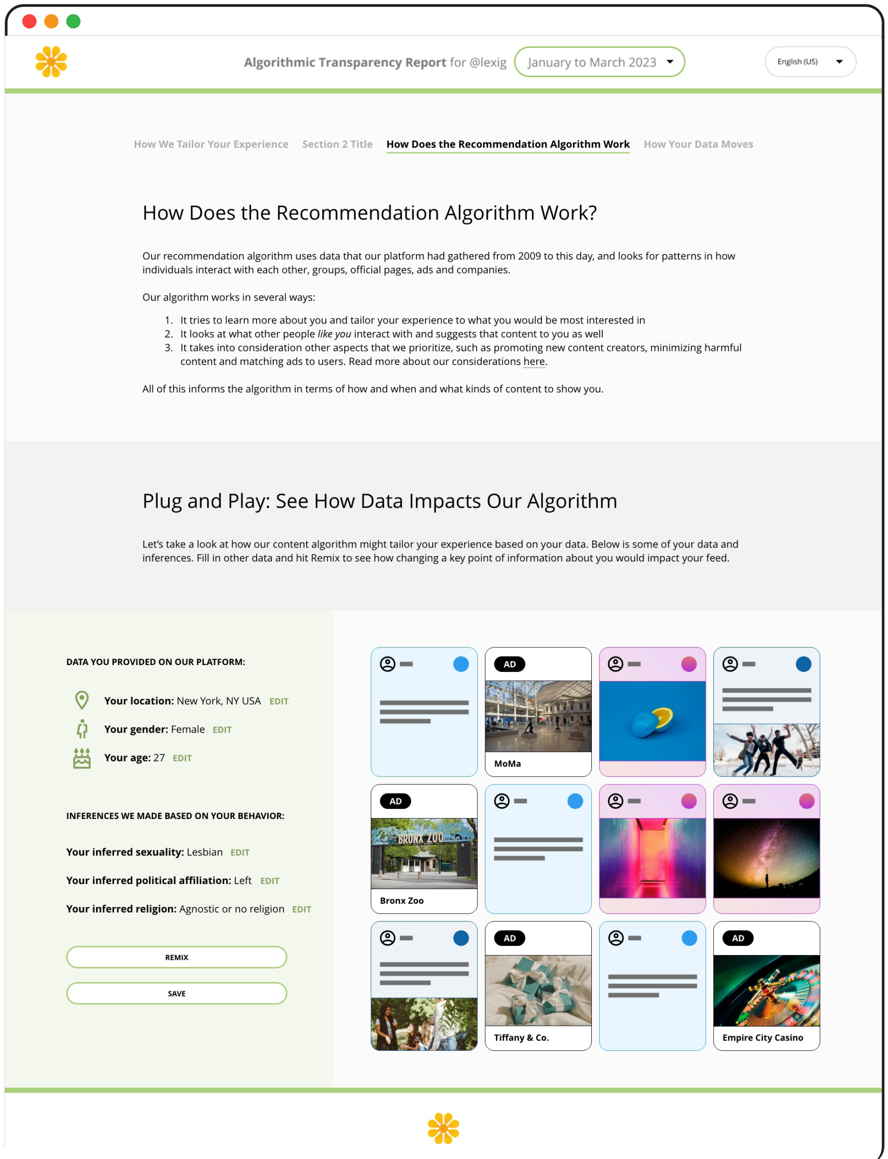
Design Motivation

In Study 1, participants indicated that they would like to know more about the relationship between input and output of a recommendation system. In the third prototype, we created an interactive feature where participants could alter their hypothetical input data and learn about how it impacts the output.

It is worth noting that recommendation algorithms are frequently too complex to make a direct connection between a single piece of input and output (Burrell, 2016; Nicholas, 2020). However, in Study 1 we found that participants cared little about being able to track the exact processing each piece of information undergoes, or in other words, *how the algorithm works*, and were rather more interested in an overview of how it impacts them.

Implemented Concepts

The interactive "Plug and Play" feature in this prototype set out to allow participants to play with some of their input data and observe the resulting output of a recommendation system (see Figure 7). Even if not precise, some presentation of the relationship between the two may contribute to explaining to users how the algorithm works. This concept is also in line with prior explainability efforts such as AI Explorables (AI Explorables | Google, n.d.) and the TensorFlow What-if Tool (Model Understanding with the What-If Tool Dashboard | TensorBoard, n.d.). In addition to the focus of this prototype



▲ **Figure 7. Prototype 3: How The Recommendation Algorithm Works.** This prototype was the third that was presented to users in Study 2, and which included an explanation of how the platform's recommendation algorithm works, followed by an interactive "Plug and Play" feature.

Source - CDT

to communicate the relationship between input and output, we also incorporated graphical (as opposed to textual) representation, interactivity and personalization. The full prototype is presented as Figure 7.

Lastly, this prototype was intended to test an opportunity to control and change data, given that some data may have significant impact on what users see. As suggested by prior work on “Actionable Recourse” (Ustun et al., 2019), in this prototype we too propose providing users with control over input variables that platforms use, and give them an opportunity to change the decision-making of a model.

Responses

> **“They’re giving me access to the actual variables used and letting me see how they work.” (P3)**

Before the “Plug and Play” feature, we included a paragraph that described, in bullets, the main ways that algorithms determine what people will see. Participants were not very interested in this explanation and tended to skip right to the next, more interactive part of the prototype. That is not to say that such information should not be included, but rather, that participants were naturally drawn to the more visual and interactive aspects of the report.

Participants appreciated that the “Plug and Play” feature was “*giving [them] access to the actual variables used and letting [them] see how they work*” (P3), as a way to learn about the recommendation algorithm. P3 further explained: “*I love this plug and play functionality [...] It really lets me experiment with why I’m seeing some random thing in my feed. Selfishly, it also gives me the tools I need to manipulate my feed in ways that I would prefer.*” P14 agreed that having this functionality would allow them to “*fine tune [their feed] to who [they] actually [are] and [to] what content [they] actually want to see.*”

The “Plug and Play” feature, in the format we presented it in, was also perceived by participants as somewhat limited in its exploration capabilities over time. Some participants described the feature as something that they would “*probably [play with] once*” (P8). Nevertheless, if the goal is to inform users about how an algorithm works and what kind of variables are included in the decision making, a single “*fun little interactive component that would demystify an algorithm*” (P8) could be sufficient.

Finally, a couple of participants found that this feature could be a useful tool to “*understand more that social media can become an echo chamber*” (P23), and to make space for alternative views by giving them a tool to “*see how [their] friends interact with the world when they’re coming at something from a different political or religious perspective*” (P15).

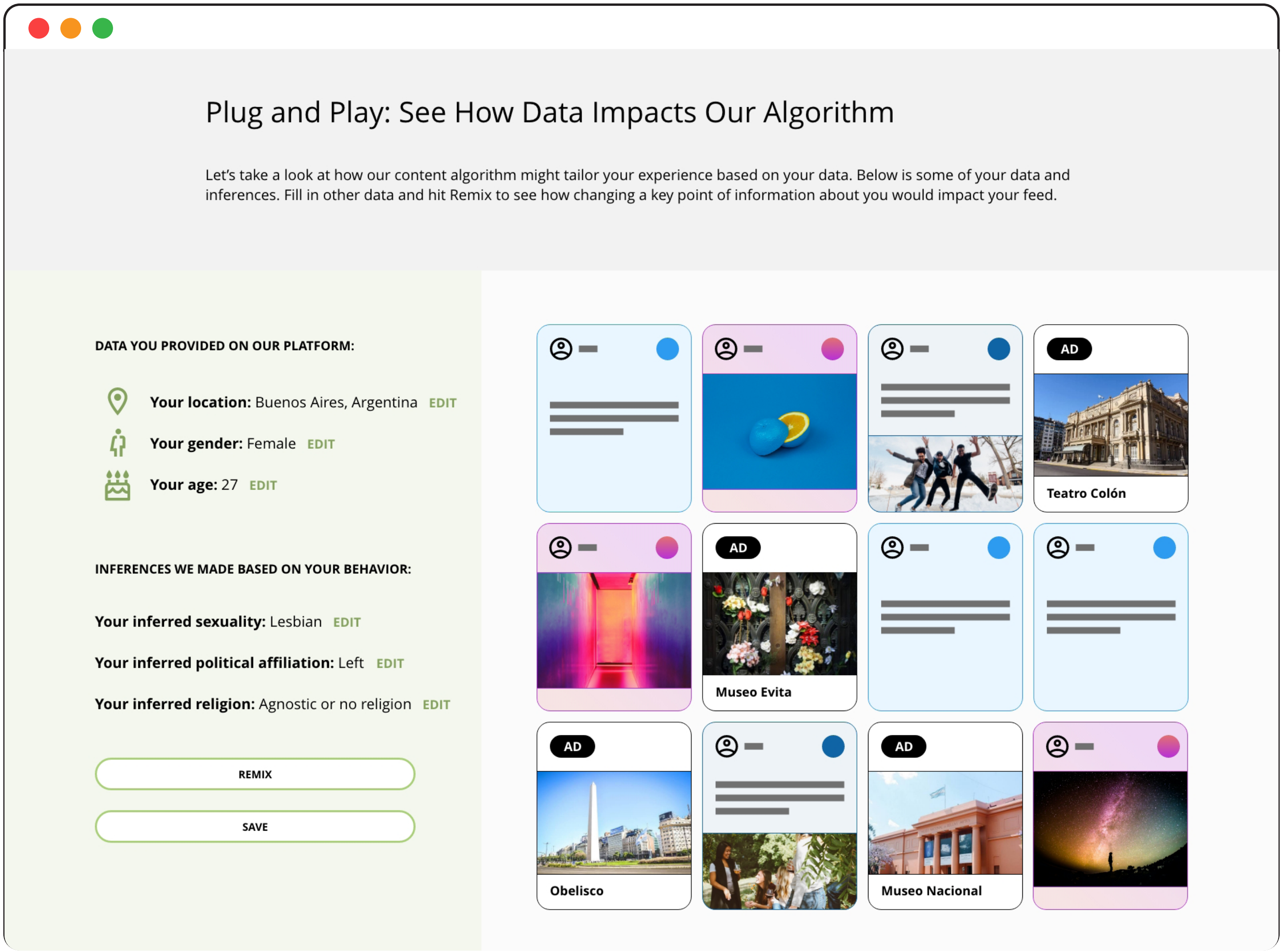


Figure 8. Plug and Play. The Plug and Play feature shown in this figure suggests an interface that would allow users to learn about the relationship between the algorithm's input and output in a personalized way.

Source - CDT

PROTOTYPE 4: DATA MOVEMENT ON AND OFF THE PLATFORM

Design Motivation

In this final prototype, we attempted to tackle participant requests from Study 1 to better understand where data that informs a recommendation algorithm is coming from, and how data that is collected on a platform presumably for recommendation algorithms may also be shared with third parties.

Implemented Concepts

The focus of this prototype was on a topic that is rarely discussed in the context of algorithmic transparency: how data is obtained to support recommendation algorithms and how data that is collected to feed into recommendation algorithms on a specific platform is shared elsewhere. We also attempted to add some aspects of control and interactivity to this shared information. For the full prototype, see Figure 9.

The prototype presents: (1) hypothetical data types that a platform collects within and outside the platform and uses as input to its recommendation algorithm; and (2) hypothetical data types that a platform collects to use for recommendation algorithms, but also shares with partners and third parties for similar or other purposes (see Figure 10).

Responses

> **"It's honest with you that, yeah, we do sell some of your information so you can go ahead and get a more personalized view. Obviously, I think that's common sense nowadays that all social media does that. So the fact that they're owning it and being honest [about] it, I like that. More importantly, [there is] the ability to change it [...] It's something that I wish all social media [would do], but it really, it puts the power back [into my hands]." (P22)**

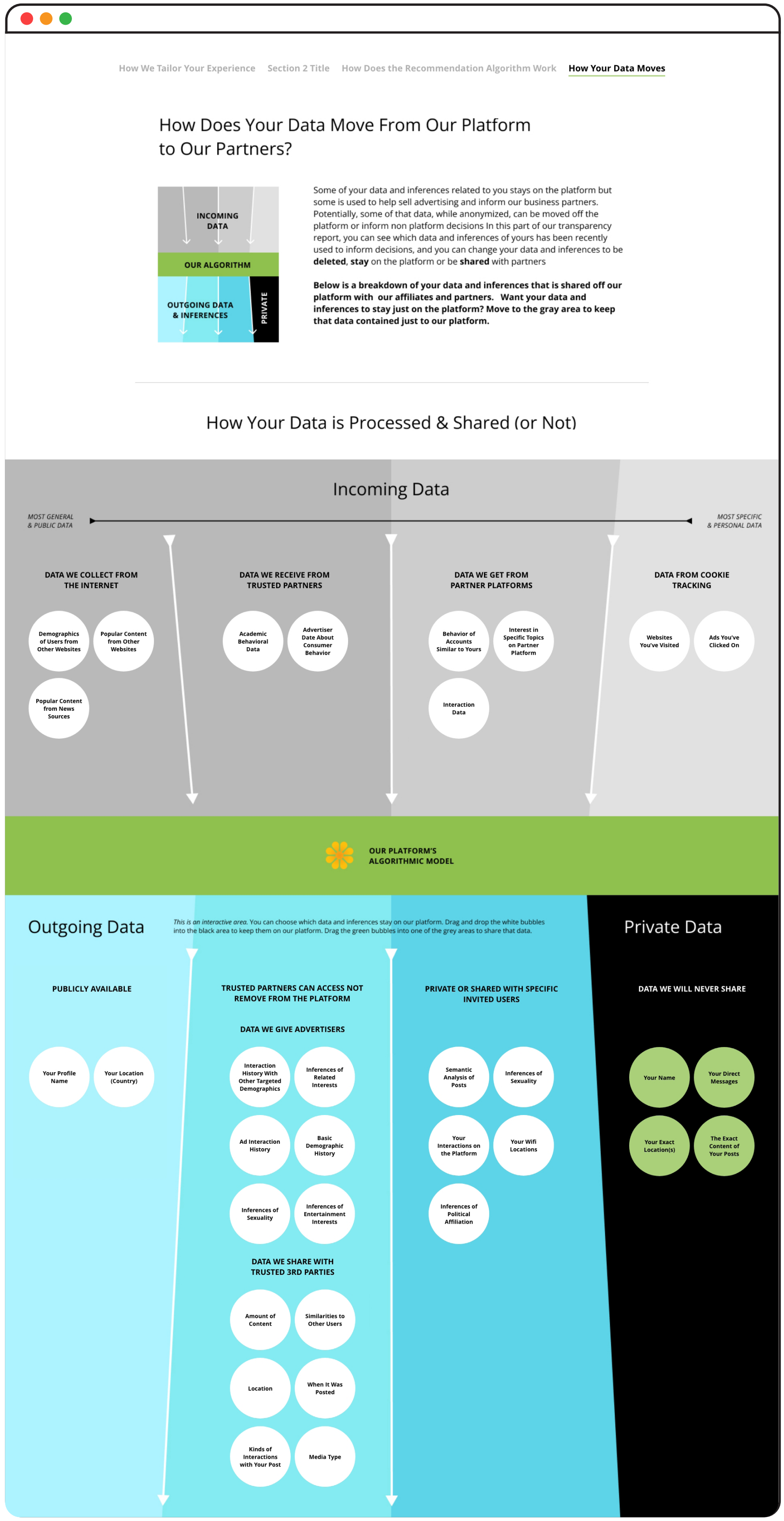


Figure 9. Prototype 4: Data Movement On and Off the Platform. The final prototype that was presented to users in Study 2, that focused on communicating how data is obtained and transferred from and to third parties.

Source - CDT

> **"This is transparency to me [...] It feels like transparency because they're saying, 'all right, yes, we do share your data with partners, but we anonymize it [...] and we're going to tell you what parts of your data have been used, have been shared with our partners.' That is exactly what I would want." (P11)**

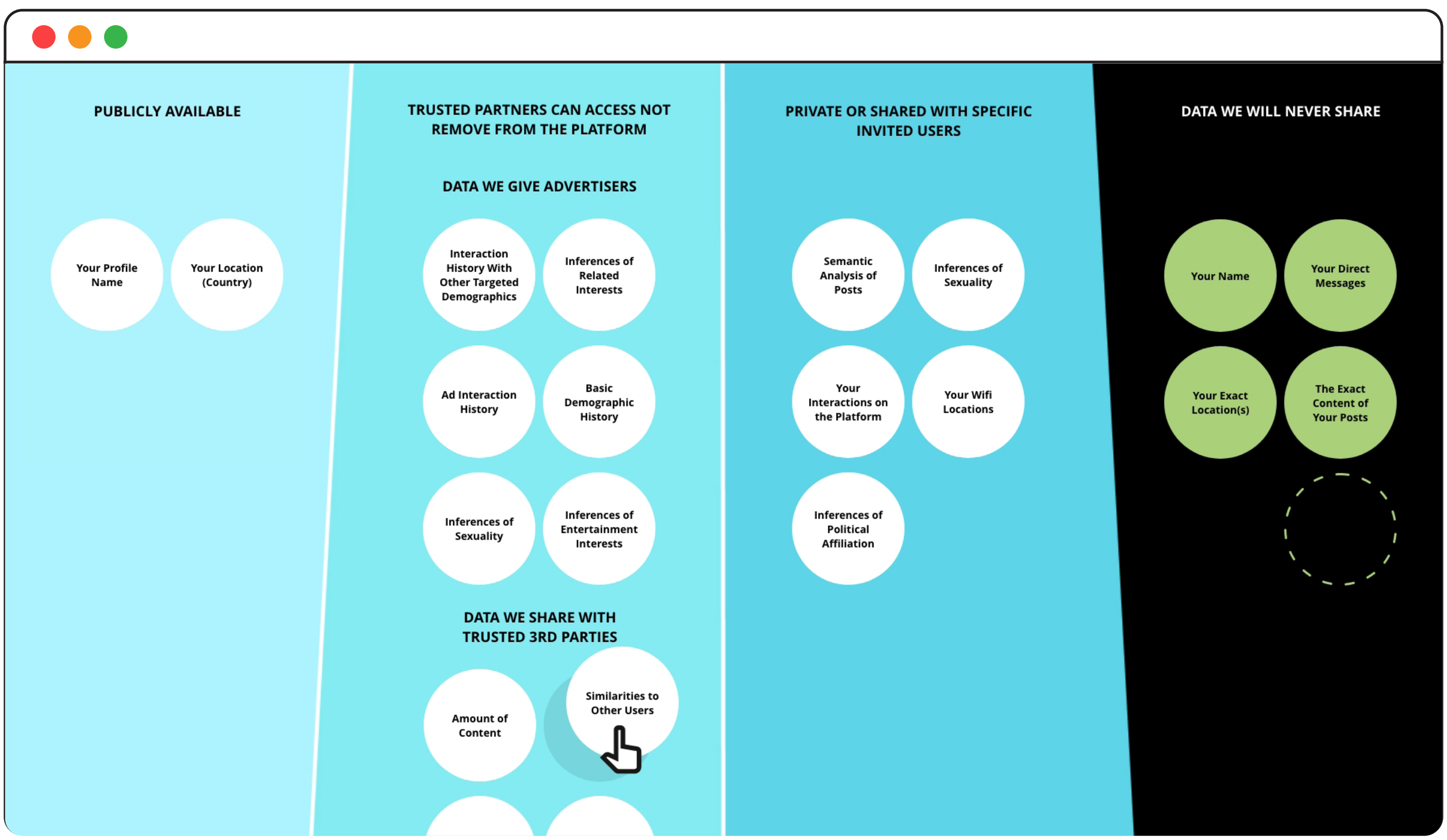
In addition to the content itself, the aspects of this prototype that stood out to people most were the visual representation, the granularity of the information about what data is shared with other platforms to use, and the ability to control this sharing.

Participants thought the visual representation in this prototype "*makes it very easy to see each tidbit of data*" (P1), and as a result "*makes it easier to take in*" (P14). Here too, the visualization stood in contrast to "*normal privacy data pages, which are just walls of text that almost nobody ever reads*" (P17).

Most participants found it "*useful to have [the data] broken down into categories*" (P17). P24 explained: "*the more I know, the better. There [are] no questions [unanswered], I'm not doubting anything.*" A couple of participants wanted even more details, suggesting "*an option to hover over a bubble so that it brings up a popup [that] tells you exactly what [the information type] is*" (P1).

▼ **Figure 10. Outgoing Data.** This part of the prototype suggests that platforms may share with users how information the platform collected to inform algorithms is also shared with partners. An interactive control feature would allow users to change how different data types are shared externally with third parties.

Source - CDT



Like in other prototypes, the ability to impact how the platform collects and shares data strongly resonated with participants. P11 elaborated: "*The element of control is valuable. It makes me feel safer. It makes me feel more in control. And without more explanation, there are things that I would definitely drag and drop.*" The ability to say to platforms "*Nope, I'm sorry, I don't want you doing that*" (P3) was perceived as novel, as "*[usually] Facebook or some other apps just tell you that they've collected [data] and give you a brief reason why. But [the prototype] really makes it a customized feature, where you can control some of the things you see and some of the information that goes out to others. It [gives] the power 100% back to the user*" (P22).

"[Y]ou can control some of the things you see and some of the information that goes out to others. It [gives] the power 100% back to the user."

The relationship between the level of detail and control provided here is somewhat complicated and reflects on a larger tension across all prototypes between giving users enough information, but not overwhelming them, as also raised in prior research (Ananny & Crawford, 2018). On one hand, participants need detailed information to make sense of how things work; on the other, controlling these details one-by-one may become tedious and overwhelming. To resolve this contradiction, some users suggested that platforms might offer "*a single button*" (P17) or a "*toggle*" (P4) setting that would rearrange detailed information in predefined setting categories. For example, in our prototype, adding a "high privacy, low personalization" setting could automatically move the bits of information into the "private" sphere, and allow users to further customize their preferences from there.

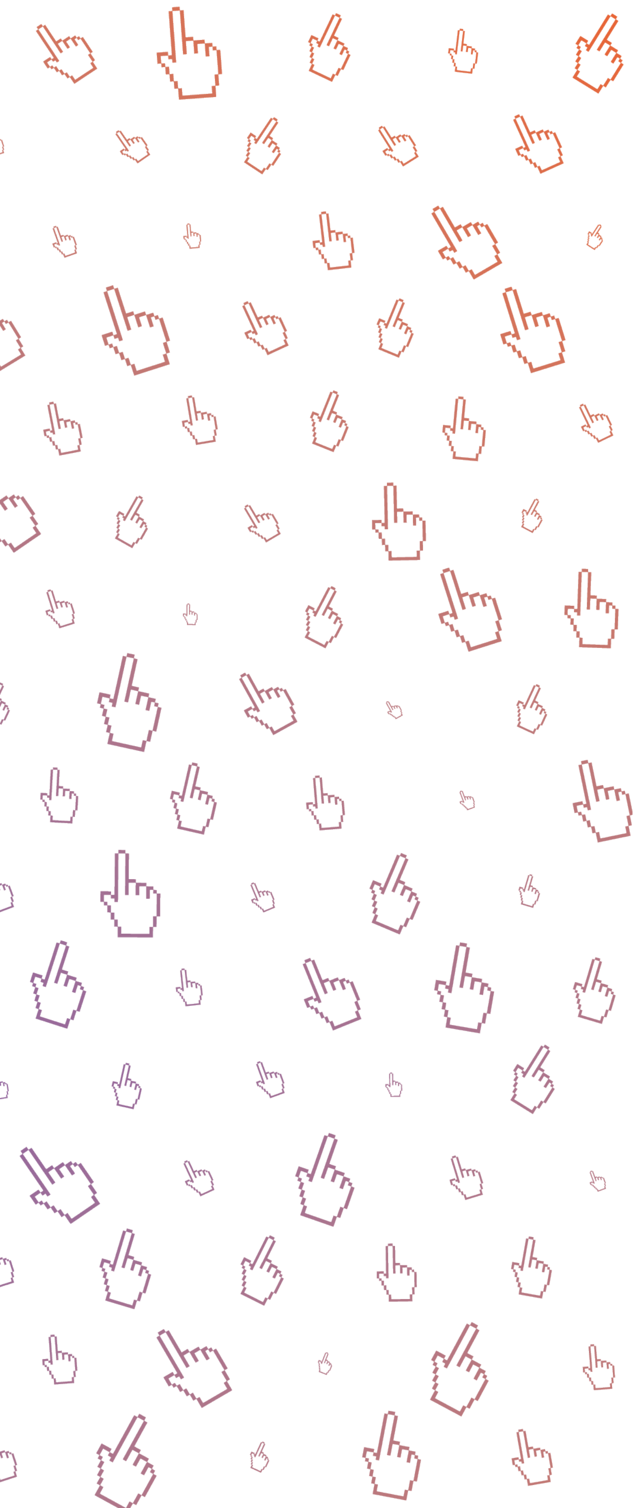
Recommendations

In recent years, users, civil society, and governments have been urging social media companies to maintain and improve transparency on all fronts, along with much effort to standardize some of these requirements through legislation (Vogus & Llansó, 2021). Given how little users, researchers and policymakers know about how social media algorithms work, transparency will promote more intelligent conversations about recommendation algorithms that people use on a daily basis. In addition to our primary goal of advocating for users and their rights, improving transparency can also benefit platforms themselves by allowing them to gain their users' trust.

Yet, the question of exactly what to report on and how to do so is not easy to answer—algorithmic transparency reports need to strike a complex balance between what users want to know, what legislators, researchers and auditors want to know, and what platforms would like to share with the public. In this report, we focus on the elements that are most critical to *users*. Users deserve to have greater agency over their platform experiences, much like they can control what stations or shows they watch, or what books they read. Social media platforms should strive towards a similar sense of transparency and control in their recommendation algorithms.

Users have a unique perspective on what would be valuable and important *for them* to know. Nevertheless, their perspectives are not always included in the creation and design of transparency reports, and, as a result, a lot of the information that is shared is either overly complicated and not very approachable, or perceived as obvious and even condescending. Given that users may already be somewhat skeptical of platforms' motives, platforms should pay extra attention to user needs and address them adequately in the context of recommendation algorithm transparency reports.

Based on Study 1 and Study 2, we summarize the most critical information for platforms to communicate to users in recommendation algorithm transparency reports.



Content: Emphasis on Personal Data and How it Informs Recommendation Algorithms

Participants were unanimously most concerned about what data is being collected about them to be used in recommendation algorithms, whether that is data they provide, inferences made about them, or personal data gathered from external sources. In the current climate of lowered trust in social media platforms, when users do not explicitly receive this information they tend to assume the worst: that they are being surveilled, that every move they make online is recorded, and that all of that data is used maliciously.[2] Thus, platforms should make a significant effort in communicating how their recommendation algorithms make use of personal information in a way that is clear, straightforward, and engaging to users. Motivating users to learn more could improve their trust in platforms and ultimately make them more satisfied with the service.

When possible, platforms should also provide information about how specific data is being manifested in people’s content experience, as opposed to saying something generic such as “*we collect personal data to improve your experience on the platform.*” Including the rationale will likely help users understand the considerations, and may allow them to make a conscious decision to opt into a particular personalized experience that the platform offers. Participants were even open to platforms’ monetary considerations and the need to support advertisers on the platforms (e.g, “*I recognize that [platforms] need to make money. And I support them making money, because they’re giving me a good service*” (P7)), so sharing the fact that data is collected to serve ads, for example, would likely not come to participants by surprise. Rather, participants positively perceived platforms that were straightforward about it.

Further, encouraging platforms to share all the information that they collect to be used in recommendation algorithms could also provide a venue for platforms to review their practices, and an opportunity to implement data minimization and only collect what is critical for the function of a recommendation system.

Presentation: Visual, Interactive, Personalized and Controllable

Our findings suggest that recommendation algorithm reports should strive to be visual, interactive, personalized, and as controllable as possible.

2 This is along the lines of CDT’s prior work that suggested that lack of transparency around shadowbanning leads to mis- and disinformation about how platforms operate (Nicholas, 2022).

VISUAL AND INTERACTIVE

Our findings suggest that an interactive presentation of algorithmic transparency reports is key to appeal to broader audiences, in line with prior work that shows the importance and strength of presenting information in a visualized and interactive way (D'Ignazio & Klein, 2020; Friendly, 2006). Interactive presentations do not have to be complex interfaces that require significant resources—our work shows that simple graphs and approachable data-driven representations of information were sufficient to draw participants' attention and interest.

PERSONALIZED

Users, not surprisingly, were most interested in information that was directly about them or impacted them. Because of that, the idea of having personalized reports was perceived as appealing, especially in the context of recommendation algorithms that are personalized by nature. Having personalized reports allowed them to directly observe how their profile is set up and change it as needed. That said, personalized reports require extensive resources that smaller platforms might not have. Our findings from Study 1 suggest that participants' expectations from a transparency report were to see information that applied to all users; thus, sharing general information that is detailed and approachable would like be satisfatory and increase their trust.

CONTROLLABLE

Most participants valued the ability to control recommendation algorithms that were personalized for them, and, as a result, to impact what they see. While research suggests that default settings should move towards better practices (Watson et al., 2015), there is also evidence that users increasingly seek control and adaptability based on their personal values and needs (Prince, 2018). The level and detail of control would likely vary between platforms, with platforms that have more resources potentially providing more granular features and control mechanisms for their recommendation algorithms.

Beyond control functionality, participants expressed a need for a high-level reporting of *what can or cannot* be controlled on a platform. Participants sought information about what *cannot* be altered or personalized without undermining platform functions just as much as what *can* be done. Thus, we recommend that an algorithmic transparency report include clear communication of where participants can make changes or opt out, and where they cannot.

Language: Specific, Direct and Demonstrated

Recommendation algorithm transparency reports should be specific, direct, and demonstrated. Participants’ preference was for data-driven information that is as straightforward and specific as possible. For example, attempts to explain how platforms enhance user experience or why their algorithm design is ethical were received as disingenuous and out of place in a transparency report. Instead, participants wanted the data itself, so that they could make their own judgments about the cost or impacts on the user and whether that tradeoff was worth it.

This preference was also reflected in our prototypes—participants were fond of the presentation of information in prototype 1 which primarily presented numbers and graphs. In prototype 2, we included a paragraph that attempted to explain in lay terms a “story” about what inferences the platform had made to feed into users’ content. Participants found that format unsettling and creepy.

Thus, we recommend that platforms avoid lengthy descriptions, and instead take a data-driven approach to sharing information about how their recommendation algorithms work, while using frequent examples. Any additional narrative should be kept to the minimum needed to make the presented data understandable. A data-driven representation may also make the sheer amount of information more digestible for users, as suggested by responses to prototype 4, that presented how information is gathered from and shared with partner platforms.

Conclusion

We use this work to highlight the importance of recommendation algorithm transparency to users. Current information about how recommendation algorithms work and make use of personal data is sparse, and the information that is shared is difficult to find. When users are kept in the dark, they tend to assume the worst about how technology works, as suggested in this and prior research efforts (Nicholas, 2022).

Also, as other research has further explored, the lack of transparency about how content recommendation algorithms work has an even more significant impact on marginalized communities (Antonie, 2022; Asian American Disinformation Table, 2022). For example, social media content algorithms have been shown to disproportionately flag speech by LGBTQ+ communities by not considering the context (Oliva et al., 2021), or to unintentionally demote content based on race, gender and political affiliation (Haimson et al., 2021).

We encourage platforms to voluntarily share information with users in a way that is compelling, approachable and straightforward to gain their interest and trust, and to allow users to hold platforms accountable for their practices.

The combined two studies in this report present initial user insights into how platforms could provide recommendation algorithm reports that are designed for users. They also highlight some of the challenges, as well as limitations, of such reports. Nevertheless, our findings are just a starting point—in this work we tested a limited scope of topics and implementations that specifically considered everyday users as the audience. User bases of different platforms end up having very different wants and needs, and thus any service is going to need to tailor its reporting to its particular products and use of recommendation algorithms, and to its particular user base.

As platforms vary in size, resources, goals, and most importantly, their audiences and user-bases, we encourage platforms to implement similar co-design approaches and include users in the process of creating and iterating on algorithmic transparency reports that are meaningfully useful for their users.

Further, users are not the only audience that should be included in the design process of transparency reports—researchers, auditors, policymakers and others also need access to information that would address their needs (Vogus, 2022) and should be addressed in a separate effort.

We hope to have set an example of a co-design process that involves users in the process of gaining insight into how information can be shared in a meaningful way, and to continue the conversation with stakeholders, policymakers, civil society, and users towards building more transparent social media platforms in the future.

Appendix: Methodology

The study included a total of two hours of co-designing with people (one 75 minute session and one 45 minute session). In these sessions, participants were invited to participate in a range of activities towards co-designing algorithmic transparency reports. The co-design process included:

1. **Formative Work**—Literature review and framework forming for Study 1;
2. **Study 1**—Learning about current perceptions regarding what people would like to know about recommendation algorithms (using Card Sorting and Semi-Structured Interviews);
3. **Design Process**—An internal design process which built on the findings of Study 1 to form tangible prototypes that manifest some of the raised topics and ideas; and
4. **Study 2**—Reflections on ideas of future recommendation algorithm transparency reports (using Experience Prototyping).

Formative Work

The study was based on several selected topics related to recommendation algorithms. These did not necessarily form a comprehensive list of all possible topics, but rather an initial set to begin the conversation with participants. The range of activities based on these topics were designed to also leave space for unknown topics of importance that might surface in the co-design process.

We primarily relied on two frameworks from prior research to form the list of topics: Koene et al.’s framework that discusses what algorithmic accountability should include (Koene et al., 2019) and Schelenz et al.’s work on best transparency practices for personalized algorithms (Schelenz et al., 2020). Through internal iteration and reflection, we narrowed down to the topics that seemed most relevant in the context of recommendation algorithm transparency reports. The following topics were defined as ones that would be explored in Study 1:

- Personal data types that are collected and used in recommendation algorithms;
- Inferences that platform might make to feed into recommendation algorithms;
- The goals of recommendation algorithms and who benefits from them;

- Aspects that impact recommendation algorithms, including how algorithms are being developed, trained and tested (variables, biases, etc.);
- The ways in which recommendation algorithms impact individuals' experiences; and
- Aspects of control and choice that users have or should have with recommendation algorithms.

We kept these topics in mind in the process of structuring the procedure for Study 1, and attempted to include them through a range of design research activities.

Study 1

PARTICIPANTS

There were 30 participants who took part in the study, a sample that was sufficient to reveal patterns in qualitative interviews and activities. They were recruited via "Prolific," an online research platform, and were compensated $20 for their participation in Study 1. In order to fit the needs of the study, we screened participants according to the following criteria: (1) Participants had to be fluent in English, (2) to have access to a computer (as we used a digital whiteboard software, "Mural," to conduct the study), and (3) to be willing to participate in one-on-one virtual calls. For quality assurance, we also screened for 90% and up approval rate on Prolific, and at least 30 prior submissions on Prolific.

To ensure diversity in our sample, we asked for participants' demographic information, which included age range, gender, race, disability, education and political orientation. We also asked them to answer two short 4-item and 3-item scales to control for attitudes towards technology (M=5.6 of 7) and attitudes towards privacy (M=4.45 of 7), adapted from Burbach et al. (Burbach et al., 2019).

In the study, 15 participants identified as female, 14 identified as male, and one as non-binary. A total of seven participants (about 23%) identified as people of color. Seven participants reported being politically right-leaning, 20 were left-leaning, and three selected "neither." Three participants (a total of 10%) reported having a disability. Eight participants were 18-29 years old, 11 were 30-44, six were 45-64, and five were 65+. One participant had high school education, four had some college education, 19 had a college degree, five had a Master's degree, and one had a Ph.D degree.

PROCEDURE

We piloted the designed procedure internally and externally with five pilot participants (three internal participants, two external participants) who were not involved in the design of the study. Three of the participants had little to no background on the topic. The pilot was intended to ensure that the topics are understandable and legible, and that the activity works well for our goals. Based on the pilot, we iterated on the activities' descriptions, titles, and layout.

Activity 1: Sensitizing Activity

We began the study with a *sensitizing activity*. Prior research indicated the importance of beginning co-design sessions with sensitizing activities, in order to introduce the topic of the study, as well as to emerge participants in the topic through a familiar, personal experience (Alvarado et al., 2020). Further, our goal was to have participants reflect on the *reporting* aspect of recommendation algorithms. To make that the focus, we used the sensitizing activity to minimize reactions to the practice of data collection and usage itself. In other words, the goal of the sensitizing activity was also to shift the focus from *whether an algorithm should be used in the first place* to *how platforms should report on the different aspects of algorithms that* ***are*** *being used*.

For the sensitizing activity, participants were asked which social media platform they use the most. Once determined, they were asked to enter the platform, either using their phone or computer (whichever they usually use or prefer). Once on the platform, participants were asked to describe the types of content they see (not the content itself—e.g., "an ad about furniture," as opposed to "an ad by Wayfair that promotes their Labor Day sale"). We followed up with several questions that intended to explore the topic of recommendation systems and personalization, such as why they think they see a particular piece of content, and how they think that platform determines what to show them. Most participants brought up recommendation algorithms themselves. If they hadn't, we used the term "automated curation" to avoid using the word "algorithm" with participants who were less technologically savvy. In either case, we laid out to participants that the goal of the study is to explore "how platforms decide what to show you when you use their platform."

Activity 2: Card Sorting

Card sorting is a common design research practice that asks people to think through a particular topic by sorting cards—they are asked to "think-out-loud" as they do so, and the cards serve as a tangible expression of topics and ideas that are easier to reflect on (Wood & Wood, 2008).

We conducted a cart sorting activity using a "Mural" digital whiteboard. Each participant had their own whiteboard set up for the study, and did not have access to other participants' activity.

In the study participants were asked to organize cards on a scale from "most important for me to know about," and "least important for me to know about" with three "decks" of cards: aspects of automated curation on social media, personal data types, and inference types. The topics in the card decks originated in prior research (Koene et al., 2019, Schelenz et al., 2020), and were iterated and extended by our team based on the topics that seemed important to explore from a tech policy standpoint. Below we list the three decks, and the cards that were included in each one:

1. Aspects of Automated Curation on Social Media
    a. What automated curation systems are being used?
    b. What "inferences" are being made about me for automated curation?
    c. What kinds of steps are being taken to correct errors in automated curation?
    d. Do I have a choice or the ability to impact automated curation?
    e. Where is there no automated curation involved?
    f. What is the goal of an automated curation system?
    g. How does automated curation impact my experience?
    h. Who benefits from automated curation of feeds?
    i. What are the implications of automated "inferences" about me?
    j. How accurate are "inferences," and what kind of errors are there?
    k. What content do I NOT see due to automated curation?
    l. What content that I post is not seen BY OTHERS due to automated curation?
    m. What personal data is being collected and used in automated curation?
    n. How was the automated curation system developed and tested?
2. Personal Data Types
    a. Personal identifying data (name, email address, phone, address, etc.)
    b. Personal physical identifying data (facial recognition, fingerprint, body features etc.)
    c. Personal identity (gender, race, sexual orientation, disability)
    d. Personal interests (hobbies, books, tv, media)
    e. Digital behaviors (web browsing, shopping, mouse movement, clicks, duration)
    f. Digital public social behavior (social media posts, comments, etc.)
    g. Digital private social behavior (texting, private messaging)
    h. Political affiliation / orientation
    i. Social circles (family members, friends, colleagues)
    j. Social-economic status (class, income, assets, loans, etc.)
    k. Physical behavior (location, offline purchases)
    l. Medical info (risks, treatment, appointments)

3. Inferences Made about You
    a. Personal identity (gender, race, sexual orientation, disability)
    b. Personal interests (hobbies, books, tv, media)
    c. Social circles (family members, friends, colleagues)
    d. Political affiliation / orientation
    e. Social-economic status (class, income, assets, loans, etc.)
    f. Shopping behavior (things you might want, likely to buy, etc.)
    g. Physical behavior (where you might go, what you might do)
    h. Medical information (conditions, risks, treatments, appointments)

The goals of using card sorting in our study were to give participants:

- A tangible way to think about a range of topics related to recommendation algorithms;
- A way to help participants physically prioritize aspects they care about most; and
- A prompt for discussion about recommendation algorithms and their transparency, and a window of opportunity to explore additional topics of interest and importance.

The activity began with showing participants the topic cards for "Aspects of automated curation on social media" (see Figure 11). They were asked to organize the cards on the scale between "most important for me to know about" and "least important for me to know about", while "thinking out loud" as they do so.

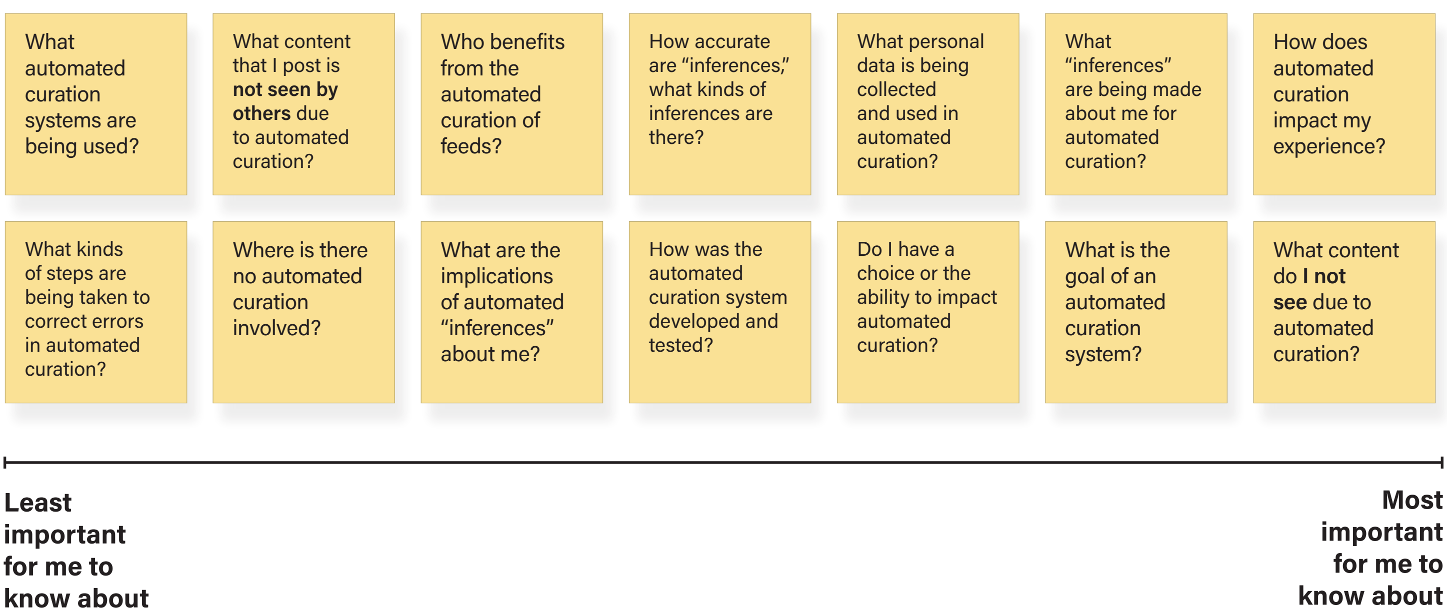


Figure 11. Aspects of Algorithmic Transparency Cards. The figure illustrated the cards that were given to participants, that laid out a range of topics related to recommendation algorithms (also referred to as automated curation). They were asked to organize these cards on a scale between aspects they would like to know about most, and those they feel are least important for them to know about.

Source - CDT

After they finished organizing the cards on the digital whiteboard (see Figure 12), we asked several follow up questions about the way that they organized the cards, such as "Why did you place x as most important?" and "Why do you think that y is least important for you to know about?" We used a semi-structured interview approach for this activity—an interview that was based on predefined questions, with follow up questions based on the conversation.

To note, the scale that participants organized the cards on was not intended to provide a quantitative measure of people's ratings of more or less important topics, but an estimated evaluation. Primarily, the organization of cards was a prompt for conversation, and intended to extract people's insights about the topics of interest. Because of that, in our analysis we did not analyze the graphs themselves, but conducted a qualitative analysis of interview responses.

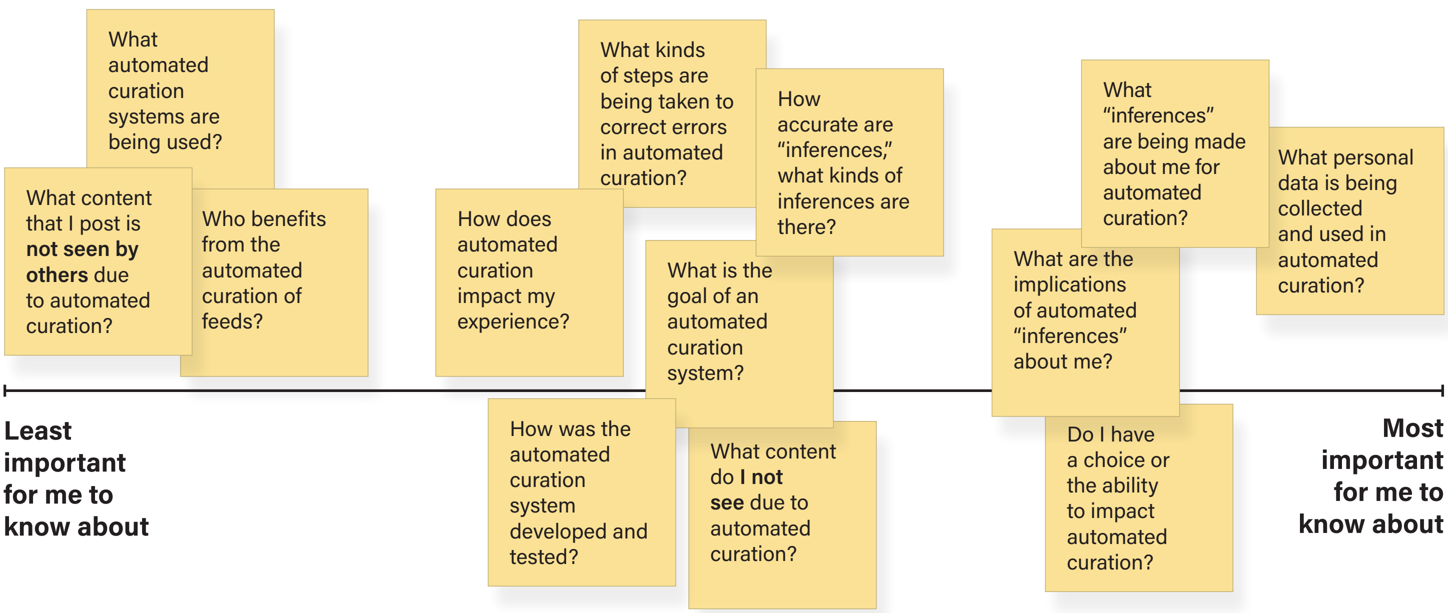


Figure 12. Organized Aspects of Algorithmic Transparency. This figure presents an example of how a participant organized recommendation algorithm-related topics using our Card Sorting methodology.

Source - CDT

Activity 3: Sketching a Report

In the third activity, participants were asked to imagine a scenario in which the main platform that they use for social media sends them a report about how their recommendation algorithm works. Participants were asked, in such a case, how would this report ideally look from their perspective?

For this activity, we created an area of the Mural whiteboard that had an empty page with a general title, and several graphical elements for participants to use in their sketch by "dragging and dropping" (see illustration in Figure 13). This was intended to support their creative thinking without learning the whiteboard software.

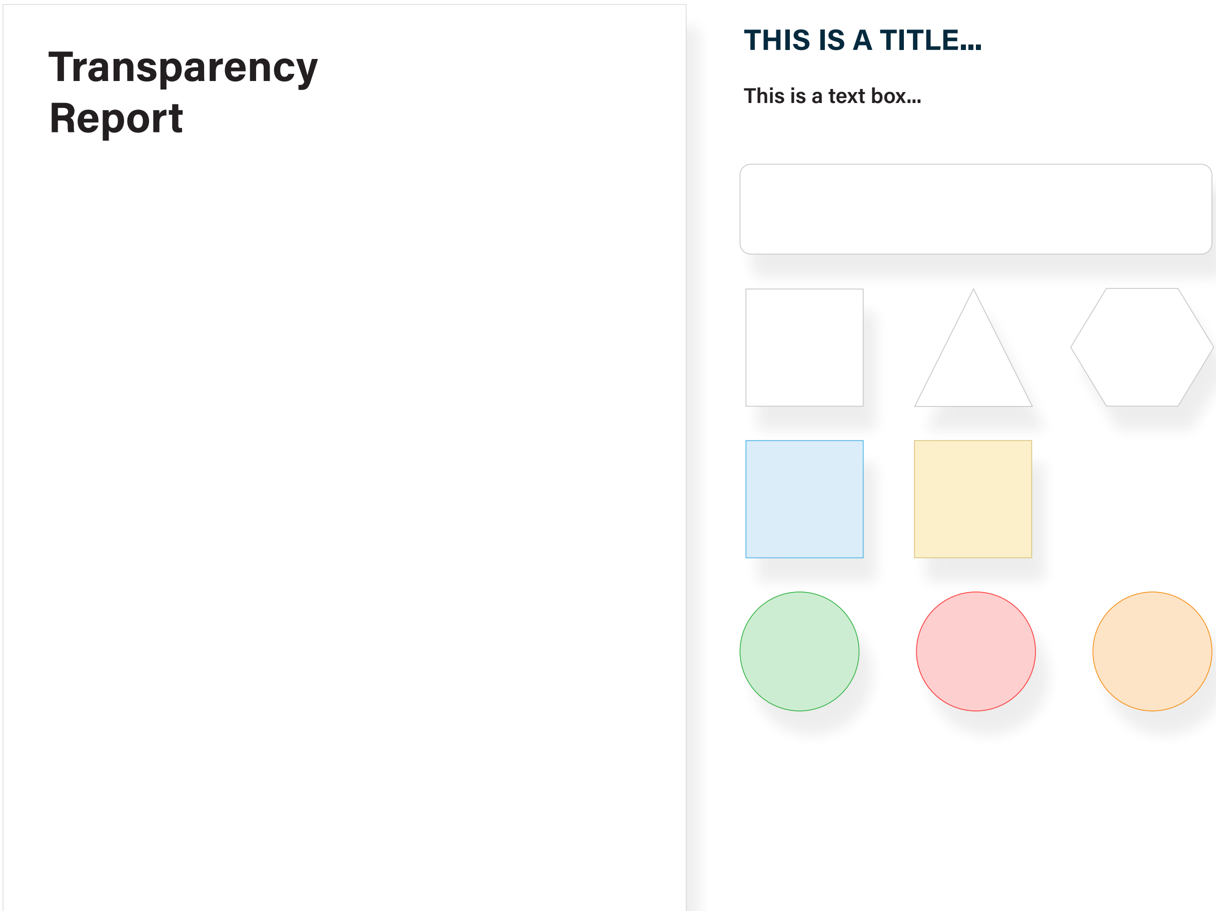
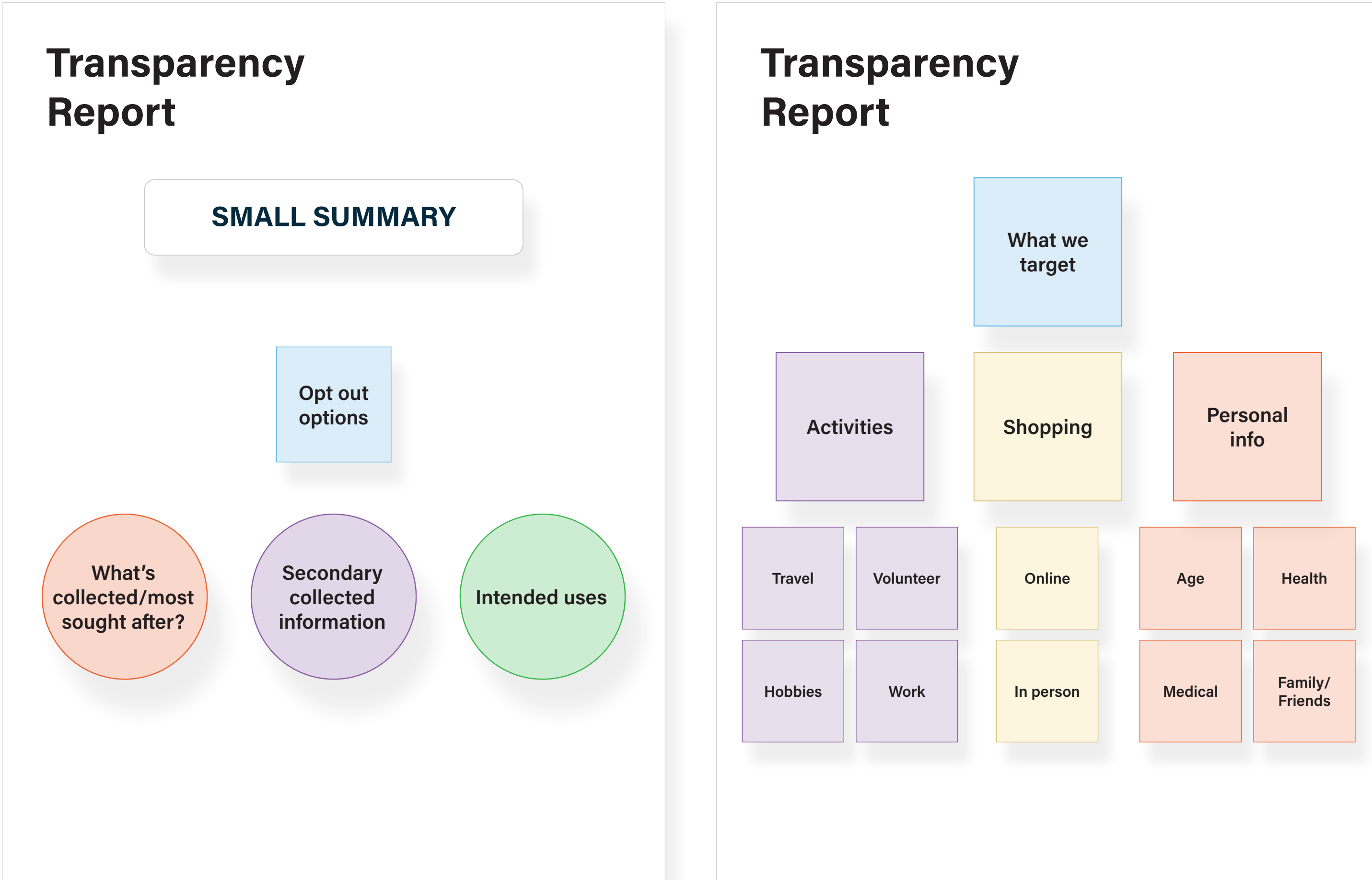


Figure 13. Transparency Report Sketching Activity. This figure illustrates the setup given to participants on the Mural whiteboard, an empty page (left) with some graphical elements to use (right). Participants were asked to sketch an ideal version of a recommendation algorithm transparency report *for them*.

Source - CDT

After giving participants instructions, the facilitator of the session “stepped away” from the virtual call by turning off their camera and microphone for five minutes. This was done to allow participants a moment to think through the prompt of what they would want to see in an algorithmic transparency report. Then, the facilitator returned and asked participants to describe and explain the sketch that they had come up with.

The goal of this sketching activity was to provide participants with another, more creative, outlet to express some of their values, needs, and desires regarding a hypothetical social media transparency report. For illustrations of the kinds of sketches some participants produced, see Figure 14.



▲ **Figure 14. Illustrations of Participant Sketches.** This figure illustrates examples of some of the sketches participants made in the sketching activity of Study 1.

Source - CDT

Activity 4: Excerpts from Social Media Privacy Policies

In this last activity, after participants gave the topic of recommendation algorithm reporting some thought, we asked them to look at text excerpts related to recommendation algorithms that were taken from current social media platforms' Privacy Policies pages (see Figure 15). The goal of this activity was to identify how participants perceive some current reporting practices: what is being done right, what doesn't work, and what adaptations and changes would benefit users in future algorithmic transparency reports. We therefore used a semi-structured interviewing approach here as well, with questions such as "What are your impressions of the information presented here?", "How useful or not useful do you think this information is?," and "Is there anything you would like to change about what is being shared in this text?"

Each participant viewed a total of three or four short two-to-five sentence excerpts, randomly selected from a predefined set of 21 excerpts. We included excerpts from Meta, Twitter, TikTok, Snapchat, Reddit, Linkedin, Discord, Google and Pinterest. Each was stripped of any information that would disclose which platform the text is taken from in order to avoid any prior preconceptions about a platform and how they share information with users. Topics included policy on how content is selected and presented to users, data collection for recommendation algorithms, how this data is used, usage of location information, usage of private message information, ad generation, and user control capabilities.

Analysis

The data from all four activities was transcribed and analyzed qualitatively using systematic thematic analysis (Braun & Clarke, 2006). Through an inductive process, qualitative codes were created based on 50% of the data, when theme saturation was reached. The second half of the data was coded deductively, and the first half was then recoded based on the final codebook. The sketches in the visual sketching activity were also included in the analysis—we screen-captured and analyzed them using qualitative graphical annotation (Bowers, 2012), where the researcher annotated the graphical qualities of each sketch (referring to layout, content, structure, etc.).

▼ **Figure 15. Excerpts from Privacy Policies.** This figure illustrates the excerpts that were presented to participants in the last activity of Study 1. All excerpts, include the ones presented here, were taken from a range of social media platforms' Privacy Policy pages.

Source - CDT

How data is used

We use information you provide to us and data we receive, including Log Data and data from third parties, to make inferences like what topics you may be interested in, how old you are, and what languages you speak. This helps us better promote and design our services for you and personalize the content we show you, including ads.

Private Messaging

When you interact with our services, we collect information that you provide us.

[...]

Of course you'll also provide us with whatever information you send through our services, such as [photos] and [messages]. Keep in mind that the users who view your [photos], [messages], and any other content can always save that content or copy it outside the app. So, the same common sense that applies to the internet at large applies to our services as well: Don't send messages or share content that you wouldn't want someone to save or share.

Design Process

In this research project we set out not only to understand people's perspectives and needs, but also to explore possible implementations of our findings in tangible prototypes of future algorithmic transparency reports. Thus, we (1) identified weak signals in the findings of Study 1, and (2) relied on these identified weak signals to develop interface prototypes that communicated and tested new design directions for future social media recommendation algorithm transparency reports.

IDENTIFICATION OF WEAK SIGNALS

The first part of the ideation process for creating prototypes was to identify weak signals in the findings of Study 1, signals that can serve as the foundation for creating prototypes for Study 2. "Weak signals" is a term used across business, futures and design communities to indicate an early sign of interest that currently has little or no impact, but that has the potential to have significant impact or change in the future (Hiltunen, 2006). Weak signals are an early indication of impact and change, and thus they may not yet be actionable, but should be considered for further exploration to fully understand. We focused on three weak signals that were most prominent in the findings of Study 1, and which served as the basis for our prototypes.

Weak Signal 1: Direct Connection Between Data, Why It's Collected and How It's Used in Recommendation Algorithms

When presented with data that was collected about the user as input for recommendation algorithms, participants expressed a desire to learn about the direct connection between the information that is collected about them to feed algorithms, and the purpose it serves ("*[platforms] say they collect [data] and give disclaimers, but don't really say why they need it or what it is for*" (P10)). This was even more salient when the input data were inferences made by the platform. In contrast to information that a user willingly hands out, participants believed platforms should make a stronger case for exactly why an inference is being drawn and what the platform plans to use it for.

Weak Signal 2: Sharing Data for Recommendation Algorithms To and From Third Parties

Our findings suggest that there is potentially a rich area for exploration in how platforms can illustrate their collaborations with third parties; which data platforms externally acquire to feed into their recommendation algorithms (and where from), and which data collected by a platform is later shared with third parties ("*I'm much more* *concerned about sharing of data with third parties. I can decide on an individual basis whether I trust a business or organization with my data. If they in turn, sell it on though, I would not even know about that or be able to decide whether I trusted that organization or company with the data. So a lot of this is about sharing for me.*" (P3); "*As long as [platforms are] not pulling information from outside sources, I'm okay with [it]*" (P27).

Weak Signal 3: Presenting Reports in a Personalized Way

While most participants believed that a recommendation algorithm transparency report will not likely be personalized, and does not have to be personalized, they still expressed interest in having the information shared be as tailored as possible ("*If we're going with the [transparency report] fantasy, then it would be specific to me. What do I click on that makes you share something with me? What kind of stuff do I interact with that would make you show me an ad, or make you recommend this person for me to follow?*" (P15)).

PROTOTYPING

Once we had a list of weak signals and possible interactions they could lead to in an algorithmic transparency report, two design researchers on the team brainstormed and sketched ideas based on each signal, while frequently referring back to all of the findings from Study 1. In addition, we conducted a design research audit of transparency pages and disclosures along with policy pages from Twitter (About Twitter's Account Suggestions, n.d.), Meta (Types of Content We Demote | Transparency Center, n.d.) and Instagram (Mosseri, 2021) along with dashboards inspired by Google Analytics (Google Analytics, n.d.) and Twitter Analytics (Twitter Analytics, n.d.).

When creating the prototypes and including data visualization for the various aspects that participants indicated, we relied on data visualization theory, such as Johanna Drucker's work about how graphic design and critical literacies enhance understanding of data (Drucker, 2014): "*The task of making knowledge visible does not depend on an assumption that images represent things in the world. Graphics make and construct knowledge in a direct and primary way. Most information visualizations are acts of interpretation masquerading as presentation. In other words, they are images that act as if they are just showing us what is, but in actuality, they are argument made in graphical form*" (Drucker, 2014).

Other inspirations that guided our design have been user experience knowledge and best practices, such as Tufte's principles of scale, where scale has been shown or rendered within an image or visualization for scientific evidence and purposes (Tufte, 1985). In our case, it was important to render the prescription of scale of one's data, along with how that data was analyzed, grouped, inferred and then used for decision making to create understanding for the user, especially when representing new material. To create understanding and readable or legible data, the user has to understand how the data is represented, and how that relates to what is being visualized. In this case, because we were visualizing a user's practical data, every section included some aspect of personal data that the user would recognize and could use as a logical 'measuring stick' to then understand the rest of the visualizations.

After several iterations on the brainstormed ideas that resulted in initial graphical representations of these interactions, a graphical designer created designed versions of the experience wireframes. These designed versions were created to be presented to participants. As part of the co-design process, the goal was to allow participants to re-evaluate their own ideas, this time in tangible form, and to reflect on the tradeoffs that emerged in the implemented version of some of their hypothetical ideas.

Study 2

PARTICIPANTS

As is common in co-design processes, we recruited participants for Study 2 from the pool of participants who participated in Study 1 because they had already engaged with the topic and reflected on it. So, a second session potentially allows them to deepen their thinking about the topic by re-examining their own reflections using tangible prototypes.

All participants who participated in Study 1 were invited to participate in Study 2 using the Prolific platform. Of the 30 participants who were invited, a total of 16 participants participated in Study 2. They were compensated an additional $20 for their participation in Study 2. In Study 2, seven participants identified as female, eight identified as male, and one as non-binary. A total of five participants (about 31%) were people of color. Five participants reported being politically right-leaning, ten were left-leaning, and two selected "neither." One participant reported having a disability. four were between 18-29 years old, eight were 30-44 years old, two were 45-64 years old, and two were 65+ years old. One had high school education, three had some college education, nine completed their college education, two had a Master's degree, and one had a Ph.D.

PROCEDURE

We used a standard Experience Prototyping methodology for Study 2 (Buchenau & Fulton, 2000). Experience prototyping is a method that presents prototypes of hypothetical user interfaces that people might experience in the future, and asks them to reflect on what they see. All prototypes were placed on a "Mural" digital whiteboard. During the interview, participants were invited to join the whiteboard, and to examine the four designed prototypes, one at a time (in a consistent, logical order, as described in the report). They were asked to share their initial responses to each prototype, and to attempt to "think out loud" as they look through them. After the open ended prompt, we continued with our semi-structured interview about the prototypes, that included questions such as: "What is your impression of the personalized aspect of this report?" and "What do you think about the granularity of the information presented here?"

During the Experience Prototyping sessions, we let participants know that the interviewer did not design the prototypes, and that CDT is an independent, non-partisan organization. The goal was to encourage a more objective response in which participants were not worried about hurting the interviewer's feelings, or to accommodate what they think may be the study expectations. Interviews were audio recorded, transcribed and analyzed qualitatively using thematic coding, similarly to our analysis approach of Study 1. All the finalized prototypes can be found in a corrosponding online repository on the CDT website.

References

Allan, A. (2022). A Series on the EU Digital Services Act: Due Diligence in Content Moderation. *Center for Democracy and Technology*. [https://perma.cc/6AYE-P9AT]

Alvarado, O., Storms, E., Geerts, D., & Verbert, K. (2020). Foregrounding Algorithms: Preparing Users for Co-design with Sensitizing Activities. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–7. [https://perma.cc/JYC9-6PDX]

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. [https://perma.cc/K586-E3H2]

Antonie, A. P. I. (2022, August 15). Hiding OUT: A Case for Queer Experiences Informing Data Privacy Laws. Public Knowledge. [perma.cc/A4NM-C6QD]

Asian American Disinformation Table. (2022). *Power, Platforms, Politics: Asian Americans and Disinformation Landscape Report*. [https://perma.cc/97P7-BZTX]

Bechmann, A., & Nielbo, K. L. (2018). Are We Exposed to the Same "News" in the News Feed? *Digital Journalism*, 6(8), 990–1002. [https://perma.cc/W8J7-KMA8]

Boczkowski, P. J., Mitchelstein, E., & Matassi, M. (2018). "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media & Society*, 20(10), 3523–3539. [https://perma.cc/SJ5K-NK2E]

Bowers, J. (2012). The logic of annotated portfolios: Communicating the value of "research through design." *Proceedings of the Designing Interactive Systems Conference*, 68-77. [https://perma.cc/AQB6-LFHQ]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology* 3(2): 77-101. [https://perma.cc/FX9D-8Y65]

Buchenau, M., & Suri, J. (2000). Experience Prototyping. *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. [https://perma.cc/UF6N-MBEF]

Burbach, L., Lidynia, C., Brauner, P., & Ziefle, M. (2019). Data protectors, benefit maximizers, or facts enthusiasts: Identifying user profiles for life-logging technologies. *Computers in Human Behavior*, 99, 9–21. [https://perma.cc/Y24J-2AV2]

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512. [https://perma.cc/2R6S-QHCS]

Chen, J., Geyer, W., Dugan, C., Muller, M., & Guy, I. (2009, April). Make new friends, but keep the old: recommending people on social networking sites. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 201-210). [https://perma.cc/C43X-ZAJD]

Chowdhury, R., & Williams, J. (2021). Introducing Twitter's first algorithmic bias bounty challenge. *Twitter*. [https://perma.cc/ULJ2-MEE4]

Davis, J., & Aggarwal, V. A. (2020, July 21). How Spotify and TikTok Beat Their Copycats. *Harvard Business Review*. [https://perma.cc/J74E-5DC8]

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. [https://perma.cc/8TSS-XXPQ]

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press. [https://perma.cc/AJ4T-UU2H]

Drucker, J. (2014). Graphesis: Visual forms of knowledge production. *Harvard University Press Cambridge, MA*. [https://perma.cc/B3EC-84QU]

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding Explainability: Towards Social Transparency in AI Systems. *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). Association for Computing Machinery. [https://perma.cc/B6QZ-M2KJ]

Elsesser, K. (2021). Here's How Instagram Harms Young Women According To Research. *Forbes*. [https://perma.cc/39B9-5NPQ]

Friendly, M. (2006). A Brief History of Data Visualization. *In Handbook of data visualization* (pp. 15-56). Springer, Berlin, Heidelberg. [https://perma.cc/KPA7-GW9C]

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for Datasets. *arXiv*. [https://perma.cc/PG6X-TH2Q]

Golembewski, M., & Selby, M. (2010). Ideation decks: A card-based design ideation tool. *Proceedings of the 8th ACM Conference on Designing Interactive Systems - DIS '10*, 89. [https://perma.cc/6CZW-ZS8A]

Google. (n.d.). Google Analytics. Retrieved September 14, 2022 [https://perma.cc/TZR5-2DPK]

Green, N., Procope, C., Cheema, A., & Adediji, A. (2021). System Cards, a new resource for understanding how AI systems work. *Meta AI*. [https://perma.cc/6Z89-KT83]

Haimson, O. L., Delmonaco, D., Nie, P., & Wegner, A. (2021). Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 466:1-466:35. [https://perma.cc/HT7W-F7LW]

Hiltunen, E. (2006). Was it a wild card or just our blindness to gradual change. *Journal of Futures Studies*, 11(2), 61-74. [https://perma.cc/3QT8-K4KS]

Hsu, T. (2018, December 17). Voter Suppression and Racial Targeting: In Facebook's and Twitter's Words. *The New York Times*. [https://perma.cc/UR6H-VE7Q]

Hughes, B. (2022, March 22). Food Businesses Lose Faith in Instagram After Algorithm Changes. *The New York Times*. [https://perma.cc/L2EE-7KFQ]

Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency* (Panel for the Future of Science and Technology, p. 124). European Parliamentary Research Services. [https://perma.cc/F8RQ-3Z9X]

Leetaru, K. (2018). The Data Brokers So Powerful Even Facebook Bought Their Data—But They Got Me Wildly Wrong. *Forbes*. [https://perma.cc/2RNH-U94N]

Lipton, Z. C. (2017). The Mythos of Model Interpretability (arXiv:1606.03490). *arXiv*. [https://perma.cc/HG4M-44M7]

Llanso, E. (2021). Santa Clara Principles 2.0: Civil society recommendations for how companies, states should protect free expression rights online. *Center for Democracy and Technology.* [https://perma.cc/E9VF-SDNB]

Llansó, E., & Morgan, S. (2014, November 5). Getting Specific About Transparency, Privacy, and Free Expression Online. *Center for Democracy and Technology.* [https://perma.cc/3LVQ-2ZUJ]

Lupton, D. (2016). The Quantified Self. *John Wiley & Sons.* [https://perma.cc/QRE7-JM5J]

Luria, M., & Candy, S. (2022). Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. *CHI Conference on Human Factors in Computing Systems,* 1–13. [https://perma.cc/SSW8-NP69]

Meta. (n.d.). *Types of content we demote | Transparency Center*. Retrieved September 14, 2022. [https://perma.cc/P2M7-JLJL]

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. [https://perma.cc/5G8N-F2WY]

Mosseri, A. (2021, June 8). *Shedding More Light on How Instagram Works.* Instagram. [https://perma.cc/3WPV-5Q8M]

Newton, C. (2017, February 7). Twitter begins filtering abusive tweets out of your replies. *The Verge.* [https://perma.cc/YL9P-62N6]

Nicholas, G. (2020). Explaining Algorithmic Decisions. *Georgetown Law Technology Review*, 4(711), 20. [https://perma.cc/S7BJ-JKKW]

Nicholas, G. (2022). *Shedding Light on Shadowbanning.* Center for Democracy & Technology. [https://perma.cc/4HGN-U7VX]

Oliva, T. D., Marcelo, A. D., & Gomes, A. (2021). Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture*, 25(2), 700–732. [https://perma.cc/57MX-5PCW]

Patil Swati, P., Pawar, B. V., & Patil Ajay, S. (2013). Search Engine Optimization: A Study. *Research Journal of Computer and Information Technology Sciences*, 1(1), 10–13. [https://perma.cc/5T4Y-ZC4W]

People + AI Research (PAIR). AI Explorables. (n.d.). Retrieved August 24, 2022. [https://perma.cc/C87Y-MCWN]

Poell, T., Van, J., & During, D. (2015). Social media and activist communication. *The Routledge Companion to Alternative and Community Media, London: Routledge.*, 527–537. [https://perma.cc/YMB4-Z3TV]

Prince, C. (2018). Do consumers want to control their personal data? Empirical evidence. *International Journal of Human-Computer Studies*, 110, 21–32. [https://perma.cc/NL65-97NT]

Richards, J., Piorkowski, D., Hind, M., Houde, S., & Mojsilović, A. (2020). A Methodology for Creating AI FactSheets. *arXiv*. [https://perma.cc/X9JU-8RVJ]

Schelenz, L., Segal, A., & Gal, K. (2020). Best Practices for Transparency in Machine Generated Personalization. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 23–28. [https://perma.cc/C3U4-JTXQ]

Shenkman, C., Franklin, S. B., Nojeim, G., & Thakur, D. (2021). *Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers. Center for Democracy & Technology.* [https://perma.cc/V7G6-R9HB]

Singh, S., & Bankston, K. (2018). The Transparency Reporting Toolkit: Content Takedown Reporting. *New America.* [https://perma.cc/EP3P-9QPK]

Sokol, K., & Flach, P. (2020). Explainability fact sheets: A framework for systematic assessment of explainable approaches. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 56–67. [https://perma.cc/5D2B-2CN2]

Steele, B. (2021). Spotify's 2021 Wrapped is here to chronicle your year in music. *Engadget*. [https://perma.cc/E6J3-JXA4]

Steen, M. (2013). Co-Design as a Process of Joint Inquiry and Imagination. *Design Issues*, 29(2), 16–28. [https://perma.cc/T9GP-PNRZ]

Stern, J. (2022). You Can Now View Your Instagram Feed in Chronological Order. *WSJ.* [https://perma.cc/B3T9-6T6R]

Szymielewicz, K., Bacciarelli, A., Hidvegi, F., Foryciarz, A., Pénicaud, S., & Spielkamp, M. (2020). Where do algorithmic accountability and explainability frameworks take us in the real world? From theory to practice. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 689. [https://perma.cc/T36R-RZ5Q]

TensorFlow. (n.d.). *Model Understanding with the What-If Tool Dashboard* | TensorBoard. [https://perma.cc/63NL-MZ3A]fg

Tintarev, N., & Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. *2007 IEEE 23rd International Conference on Data Engineering Workshop*, 801–810. [https://perma.cc/GJW9-XAZ7]

Transparency Reporting Index—Access Now's global database. (n.d.). *Access Now*. [https://perma.cc/749W-E9PK]

Tufte, E. R. (1985). The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3), 15. [https://perma.cc/M8XF-C4WB]

Twitter. *About Twitter's account suggestions*. (n.d.). [https://perma.cc/4RKS-6SPR]

Twitter. *Twitter Analytics.* (n.d.). Retrieved September 14, 2022. [https://perma.cc/4BFY-2V5T]

Ustun, B., Spangher, A., & Liu, Y. (2019). Actionable Recourse in Linear Classification. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 10–19. [https://perma.cc/ZX98-FAKT]

Valencia, S., Luria, M., Pavel, A., Bigham, J. P., & Admoni, H. (2021). Co-designing Socially Assistive Sidekicks for Motion-based AAC. P*roceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 24–33. [https://perma.cc/BBT5-PW6P]

Vogus, C. (2022). *Improving Researcher Access to Digital Data: A Workshop Report.* Center for Democracy & Technology. [https://perma.cc/T8C4-YX5Z]

Vogus, C., & Llansó, E. (2021). *Making Transparency Meaningful: A Framework for Policymakers.* Center for Democracy & Technology. [https://perma.cc/R3NH-U4XS]

Wall Street Journal. (2021). *Investigation: How TikTok's Algorithm Figures Out Your Deepest Desires.* WSJ. [https://perma.cc/6LGS-PLCP]

Watson, J., Lipford, H. R., & Besmer, A. (2015). Mapping User Preference to Privacy Default Settings. *ACM Transactions on Computer-Human Interaction*, 22(6), 1–20. [https://perma.cc/9Y99-VFQL]

Westerlund, B., Lindquist, S., Sundblad, Y., & Mackay, W. (2003). Co-designing methods for designing with and for families. *The 5th European Academy Of Design Conference, Barcelona, 28–30 April 2003.* [https://perma.cc/SQP6-W2VB]

Wood, J. R., & Wood, L. E. (2008). Card Sorting: Current Practices and Beyond. *Journal of Usability Studies.* [https://perma.cc/273U-9J9Q]

Woolery, L., Budish, R. H., & Bankston, K. (2016). *The transparency reporting Toolkit: Best practices for reporting on US government requests for user information.* The Berkman Klein Center for Internet & Society. [https://perma.cc/HX7B-FL6L]

Yang, M. (2021). Explaining "Explainability." Center for Democracy and Technology. [https://perma.cc/9EHZ-E82H]

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

CENTER FOR
DEMOCRACY
& TECHNOLOGY