Not Just Algorithms

Strategically Addressing Consumer Impacts in Information Retrieval

Ekstrand, Michael D.; Beattie, Lex; Pera, Maria Soledad; Cramer, Henriette

DOI
10.1007/978-3-031-56066-8_25

Publication date
2024

Document Version
Final published version

Published in
Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Proceedings

Citation (APA)
Ekstrand, M. D., Beattie, L., Pera, M. S., & Cramer, H. (2024). Not Just Algorithms: Strategically Addressing Consumer Impacts in Information Retrieval. In N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, & I. Ounis (Eds.), *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Proceedings* (pp. 314-335). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 14611 LNCS). Springer. https://doi.org/10.1007/978-3-031-56066-8_25

Not Just Algorithms: Strategically Addressing Consumer Impacts in Information Retrieval

Michael D. Ekstrand[1](✉), Lex Beattie[2], Maria Soledad Pera[3], and Henriette Cramer[1,2,3]

[1] Drexel University, Philadelphia, PA, USA
mdekstrand@drexel.edu
[2] Spotify, Seattle, USA
lex@spotify.com
[3] Delft University of Technology, Delft, NL, The Netherlands
m.s.pera@tudelft.nl

Abstract. Information Retrieval (IR) systems have a wide range of impacts on *consumers*. We offer maps to help identify goals IR systems could—or should—strive for, and guide the process of *scoping* how to gauge a wide range of consumer-side impacts and the possible interventions needed to address these effects. Grounded in prior work on scoping algorithmic impact efforts, our goal is to promote and facilitate research that (1) is grounded in impacts on information consumers, contextualizing these impacts in the broader landscape of positive and negative consumer experience; (2) takes a broad view of the possible means of changing or improving that impact, including non-technical interventions; and (3) uses operationalizations and strategies that are well-matched to the technical, social, ethical, legal, and other dimensions of the specific problem in question.

Keywords: users · consumers · impact · harm · equity

1 Introduction

Search engines, recommender systems, and related information retrieval (IR) tools are embedded in the daily lives of individuals on all paths of life, as they facilitate access to large and diverse collections, from articles to songs to products for purchase. IR research in its sociotechnical context, along with work on information seeking, recommender systems, and relevant segments of human-computer interaction (HCI), has long been concerned with ensuring that the systems provide efficient and effective information access to all who require it. It also examines the impacts of sociotechnical systems and implications for IR.

We explore the *impacts*—defined broadly—that IR systems have on *consumers* [20] (or users [67])and how those impacts differ between different (groups of) consumers. This draws from multiple perspectives, including:

Partly supported by the National Science Foundation on Grant 17-51278.

N. Goharian et al. (Eds.): ECIR 2024, LNCS 14611, pp. 314–335, 2024.
https://doi.org/10.1007/978-3-031-56066-8_25

(i) General IR research, which has long worked to providing resources that are well-matched to users' information needs in some or all of their nuances.
(ii) Consumer fairness [20,41] seeks to ensure that different consumers' experience is in some sense *fair* in its qualitative and/or quantitative dimensions [44], e.g., utility and representation.
(iii) Audience-specific IR looks to build systems that meet the differing needs of particular groups, such as children [80,90,91,106], autistic users [11,70,83], users experiencing dyslexia [18,50], or language learners [31,79].
(iv) Harm-aware IR, including content moderation actions (e.g., reducing exposure to content that is exploitative, or related to criminal or violent groups [105]) or industry interventions like "compassionate search" designs that redirect searches to resources developed in collaboration with experts [86].

Thoroughly exploring impacts and their challenges and possibilities is a complex and multifaceted project [41]. For example, a system under-representing women in search results about the best athletes of the century would impact consumers by giving them an inaccurate picture of the information space, besides depriving women of the visibility and attention afforded to their male counterparts. Historical and current datasets and baselines may however be absent to easily assess such impacts. The interweaving systems with oft-competing objectives making up modern IR systems additionally complicate how to control for impacts. For example, external audits have examined gender disparities in computational *advertising* for domains such as jobs [32]. Ali et al. [6] found that even if advertisers work to ensure demographic parity in ad reach, e.g., people of all genders see an ad for a well-paying job, the computational and marketplace dynamics of ad platforms can induce disparities in ad visibility. For another example, IR systems trained primarily on adults' interactions may not be equally useful (or beneficial) to all groups: a *child* searching for "Sven movie" is likely looking for *Frozen*, but the search engine might treat 'sven" as a misspelling for "seven" [69] and return results for *Se7en*, an R-rated film with "grisly afterviews of horrific and bizarre killings, and strong language" [59].

Beyond statistical properties of rankings, shifts in power in the larger context are a big concern [36] but this does not diminish within-system impacts: "attention" by being recommended can deliver power over time, and getting high(er) quality information or feeling represented in results can empower [72]. Similarly, not only do the resources returned by IR systems play a role in the system's impacts, but aesthetics, imagery, and UX writing signal whether a service welcomes user groups and feels designed with that particular group in mind [72].

There are many points (technical, institutional, or procedural) in an IR system where it is possible to *intervene* to correct or amplify a system's impacts.

[1] Fig. 1 shows the components of typical IR systems; each can be adjusted when impacts are observed to be inequitable, illegal, or misaligned with societal or business goals. Different segments of research emphasize different sites and strategies for intervention: for example, fairness literature typically emphasizes algorithmic interventions, with some attention to evaluation, while work on meeting particular user needs often focuses on user interface and user experience, with some work on algorithms. Other work takes a broader view, ranging from mapping different types of harms to studying different stages in the machine learning (ML) lifecycle. There is little explicit attention to the relative strengths and capabilities of different points and types of intervention, how to compare which may be better suited to address particular impacts, and integrate the wide research and practice fields.

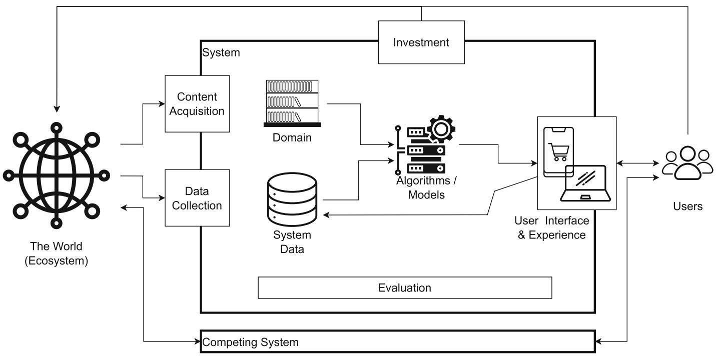

Fig. 1. The many components of IR systems and their sociotechnical context.

Effective impact intervention efforts must first identify a *goal* to pursue or problem to solve (Sect. 2), as the choice of *operationalization* (Sect. 3) and *intervention* (Sect. 4) depends on the goal [42]. Goals begin as high-level, theoretical, and qualitative objectives that are usually refined into quantitative measurements of impact—a process called *scoping* [14]—that are then matched to an appropriate technique to advance the original goal. To support this process, we provide three maps to guide and practice (summarized in Table 1): (1) a high-level overview of impact-related goals of IR-for-good efforts and tensions when considering these goals; (2) a review of operationalizations of those goals that bridge human concerns and technical possibilities; and (3) an overview of intervention points and strategies across the technical, institutional, and social aspects of an IR system in context. These maps are *non-exhaustive*, as further research will likely uncover more possible goals and strategies; our aims are to

[1] We borrow the concept of *intervention* from public health, behavioral health, and education to concisely express the idea of taking action to modify a system or its environment in order to address a problem (or enhance a positive phenomena). While this language is not widely used in IR research, we propose that it is useful for discussing how changes in a system's operation and outcomes can be effected.

provide a basis to contextualize existing work and to promote creative, interdisciplinary thinking about how to advance the benefits of IR and avoid or mitigate its potential harms, while more clearly integrating this rapidly growing field. For brevity and clarity, we focus this paper primarily on consumer-side impacts, but both the scoping process and the maps we discuss are fully applicable to impacts for other stakeholders such as providers and subjects.

We hope to see theoretical and applied work on IR impacts that: (1) **grounds** the work in specific impact goals (e.g. positive impact) or concerns (e.g. harm to avoid) that are clearly identified and well-motivated, (2) **contextualizes** those impacts with regards to the varied ways an IR system can affect its consumers, (3) **thinks expansively** about the range of possible methods and intervention points for addressing that impact, and (4) **matches** the impact to a strategy that is appropriate given technical and organizational constraints and the specific ethical, legal, economic, and other dimensions of the impact concern.

Table 1. Examples of goals, operationalizations, and strategies. Effective impact work requires clear selection and matching across these dimensions.

Examples		
Goal (Individual/group/society)	Operalization	Strategy
Recommender utility *e.g. aid discovery, or increase or diversify engagement, conversion and/or satisfaction* *Usability and access* *Representation* *Avoiding platform abuse and/or harmful content exposure*	*Metric objectives* *e.g. Equal utility* *e.g. Max grp log util* *Metric guardrails* *e.g. Set a max in engagement loss* *e.g. Reduce explicit negative user feedback* *Qual or Quant Data objectives*	*Auditing and monitoring* *Data pre-processing* *Data curation* *Feature selection* *Model enhancement* *Post-processing* *Engineering process* *Prioritization* *Evaluation* *New product development* *Design or Editorial intervention*

2 Scoping Stage 1: Goals

Our first map is a catalog of examples of impact-related goals for IR systems. The list is neither exhaustive nor a strict taxonomy; it is intended to spur expansive and contextualized thinking about what the different goals can mean in practice, and how we as a research community map the means to pursue them.

Mapping and evaluation for positive and negative impact can happen at multiple levels and may run into tensions. Defining positive goals to explicitly work towards is crucial to understanding whether the efforts involved in devising an IR system have the desired impact on individuals, organizations or wider society—and especially at what level this impact is expected to be seen. Safeguarding against negative impacts and identifying guardrails also requires definition. In practice, IR systems are often assemblages of multiple systems with different objectives, guardrails, and possibilities. For example, services like Instagram combine different models, where some are focused on new discoveries, whereas others serve up more familiar content, together assembled in a user-facing feed

[78]. Some single models use multi-objective optimization [98]. Mappings of goals then must account for such tensions, and how multiple, conflicting or mutually beneficial goals might impact operationalization and strategies. For example, Barocas et al. [12] divide potential negative impacts of ML into harms of *allocation* and *representation*; this delineation also applies to IR [28,41]. The former harms refer to withholding opportunities or resources, affecting the quality of service and information provided. The latter reinforce stigmas or stereotypes, for example over or underrepresentation of certain information or groups in results. Katzman et al. [64] point out that from operationalizations and measurements chosen, it can be unclear which harms were intended to be measured, thus advising to explicitly state the harm of interest to avoid ambiguity.

Defining Positive and Negative Impacts before Measurement. Before it can be measured, let alone improved, the impact needs to be clearly defined in its concept. IR has long sought the impact of *utility*, i.e., providing consumers with resources that are relevant to their needs. Utility is broad, however, and can involve various ways to meet the diversity of users' needs, e.g., through access to resources or increased engagement [23]. Diversity is a useful goal but needs to be defined more specifically to be meaningfully evaluated, e.g., subtopic diversity [93] considers the different things a query might mean, whereas calibration [104] aims for results that represent the breadth of users' interests, and other systems may aim to diversify interests or information provided [46,110]. Other work looks at utility being distributed fairly or equitably [45,73], to ensure the system is useful to all of its users instead of systematically under-serving some user groups (genders, ethnicities, etc.) Deldjoo et al. [35] observe in a thorough review that fairness research typically assumes that a clear definition is already available, "thus rendering the problem as one of designing algorithms to optimize a given metric"; this skips the explicit goal setting process.

Impacts go beyond resources and utility. *Representational harms* [30,64] to potentially assess include: inaccurately representing users' interests and information needs internally, preventing certain user groups from systematically having less-accurate representations (e.g. user embeddings or other user models that may lead to stereotyped recommendations [19]); perpetuating harmful or unnecessary stereotypes [85,94]; misrepresenting identities, e.g., misgendering; modeling users' interests by making assumptions based on demographic or other attributes instead of their preferences, i.e., "box products, not people" [97]; and misrepresenting the space of content, creators, subjects, etc. to consumers through the composition and presentation of results, particularly over- or under-representing certain types of content or people, or the resources and the results themselves [94]. In particular settings, specific representational harms may occur. For instance, Katzman et al. [64] further map different types of representational harms in image tagging.

Usability and Impact for Different Consumers. While IR systems may provide access to information, systems need to be usable in the first place; ensuring different groups can access and use the system is an important class of impact goals. Many designs assume a default user, typically an educated adult; such systems may not be usable for children [10], second-language learners [77], disabled users [70], people with low-bandwidth internet connections [84], or others who for whatever reason have difficulty using the default system. The system's results may harm either consumers generally or particular consumer groups in various ways as well. Self-harm advocacy [27], disinformation, and forms of deceptive or illegal material affect many users. Other content impacts differ from user to user. Milton and Pera [76] showcase how the subliminal stimulus presented by popular search engines varies throughout the information seeking process for searchers affected by mental health issues when compared to "typical" searchers. Information can also interact in often unpredictable ways with users' emotional state and context, such as continuing to recommend baby products when a consumer has suffered a pregnancy loss. Such content can both cause direct emotional harm and "erase" that such human experiences are not aberrations [8].

From Individual Consumers to Society. Impacts do not end with individuals, they may have a ripple effect on organizations, movements, and society. For example, Helberger [56] discuss the role of diverse IR results in supporting users' engagement in democratic processes. Decades ago, Belkin and Robertson [16] voiced the possibility for IR technology to manipulate users, particularly regarding ideological or political beliefs. Similarly, concerns have been raised about the potential for recommender systems (and personalized IR in general) to isolate individuals in different segments of the information space [75,87,109], or indulging or reinforcing beliefs that harm consumers' relationships with others [66]. The exact societal impacts around topics such as polarization are contested depending on specific study settings and effects; setting goals at different levels with different groups leads to different operationalizations. This concern is not limited to IR work; early work around "internet communication" (1996, [33]) observed that while much insight was provided by ongoing studies, differences in units of analysis lead to difficulty in cross-study comparisons as well as theoretical integration (see for a similar discussion around statistics related to patient outcomes in medicine [7]). Careful definition is essential.

3 Scoping Stage 2: Operationalization

Operationalization requires translating a theoretical construct into conceptualization that can be observed and assessed using a specific quantitative and/or qualitative research design to guide and evaluate impact efforts. This step requires careful creativity in choosing or designing metrics and protocols, as operationalization is highly dependent on the goal in all of its technical and sociotechnical nuance [60]. In this section, we present a series of constraints that govern the operationalization process and example operationalizations for three common measurement objectives, focusing on quantitative methods; the full reality of mapping the field and methods' impact is more complex. Patton [89], for example, discusses how methods from social work can be used to anticipate how AI and UX solutions impact diverse communities. In 1999, Sofaer [102] also provided a variety of examples of the usefulness of qualitative methods in health services and policy research in a similarly methodologically 'nascent' field.

3.1 Operationalization Constraints

Operationalization constraints are marked by the need to identify how to quantitatively address a qualitative goal within the confines of an auditing scenario, a critical step in many broad system-level frameworks for algorithmic impact measurement [14]. Below we share three possible constraints to address.

Consumer Properties. Identifying consumer properties enables us to understand what will be measured for the consumer, and how that characteristic will be measured. The primary property constraint should be to define if the consumer should be measured in a group or individual setting. Addressing this constraint is a requirement for most fairness evaluation settings due to the inherent difference in measuring group versus individual fairness [101]. Individual versus group measurement specificity can extend beyond fairness and should be accounted for when operationalizing one's goal to ensure that the measurement accounts for how a goal may differ when observed at the individual or group level. When measuring goals for individuals, operationalization should target quantifying if similar individuals receive similar treatment or outcomes. If measurement is defined to target consumers at the group level, the practitioner should further identify the group attributes for measurement. We can narrow types of group attributes into three categories: binary, multi-categorical, and intersectional. It is essential to understand the group attribute type due to the fundamental differences in measuring binary, multi-categorical, and intersectional scenarios.

System Components for Evaluation. Goals can be scoped into operationalized metrics at many different points in the IR system components (Fig. 1). While end-to-end measurement is always important [39], measurement at intermediate stages is often useful. Evaluation or optimization of components of a multi-stage recommender system, for example, may target the candidate generator, ranking model, or the end-to-end system. This view moves beyond the standard fair ML model stages of pre-processing, intrinsic, and post-processing algorithmic components. Targeting the goal to be assessed in a data collection component requires unique measurement objectives and metrics to address data-collection-specific concerns such as selection or sampling bias. Potential subsequent measurement objectives and metrics will be dependent on the algorithm, model outputs, and data available for evaluation; it is therefore critical to understand the system components and how their outputs can be quantified operationalizing a goal.

Measurement Objectives. Measurement objectives are initially addressed when defining goals, but operationalizing goals requires more specific and quantitative definition of objectives. For example, in past research, evaluation metrics are often categorized into distinct groups to provide guidance for how and when to use specific metrics. Smith et al. [101] demonstrate the importance of understanding measurement objectives for fairness related evaluation by surfacing specific fairness objectives and their subsequent relevant metrics to achieve the objective. In their work, utility for consumers is further defined by two measurement objectives: utility versus merit and group utility [101]. Utility versus merit defines the utility measurement objective of "utility for consumers" is distributed "based on their merit or need", while group utility looks to "distribute utility equally between groups of consumers" [101]. These two different measurement objectives require different metrics for operationalization.

3.2 Identifying Metrics Within Operationalization Constraints

The relevant metrics for operationalization depend on the choices made when identifying the consumer properties, system components, and measurement objectives, as well as the social, ethical, economic, and other particularities of the goal. Some guidance exists for identifying relevant metrics from the research literature in light of such constraints, such as the recommendation fairness decision framework introduced by Smith et al. [101]. Unfortunately, there are many applications and types of impact where such frameworks are not yet available, requiring metric design from scratch. Measurement modeling, as showcased by Jacobs and Wallach [60] for operationalizing fairness, provides a substantive framework for designing robust metrics for measuring qualitative goals or constructs. These metrics then enable the practitioner to assess the impact of an intervention on their original qualitative goal. Additionally, using the identified metrics to evaluate the system prior to intervention can guide the choice of intervention strategy and provide a baseline for measuring improvement. In the rest of this section, we present several examples of identifying metrics based on three potential measurement objectives and constraints.

Utility. Utility is a well-known and widely-studied goal for consumer impact. There are many metrics to estimate utility to consumers [54], but even once a metric is selected, there are several ways to operationalize *equity* of utility to ensure groups of consumers aren't being systematically under-served. One way is to compute the *difference* in utility between groups ($\Upsilon_{g_1} - \Upsilon_{g_2}$). Another is to compute the *total groupwise log utility* [$\sum_g \log \Upsilon_g$, 111], so maximum improvement for the overall metric comes from improving utility for the lowest-utility group. Targeting differences in utility [e.g. 58,68,81] can remove inequity, but treats utility as zero-sum, sometimes with significant majority-group utility loss [68]. A positive-sum aggregate that avoids unnecessary competition [111]; since consumer utility is non-subtractable (users do not compete for the same utility) [15,41], this is more appropriate. It also extends beyond binary groups.

Representation. *Calibrated "fairness"* is an example of operationalizing representation for measurement. This concept probes if users with multiple interests are represented "fairly" in the result set by correctly reflecting their historical interests [38]; it can also be used to measure how consumer behavior affects recommendations along various axes [25,43]. This is done at the individual or group level by comparing distributions between what is recommended versus users' history in the system. Relevant metrics can change based on their intended use.

For instance, distribution comparison metrics are most often used to evaluate calibrated fairness. Kullback-Leibler Divergence (KLD) lends itself well to evaluating calibration fairness of content-pool generation by appraising the served distribution of the entire final content pool. Normalized KLD is better suited for evaluating the distribution of a ranking component since it penalizes for rank during calculation. These metrics work well for evaluating and prioritizing interventions, but may not suit evaluation and optimization. In the case that one metric should be used for evaluation and intrinsic optimization of calibrated fairness, Jensen-Shannon Divergence may be favored since the symmetry of the metric lends itself better to intrinsically optimize calibration fairness.

Diversity. Vrijenhoek et al. [110] examine different operationalizations of the qualitative goal of diversity for news recommendation in the context of supporting democratic engagement. Drawing from different theories of democracy, they structure five measurement objectives with specific metric operationalizations like "fragmentation", which "denotes the amount of overlap between news story chains shown to different users" [110]. To further concretize this metric, they choose between two possible metrics Kendall Tau Rank Distance (KTRD) and Rank Biased Overlap (RBO) [110], selecting RBO to avoid measurement limitations of KTRD, which does not penalize by rank; thus making it a more suitable metric for evaluating candidate pools than final ranked lists for fragmentation [110]. However, RBO does not satisfy their original measurement objective of fragmentation, leading them to use an adaptation that subtracts RBO from one. The documentation of their process showcases the need to not only account for the system component but also the limitations of current metrics, resulting in a specially designed implementation for the measurement of the original objective.

4 Intervention Strategies

There is a range of strategies for advancing consumer impact goals and operationalizations, with interventions possible anywhere in the system as well as at many points in the broader sociotechnical context where the system operates (Fig. 1). These strategies can fall into one of these broad and non-mutually - exclusive categories: (1) **system interventions** to improve the impacts of an IR system (2) **process interventions** that change how an organization builds and maintains the system (3) **marketplace and ecosystem interventions** that introduce new products or intervene in the ecosystem that provides the IR system's inputs and consumes its output

. Some interventions directly employ an operationalization (Sect. 3) of an impact goal, e.g. as an objective function for training a model; others do not, but the operationalized goals are still crucial for identifying and evaluating the intervention. Effective interventions need to be well-matched to the particular goal and operationalization, as well as to the resources and constraints of the organization.

4.1 System Interventions

As we discuss in this section, decisions across the entire architecture and lifecycle of an IR systems affect the system's impact on people [29,112].

Data Interventions. Data is fundamental to IR systems; the corpus and the data used to train algorithms and models that power the system [108] affect its operation and impacts. Manipulating the data is then a viable strategy for adjusting system impacts. For example, injecting fake user profiles [96] and removing spam reviews [100] can improve recommender system impacts. For a summary of manipulation techniques at data pre-processing time to mitigate discrimination of protected groups in IR-powered disaster management applications see [114]. Chen et al. [26] examines different manifestations of bias in ML classifiers and distinguishes those best addressed through more or better data, and those more amenable to model-based interventions. Techniques to reduce position bias in CTR estimation through interventions [61] or re-analysis [5] may also apply to biases in user response that correlate with group membership.

Sometimes, data needs to be specifically collected or engineered. Allocating a budget for data improvement and prioritizing data needed to serve underserved groups can help. Goel and Faltings [52] provide a mechanism for crowdsourcing data while respecting fairness objectives in the data's coverage; this is a potential building block for data interventions. Interestingly, interventions designed to positively impact in one aspect can have unintended side effects. E.g., differentially-private training mechanisms can exacerbate data biases [47], making the evaluation of data strategies, and especially their combinations, crucial.

Algorithmic Interventions. Modifying the algorithms employed in an IR system is a common strategy for addressing impact. For example, much consumer fairness research augments the loss function an inter-user equity objective [58,111,113], sometimes through a regularization term [62,115]. Penalizing dependence between recommended items and user attributes [62] may be a viable alternative for reducing stereotypes. New fairness-aware metrics can be directly optimized [51,113]. Aggregations can be changed; using a maximin objective function in social network information spreading—instead of maximum or total spread—can reduce disparities in who has access to information [49].

There are many ways to modify algorithms besides adjusting the training objective, such as reranking [48,68], multi-objective optimization [74], and changing neighborhood selection [22]. Adversarial learning to learn user embeddings that cannot predict consumers' sensitive attributes such as gender [19] are promising both as a fair representation learning approach and to address potential stereotypes in recommendations. Existing techniques can be repurposed for indirectly improving consumer impacts. Diversification techniques can generate rankings that are fairer to a broader range of users [71]; diversification can support consumers in democratic participation under differing political theories [110]. MMR [23] has proven useful for fair image search [24,63]. Such indirect approaches have significant promise—existing computational machinery can likely be repurposed for a variety of impact-related aims.

Design or Editorial Interventions. The design of the user interface and experience presents many possibilities for addressing a system's impacts, particularly by better matching the system to users' specific needs and abilities. One example is Pinterest's inclusive search that combines both UX features and an inclusive model development process as well as editorial expertise to make the best of the strengths of both: "*Pinners can search for a broad hair term like 'summer hairstyles' [...] and narrow their results by selecting one of the six hair patterns[...]. Pinterest has detected a hair pattern (e.g. coily, curly, protective) in over 500 million images on our platform*" [92]. Deldjoo et al. [34] proposed a child-oriented TV/movie recommendation interface using tangible interaction: the child holds up a toy truck to get recommendations for shows about trucks.

Design can make systems or products more usable to a broader set of consumers; to provide interfaces tailored to the needs of particular classes of users; or to dynamically adapt the system and its interface for the current user's needs [e.g. 9,53]. Dynamic adaptation can serve children [37,69,99], or users of different physical abilities [88]. Adaptivity or new interaction designs are not without risk and potential negative impact, particularly as they may require additional data collection or inference. As with any strategy, they can lead to new types of abuse or inadvertent errors, and thus iterative goal setting is necessary.

4.2 Organization, Process, and Evaluation Interventions

Organizational dynamics are critical to consumer impact work in practice. Rakova et al. [95] identify four themes in organizational transitions towards work practices that include AI impact and responsibility: anticipating rather than reacting to issues, providing internal structure to address concerns, aligning on success, and resolving tensions within the organization. These are key to creating an organization where identifying, monitoring, and addressing consumer impacts is standard practice. Organizational structures and practices are a possible, yet often overlooked, point of intervention themselves. Changing *how* a system is built and evaluated will have further effects on *what* is built and its impacts.

Evaluation Interventions. Regularly auditing for scoped impact objectives and other forms of impact and equity [57] can identify problems and help detect regressions on past impact improvements. Other interventions also begin with evaluation; documenting and measuring system impact is crucial to informing the selection and design of any intervention strategy, as well as evaluating whether the chosen strategy is effectively achieving its objective. Further, the evaluation process itself can impact outcomes downstream and can be a site of intervention in its own right: changing how the system is evaluated, or how the evaluation is analyzed, affects both human and computational decisions in its ongoing development and maintenance. Disaggregated [13] and distributional [40] evaluations can identify inequities between consumers [45,73] and quantify broader consumer experiences; mediation analysis [73] can identify intervention targets for addressing inequities. While most commonly seen for assessing equity of utility, these can be applied to any measure of user experience.

Evaluation should go beyond quantitative measurements to include qualitative studies that note how consumers experience the system and what they want from it [55]. In practice, assessment is difficult, as the literature provides little guidance for navigating the decisions involved in designing an evaluation [14].

Engineering Process. The engineering process itself admits strategies for addressing impact inequities [57,95]. In every development cycle, product teams make decisions about what to prioritize and how to allocate resources; investing in efforts that address impacts or enhance the system for under-served consumers is a tool for improving impacts and equity. Evaluations that reveal *why* an impact occurs (e.g. mediation analysis) provide valuable inputs for planning processes.

4.3 (New) Product/Marketplace and Ecosystem Interventions

In some cases, impacts can best be addressed through **new systems** targeted at needs that are not well-met by existing systems. While this often is not adequate to address the negative impacts of existing IR systems, new possibilities for consumer equity emerge by thinking more broadly than a single system to consider whether people have equitable access to information across a set of product offerings or the broader marketplace and ecosystem of services. Many services implicitly or explicitly assume a "default" user and provide an experience that may not suit all consumers. Children are a key example; some services such as Netflix provide distinct services targeting specifically for children and/or families [82], while new startups are also trying to fill the gap between adult and child information access [1–3]. These new firms and their products can be seen as marketplace interventions: new products enter the market to meet needs that are not addressed by existing platforms, helping consumers directly and applying market pressure on existing firms to better meet those needs.

Other impact-related marketplace activities are possible, such as contracting with content providers to expand or improve the content that can be provided to consumers, or investing in local expertise to improve the metadata for content from a region or community. Tsioutsiouliklis et al. [107] attempted to improve the fairness of PageRank by recommending new links that, when included in the web's link graph, would result in more fair results. There are more opportunities to invest in a broad ecosystem in which IR systems can have better impacts, but for the research community it is especially important to invest in the mapping of what we do (not) know about the impact of different strategies in specific situations beyond the work of the publishing academic community.

5 Conclusion

Identifying, monitoring, and improving the impacts of for IR system requires *impact objectives* to be clearly scoped in a well-founded manner and matched to suitable intervention strategies, while accounting for the system's broader socio-technical impact. Landoni et al. [65] identify four aspects for evaluating technical advances: the advance itself (i.e., the intervention); the group of consumers it is intended to serve; the task those consumers perform; and the context or environment in which the advance will operate. These factors provide a starting point for reasoning about impact interventions for consumers and other stakeholders, and in this paper we have provided broad maps to help structure that reasoning.

Ensuring IR systems are equitable and beneficial for all of their consumers is a creative, yet challenging, effort that must be grounded in expansive and interdisciplinary thinking about the wide range of impacts and the possibilities for identifying, assessing, and addressing them. Any impact study or intervention must also be anchored to the context in which it is performed, as what may be appropriate in one context, may not be in another. Contextual impacts include:

(1) **Regulatory environment**, i.e., different jurisdictions and information domains have different external regulatory requirements.
(2) **Costs vs impact**, i.e., domains have different costs both to produce and consume resources, and these costs affect the impact of different actions.
(3) **Business arrangements and momentum**, i.e., contractual obligations, content availability, etc. can affect what can (or must) be done.
(4) **Infrastructure and staffing**, i.e., it can be much more effective to pick a strategy that leverages existing resources or momentum.
(5) **Impact type**, i.e., the specific kind of impact considered makes a significant difference in the appropriateness and usefulness of an intervention.

Sticking to one approach (e.g. adjusting a dataset or introducing regularization terms to model objectives) might work in some cases, but it is crucial to identify whether other intervention sites (e.g. UX) or combinations of sites and strategies might have a larger impact before picking just one.

The choice of *where* in the IR systems and its sociotechnical context to measure and/or intervene, *how* to intervene at that point, and *who* makes those determinations needs to be well-matched to the specific impact objective and details of the application, domain, market, and users. This also requires careful consideration of multiple perspectives, particularly the voices of those harmed by the system. Key questions include:

– Who is involved in the process and what are their roles and responsibilities (including who gets to participate in imagining what the possibilities could be vs. only get to live with its consequences [17])?
– How are impacts translated into specific objectives and decisions?
– How are different kinds of impacts prioritized?

More research is needed to understand how to implement the broad range of possible interventions (not limited to those we discussed) most effectively, how to select effective interventions, and how the consumer impact problem space decomposes into subproblems. Such research will identify when different interventions may or may not be appropriate, generate new ideas for engaging with impacts not yet considered, and develop a map coupled with evidence-based guidance for future practice. Beyond *more* research, efforts are especially needed to integrate and compare all that research—including non-academic research that feeds into existing practice—to ensure the research community has a more concerted impact itself.

We invite the broader communities of research and practice studying IR, HCI, algorithmic impact, and related topics to join us in this cartographic adventure.

References

1. ABC Mouse (August 2022). https://www.abcmouse.com/
2. Biblionasium (August 2022). https://www.biblionasium.com/
3. Pickatale (August 2022). https://pickatale.com/
4. Abdollahpouri, H., et al.: Multistakeholder recommendation: survey and research directions. User Model. User-Adapted Interact. 30(1), 127–158 (2020). https://doi.org/10.1007/s11257-019-09256-1, ISSN 0924–1868
5. Agarwal, A., Zaitsev, I., Wang, X., Li, C., Najork, M., Joachims, T.: Estimating position bias without intrusive interventions. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 474–482, Association for Computing Machinery, New York. (Jan 2019), https://doi.org/10.1145/3289600.3291017
6. Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., Rieke, A.: Discrimination through optimization: how facebook's ad delivery can lead to biased outcomes. Proc. ACM Hum.-Comput. Interact. **3**(CSCW), 1–30 (2019). https://doi.org/10.1145/3359301
7. Altman, D.G., Bland, J.M.: Statistics notes: units of analysis. BMJ **314**(7098), 1874 (1997). https://doi.org/10.1136/bmj.314.7098.1874, ISSN 0959–8138, 1468–5833
8. Andalibi, N., Garcia, P.: Sensemaking and coping after pregnancy loss. Proc. ACM Hum.-Comput. Interact. **5**(CSCW1), 1–32 (2021). https://doi.org/10.1145/3449201, ISSN 2573–0142
9. Antona, M., Savidis, A., Stephanidis, C.: A process-oriented interactive design environment for automatic user-interface adaptation. Inter. J. Hum.-Comput. Interact. **20**(2), 79–116 (2006). https://doi.org/10.1207/s15327590ijhc2002_2, ISSN 1044–7318
10. Anuyah, O., Milton, A., Green, M., Pera, M.S.: An empirical analysis of search engines' response to web search queries associated with the classroom setting. Aslib J. Inform. Manag. **72**(1), 88–111 (2020). https://doi.org/10.1108/AJIM-06-2019-0143, ISSN 2050–3806
11. Banskota, A., Ng, Y.K.: Recommending video games to adults with autism spectrum disorder for social-skill enhancement. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, pp. 14–22, Association for Computing Machinery, New York (Jul 2020), https://doi.org/10.1145/3340631.3394867
12. Barocas, S., Crawford, K., Shapiro, A., Wallach, H.: The problem wtih bias: Allocative versus representational harms in machine learning. In: 9th Annual Conference of the Special Interest Group for Computing, Information and Society (2017)

13. Barocas, S., et al.: Designing disaggregated evaluations of AI systems: choices, considerations, and tradeoffs. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 368–378, Association for Computing Machinery, New York (Jul 2021). https://doi.org/10.1145/3461702.3462610
14. Beattie, L., Taber, D., Cramer, H.: Challenges in translating research to practice for evaluating fairness and bias in recommendation systems. In: Proceedings of the 16th ACM Conference on Recommender Systems, pp. 528–530, Association for Computing Machinery, New York (Sep 2022). https://doi.org/10.1145/3523227.3547403
15. Becker, C.D., Ostrom, E.: Human ecology and resource sustainability: the importance of institutional diversity. Annal Rev. Ecol. Systematics **26**(1), 113–133 (1995), https://doi.org/10.1146/annurev.es.26.110195.000553, ISSN 0066–4162
16. Belkin, N.J., Robertson, S.E.: Some ethical and political implications of theoretical research in information science. In: Proceedings of the ASIS Annual Meeting (1976). https://www.researchgate.net/publication/255563562
17. Benjamin, R.: Race after Technology: Abolitionist Tools for the New Jim Code. Polity (2019), ISBN 978-1-5095-2640-6
18. Berget, G., Sandnes, F.E.: Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of Google. J. Assoc. Inform. Sci. Technol. **67**(10), 2320–2328 (2016). https://doi.org/10.1002/asi.23572, ISSN 2330–1643
19. Beutel, A., Chen, J., Zhao, Z., Chi, E.H.: Data decisions and theoretical implications when adversarially learning fair representations (Jul 2017). http://arxiv.org/abs/1707.00075
20. Boratto, L., Fenu, G., Marras, M., Medda, G.: Consumer fairness in recommender systems: contextualizing definitions and mitigations. In: Hagen, M., et al. (eds.) ECIR 2022. LNCS, vol. 13185, pp. 552–566. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99736-6_37
21. Burke, R.: Multisided fairness for recommendation (Jul 2017). http://arxiv.org/abs/1707.00093
22. Burke, R., Sonboli, N., Ordonez-Gauger, A.: Balanced neighborhoods for multi-sided fairness in recommendation. In: Friedler, S.A., Wilson, C. (eds.) Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, vol. 81, pp. 202–214. PMLR (2018). http://proceedings.mlr.press/v81/burke18a.html
23. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–336. Association for Computing Machinery, New York (1998). https://doi.org/10.1145/290941.291025
24. Celis, L.E., Keswani, V.: Implicit diversity in image summarization. Proc. ACM Hum.-Comput. Interact. **4**(CSCW2), 1–28 (2020). https://doi.org/10.1145/3415210
25. Channamsetty, S., Ekstrand, M.D.: Recommender response to diversity and popularity bias in user profiles. In: Proceedings of the 30th Florida Artificial Intelligence Research Society Conference. AAAI Press (May 2017). https://aaai.org/papers/657-flairs-2017-15524/
26. Chen, I., Johansson, F.D., Sontag, D.: Why is my classifier discriminatory? In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett,

R. (eds.) Advances in Neural Information Processing Systems, vol. 31, pp. 3539–3550, Curran Associates, Inc. (2018). http://papers.nips.cc/paper/7613-why-is-my-classifier-discriminatory.pdf
27. Cheng, Q., Yom-Tov, E.: Do search engine helpline notices aid in preventing suicide? analysis of archival data. J. Med. Internet Res. **21**(3), e12235 (2019). https://doi.org/10.2196/12235, ISSN 1438–8871
28. Cramer, H., Garcia-Gathright, J., Reddy, S., Springer, A., Takeo Bouyer, R.: Translation, tracks & data: an algorithmic bias effort in practice. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA 2019, pp. 1–8. Association for Computing Machinery, New York (May 2019). https://doi.org/10.1145/3290607.3299057, ISBN 978-1-4503-5971-9
29. Cramer, H., et al.: Challenges of incorporating algorithmic fairness into practice: a tutorial at FAccT 2019 (2019). https://algorithmicbiasinpractice.files.wordpress.com/
30. Crawford, K.: The trouble with bias (Dec 2017). https://youtu.be/fMym_BKWQzk
31. Dabran-Zivan, S., Baram-Tsabari, A., Shapira, R., Yitshaki, M., Dvorzhitskaia, D., Grinberg, N.: "Is COVID-19 a hoax?": auditing the quality of COVID-19 conspiracy-related information and misinformation in Google search results in four languages. Internet Res. **33**(5), 1774–1801 (2023). https://doi.org/10.1108/INTR-07-2022-0560, ISSN 1066–2243
32. Datta, A., Tschantz, M.C., Datta, A.: Automated experiments on ad privacy settings. Proc. Priv. Enhancing Technol. **2015**(1), 92–112 (2015). https://doi.org/10.1515/popets-2015-0007, ISSN 2299–0984
33. December, J.: Units of analysis for internet communication. J. Comput.-Mediated Commun. **1**(4), JCMC143 (1996). https://doi.org/10.1111/j.1083-6101.1996.tb00173.x, ISSN 1083–6101
34. Deldjoo, Y., et al.: Enhancing children's experience with recommendation systems. In: Proceedings of the International Workshop on Children & Recommender Systems (2017). https://yasdel.github.io/files/KidRec17_deldjoo.pdf
35. Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., Zanzonelli, D.: Fairness in recommender systems: Research landscape and future directions. User Model. User-Adapted Interact. (2023). https://doi.org/10.1007/s11257-023-09364-z, ISSN 1573–1391
36. D'Ignazio, C., Klein, L.F.: Data Feminism. MIT Press (2020). https://data-feminism.mitpress.mit.edu/, ISBN 978-0-262-04400-4
37. Downs, B., Pera, M.S., Wright, K.L., Kennington, C., Fails, J.A.: KidSpell: making a difference in spellchecking for children. Inter. J. Child-Comput. Interact., 100373 (2021). https://doi.org/10.1016/j.ijcci.2021.100373, ISSN 2212–8689
38. Dragovic, N., Azpiazu, I.M., Pera, M.S.: From recommendation to curation: when the system becomes your personal docent. In: Proceedings of 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2018), pp. 37–44 (Oct 2018). http://ceur-ws.org/Vol-2225/paper6.pdf
39. Dwork, C., Ilvento, C.: Fairness under composition. In: Blum, A. (ed.) 10th Innovations in Theoretical Computer Science Conference (ITCS 2019), Leibniz International Proceedings in Informatics (LIPIcs), vol. 124, pp. 33:1–33:20, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2019). https://doi.org/10.4230/LIPICS.ITCS.2019.33
40. Ekstrand, M.D., Carterette, B., Diaz, F.: Distributionally-informed recommender system evaluation. ACM Trans. Recommender Syst. (2023). https://doi.org/10.1145/3613455

41. Ekstrand, M.D., Das, A., Burke, R., Diaz, F.: Fairness in information access systems. Foundat. Trends® Inform. Retrieval **16**(1–2), 1–177 (2022). https://doi.org/10.1561/1500000079, ISSN 1554–0669
42. Ekstrand, M.D., Das, A., Burke, R., Diaz, F.: Fairness in recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 679–707, Springer, US (2022). https://doi.org/10.1007/978-1-0716-2197-4_18, ISBN 978-1-07-162197-4
43. Ekstrand, M.D., Kluver, D.: Exploring author gender in book rating and recommendation. User Model. User-Adapted Interact. **31**(3), 377–420 (2021). https://doi.org/10.1007/s11257-020-09284-2, ISSN 0924–1868
44. Ekstrand, M.D., Pera, M.S.: Matching consumer fairness objectives & strategies for RecSys (Sep 2022). http://arxiv.org/abs/2209.02662
45. Ekstrand, M.D., et al.: All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In: Friedler, S.A., Wilson, C. (eds.) Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, vol. 81, pp. 172–186. PMLR (2018). https://proceedings.mlr.press/v81/ekstrand18b.html
46. Epps-Darling, A., Bouyer, R.T., Cramer, H.: Artist gender representation in music streaming. In: Proceedings of the 21st International Society for Music Information Retrieval Conference, pp. 248–254. ISMIR (Oct 2020). https://program.ismir2020.net/poster_2-11.html
47. Farrand, T., Mireshghallah, F., Singh, S., Trask, A.: Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy. In: Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice, pp. 15–19. Association for Computing Machinery, New York (Nov 2020). https://doi.org/10.1145/3411501.3419419, ISBN 978-1-4503-8088-1
48. Feng, Y., Shah, C.: Has CEO gender bias really been fixed? Adversarial attacking and improving gender fairness in image search. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36(11), pp. 11882–11890 (Jun 2022). https://doi.org/10.1609/aaai.v36i11.21445, ISSN 2374–3468
49. Fish, B., Bashardoust, A., Boyd, D., Friedler, S., Scheidegger, C., Venkatasubramanian, S.: Gaps in information access in social networks? In: WWW 2019: The World Wide Web Conference, pp. 480–490. Association for Computing Machinery, New York (May 2019). https://doi.org/10.1145/3308558.3313680
50. Fourney, A., Ringel Morris, M., Ali, A., Vonessen, L.: Assessing the readability of web search results for searchers with dyslexia. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1069–1072, Association for Computing Machinery, New York. (Jun 2018), https://doi.org/10.1145/3209978.3210072
51. Gao, R., Ge, Y., Shah, C.: FAIR: Fairness-aware information retrieval evaluation. J. Assoc. Inform. Sci. Technol. **73**(10), 1461–1473 (2022). https://doi.org/10.1002/asi.24648, ISSN 2330–1643,
52. Goel, N., Faltings, B.: Crowdsourcing with fairness, diversity and budget constraints. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 297–304. Association for Computing Machinery, New York (Jan 2019). https://doi.org/10.1145/3306618.3314282
53. Gossen, T., Nitsche, M., Vos, J., Nürnberger, A.: Adaptation of a search user interface towards user needs: A prototype study with children & adults. In: Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval, pp. 1–10. Association for Computing Machinery, New York (Oct 2013). https://doi.org/10.1145/2528394.2528397, ISBN 978-1-4503-2570-7

54. Gunawardana, A., Shani, G., Yogev, S.: Evaluating recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, third edn., pp. 547–601, Springer, US (2022). https://doi.org/10.1007/978-1-0716-2197-4_15, ISBN 978-1-07-162196-7
55. Harambam, J., Bountouridis, D., Makhortykh, M., van Hoboken, J.: Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In: Proceedings of the 13th ACM Conference on Recommender Systems, pp. 69–77. Association for Computing Machinery, New York (Sep 2019), https://doi.org/10.1145/3298689.3347014
56. Helberger, N.: On the democratic role of news recommenders. Digital J. **7**(8), 993–1012 (2019). https://doi.org/10.1080/21670811.2019.1623700, ISSN 2167–0811
57. Holstein, K., Wortman Vaughan, J., Daumé, III, H., Dudik, M., Wallach, H.: Improving fairness in machine learning systems: What do industry practitioners need? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–16. Association for Computing Machinery, New York (May 2019). https://doi.org/10.1145/3290605.3300830
58. Huang, W., Labille, K., Wu, X., Lee, D., Heffernan, N.: Achieving User-Side Fairness in Contextual Bandits. Hum.-Centric Intel. Syst. **2**(3), 81–94 (2022. https://doi.org/10.1007/s44230-022-00008-w, ISSN 2667–1336
59. IMDb: Se7en (1995) - IMDb (1995). http://www.imdb.com/title/tt0114369/parentalguide
60. Jacobs, A.Z., Wallach, H.: Measurement and fairness. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 375–385, Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3442188.3445901
61. Joachims, T., Swaminathan, A., Schnabel, T.: Unbiased learning-to-rank with biased feedback. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, California (Jul 2018). https://doi.org/10.24963/ijcai.2018/738
62. Kamishima, T., Akaho, S., Asoh, H., Sakuma, J.: Recommendation independence. In: Friedler, S.A., Wilson, C. (eds.) Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, vol. 81, pp. 187–201. PMLR, New York (2018). http://proceedings.mlr.press/v81/kamishima18a.html
63. Karako, C., Manggala, P.: Using image fairness representations in diversity-based re-ranking for recommendations. In: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 23–28. Association for Computing Machinery, New York. (Jul 2018). https://doi.org/10.1145/3213586.3226206
64. Katzman, J., et al.: Taxonomizing and measuring representational harms: a look at image tagging. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37(12), pp. 14277–14285 (Jun 2023). https://doi.org/10.1609/aaai.v37i12.26670, ISSN 2374–3468
65. Landoni, M., Matteri, D., Murgia, E., Huibers, T., Pera, M.S.: Sonny, cerca! evaluating the impact of using a vocal assistant to search at school. In: Crestani, F., et al. (eds.) CLEF 2019. LNCS, vol. 11696, pp. 101–113. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28577-7_6
66. Lawrence, E.E.: On the problem of oppressive tastes in the public library. J. Documentation **76**(5), 1091–1107 (2020). https://doi.org/10.1108/JD-01-2020-0002, ISSN 0022–0418

67. Leonhardt, J., Anand, A., Khosla, M.: User fairness in recommender systems. In: Companion Proceedings of the The Web Conference 2018, pp. 101–102, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (Apr 2018). https://doi.org/10.1145/3184558.3186949
68. Li, Y., Chen, H., Fu, Z., Ge, Y., Zhang, Y.: User-oriented fairness in recommendation. In: Proceedings of the Web Conference 2021, pp. 624–632. Association for Computing Machinery, New York (Apr 2021). https://doi.org/10.1145/3442381.3449866
69. Madrazo Azpiazu, I., Dragovic, N., Anuyah, O., Pera, M.S.: Looking for the movie seven or Sven from the movie Frozen? A multi-perspective strategy for recommending queries for children. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, pp. 92–101. Association for Computing Machinery, New York (Mar 2018). https://doi.org/10.1145/3176349.3176379
70. Mauro, N., Ardissono, L., Cena, F.: Personalized recommendation of PoIs to people with autism. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, pp. 163–172. Association for Computing Machinery, New York (Jul 2020). https://doi.org/10.1145/3340631.3394845
71. McDonald, G., Macdonald, C., Ounis, I.: Search results diversification for effective fair ranking in academic search. Inform. Retrieval J. **25**(1), 1–26 (2022). https://doi.org/10.1007/s10791-021-09399-z, ISSN 1573–7659
72. McNealy, J., Cramer, H.: Trust and representation in recommender systems. In: ICA 2022 (2022)
73. Mehrotra, R., Anderson, A., Diaz, F., Sharma, A., Wallach, H., Yilmaz, E.: Auditing search engines for differential satisfaction across demographics. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 626–633, International World Wide Web Conferences Steering Committee (2017). https://doi.org/10.1145/3041021.3054197
74. Mehrotra, R., Carterette, B., Li, Y., Yao, Q., Gao, C., Kwok, J., Yang, Q., Guyon, I.: Advances in recommender systems: from multi-stakeholder marketplaces to automated RecSys. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3533–3534. Association for Computing Machinery, New York (Aug 2020). https://doi.org/10.1145/3394486.3406463, ISBN 978-1-4503-7998-4
75. Michiels, L., Leysen, J., Smets, A., Goethals, B.: What are filter bubbles really? A review of the conceptual and empirical work. In: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2022 Adjunct, pp. 274–279. Association for Computing Machinery, New York (Jul 2022). https://doi.org/10.1145/3511047.3538028, ISBN 978-1-4503-9232-7
76. Milton, A., Pera, M.S.: Into the unknown: exploration of search engines' responses to users with depression and anxiety. ACM Trans. Web **17**(4), 25:1–25:29 (Jul 2023). https://doi.org/10.1145/3580283, ISSN 1559–1131
77. Moore, M., Bias, R.G., Prentice, K., Fletcher, R., Vaughn, T.: Web usability testing with a Hispanic medically underserved population. J. Med. Library Assoc. JMLA **97**(2), 114–121 (Apr 2009). https://doi.org/10.3163/1536-5050.97.2.008, ISSN 1536–5050, 1558–9439
78. Mosseri, A.: Instagram ranking explained (May 2023). https://about.instagram.com/blog/announcements/instagram-ranking-explained/
79. Murgia, E., Abbasiantaeb, Z., Aliannejadi, M., Huibers, T., Landoni, M., Pera, M.S.: ChatGPT in the classroom: A preliminary exploration on the feasibility of adapting ChatGPT to support children's information discovery. In: Adjunct

Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, pp. 22–27. Association for Computing Machinery, New York (Jun 2023). https://doi.org/10.1145/3563359.3597399, ISBN 978-1-4503-9891-6
80. Murgia, E., Landoni, M., Huibers, T., Fails, J.A., Pera, M.S.: The seven layers of complexity of recommender systems for children in educational contexts. In: Proceedings of the Workshop on Recommendation in Complex Scenarios Co-Located with 13th ACM Conference on Recommender Systems, vol. 2449. CEUR-WS (Sep 2019). http://ceur-ws.org/Vol-2449/paper1.pdf
81. Naghiaei, M., Rahmani, H.A., Deldjoo, Y.: CPFair: personalized consumer and producer fairness re-ranking for recommender systems. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 770–779, Association for Computing Machinery, New York (Jul 2022). https://doi.org/10.1145/3477495.3531959
82. Netflix: Children & family movies (August 2022). https://www.netflix.com/browse/genre/783
83. Ng, Y.K., Pera, M.S.: Recommending social-interactive games for adults with autism spectrum disorders (ASD). In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 209–213. Association for Computing Machinery, New York (Sep 2018). https://doi.org/10.1145/3240323.3240405
84. Odinma, A.C., Butakov, S., Grakhov, E.: Improving the browsing experience in a bandwidth limited environment through traffic management. Inform. Technol. Developm. **17**(4), 306–318 (2011). https://doi.org/10.1080/02681102.2011.568224, ISSN 0268–1102
85. Olteanu, A., Diaz, F., Kazai, G.: When are search completion suggestions problematic? Proc. ACM Hum.-Comput. Interact. **4**(CSCW2), 171:1–171:25 (2020). https://doi.org/10.1145/3415242
86. Pardes, A.: Feeling stressed out? Pinterest wants to help. Wired (Jul 2019). https://www.wired.com/story/pinterest-compassionate-search/
87. Pariser, E.: The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. Penguin (May 2011), ISBN 978-1-101-51512-9
88. Partarakis, N., Doulgeraki, C., Leonidis, A., Antona, M., Stephanidis, C.: User interface adaptation of web-based services on the semantic web. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 711–719. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02710-9_79
89. Patton, D.U.: Social work thinking for UX and AI design. Interactions **27**(2), 86–89 (2020). https://doi.org/10.1145/3380535, ISSN 1072–5520
90. Pera, M.S., Murgia, E., Landoni, M., Huibers, T.: With a little help from my friends: Use of recommendations at school. In: Proceedings of ACM RecSys 2019 Late-breaking Results, CEUR-WS, vol. 2431. CEUR (2019). http://ceur-ws.org/Vol-2431/paper13.pdf
91. Pera, M.S., Ng, Y.K.: Automating readers' advisory to make book recommendations for K-12 readers. In: Proceedings of the 8th ACM Conference on Recommender Systems, pp. 9–16. Association for Computing Machinery, New York (2014). https://doi.org/10.1145/2645710.2645721
92. Pinterest: Pinterest introduces first-of-its-kind hair pattern search for inclusive beauty results (Aug 2021). https://newsroom.pinterest.com/en-gb/post/pinterest-introduces-first-of-its-kind-hair-pattern-search-for-inclusive-beauty-results
93. Qin, X., Dou, Z., Wen, J.R.: Diversifying search results using self-attention network. In: Proceedings of the 29th ACM International Conference on Information

& Knowledge Management, CIKM 2020, pp. 1265–1274. Association for Computing Machinery, New York (Oct 2020). https://doi.org/10.1145/3340531.3411914, ISBN 978-1-4503-6859-9

94. Raj, A., Ekstrand, M.D.: Fire dragon and unicorn princess: gender stereotypes and children's products in search engine responses. In: Proceedings of the 2022 SIGIR Workshop On eCommerce (Jun 2022). http://arxiv.org/abs/2206.13747
95. Rakova, B., Yang, J., Cramer, H., Chowdhury, R.: Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices (Jun 2020). http://arxiv.org/abs/2006.12358
96. Rastegarpanah, B., Gummadi, K.P., Crovella, M.: Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 231–239. ACM (Jan 2019). https://doi.org/10.1145/3289600.3291002
97. Riedl, J., Konstan, J.: Word of Mouse. Warner Books (2002), ISBN 978-0-446-53003-3
98. Rodriguez, M., Posse, C., Zhang, E.: Multiple objective optimization in recommender systems. In: Proceedings of the Sixth ACM Conference on Recommender Systems, pp. 11–18. Association for Computing Machinery, New York (Sep 2012). https://doi.org/10.1145/2365952.2365961, ISBN 978-1-4503-1270-7
99. Rothschild, M., Horiuchi, T., Maxey, M.: Evaluating "just right" in EdTech recommendation. In: KidRec '19: Workshop in International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems, Co-located with ACM IDC (2019). https://kidrec.github.io/papers/KidRec_2019_paper_6.pdf
100. Shrestha, A., Spezzano, F., Pera, M.S.: An empirical analysis of collaborative recommender systems robustness to shilling attacks. In: Proceedings of the Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems Co-Located with RecSys 2021, CEUR-WS, vol. 3012, pp. 45–57. CEUR (Oct 2021). http://ceur-ws.org/Vol-3012/OHARS2021-paper4.pdf
101. Smith, J.J., Beattie, L., Cramer, H.: Scoping fairness objectives and identifying fairness metrics for recommender systems: The practitioners' perspective. In: Proceedings of the ACM Web Conference 2023, pp. 3648–3659, Association for Computing Machinery, New York (Apr 2023). https://doi.org/10.1145/3543507.3583204, ISBN 978-1-4503-9416-1
102. Sofaer, S.: Qualitative methods: what are they and why use them? Health Serv. Res. **34**(5 Pt 2), 1101–1118 (1999). https://pubmed.ncbi.nlm.nih.gov/10591275/, ISSN 0017-9124
103. Sonboli, N., Burke, R., Ekstrand, M., Mehrotra, R.: The multisided complexity of fairness in recommender systems. AI Mag. **43**(2), 164–176 (2022). https://doi.org/10.1002/aaai.12054, ISSN 0738–4602
104. Steck, H.: Calibrated recommendations. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 154–162. Association for Computing Machinery (Sep 2018). https://doi.org/10.1145/3240323.3240372
105. Trust & Safety Professionals Association: Abuse types (Jun 2021). https://www.tspa.org/curriculum/ts-fundamentals/policy/abuse-types/
106. Tsiakas, K., Barakova, E., Khan, J.V., Markopoulos, P.: BrainHood: towards an explainable recommendation system for self-regulated cognitive training in children. In: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–6. Association for

Computing Machinery, New York (Jun 2020). https://doi.org/10.1145/3389189.3398004
107. Tsioutsiouliklis, S., Pitoura, E., Semertzidis, K., Tsaparas, P.: Link recommendations for PageRank fairness. In: Proceedings of the ACM Web Conference 2022, pp. 3541–3551. Association for Computing Machinery, New York (Apr 2022). https://doi.org/10.1145/3485447.3512249, ISBN 978-1-4503-9096-5
108. Valentim, I., Lourenço, N., Antunes, N.: The impact of data preparation on the fairness of software systems. In: 2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE), pp. 391–401 (Oct 2019). https://doi.org/10.1109/ISSRE.2019.00046, ISSN 2332–6549
109. van Alstyne, M., Brynjolfsson, E.: Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. Manag. Sci. **51**(6), 851–868 (2005). https://doi.org/10.1287/mnsc.1050.0363, ISSN 0025–1909
110. Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., Helberger, N.: Recommenders with a mission: Assessing diversity in news recommendations. In: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, pp. 173–183. Association for Computing Machinery, New York (Mar 2021). https://doi.org/10.1145/3406522.3446019, ISBN 978-1-4503-8055-3
111. Wang, L., Joachims, T.: User fairness, item fairness, and diversity for rankings in two-sided markets. In: Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 23–41. Association for Computing Machinery, New York (Jul 2021). https://doi.org/10.1145/3471158.3472260
112. Wortman Vaughan, J.: Transparency and intelligibility throughout the machine learning life cycle (Jan 2020). https://www.microsoft.com/en-us/research/video/transparency-and-intelligibility-throughout-the-machine-learning-life-cycle/
113. Wu, H., Mitra, B., Ma, C., Diaz, F., Liu, X.: Joint multisided exposure fairness for recommendation. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 703–714. ACM (Jul 2022), https://doi.org/10.1145/3477495.3532007
114. Yang, Y., Zhang, C., Fan, C., Mostafavi, A., Hu, X.: Towards fairness-aware disaster informatics: An interdisciplinary perspective. IEEE Access **8**, 201040–201054 (2020). https://doi.org/10.1109/ACCESS.2020.3035714, ISSN 2169–3536
115. Yao, S., Huang, B.: Beyond parity: Fairness objectives for collaborative filtering. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 2925–2934. Curran Associates, Inc. (2017). http://papers.nips.cc/paper/6885-beyond-parity-fairness-objectives-for-collaborative-filtering.pdf