Are Recommender Systems Serving Children? Toward Child-Aware Design and Evaluation

Ungruh, Robin

DOI
10.1145/3705328.3748752

Publication date
2025

Document Version
Final published version

Published in
RecSys2025 - Proceedings of the 19th ACM Conference on Recommender Systems

Are Recommender Systems Serving Children? Toward Child-Aware Design and Evaluation

Robin Ungruh
Delft University of Technology
Delft, Netherlands
R.Ungruh@tudelft.nl

Abstract

Recommender Systems research continuously improves recommendation strategies to meet the needs of a wide range of users and other stakeholders. However, much of this research centers on the traditional, adult user, often overlooking underrepresented demographics. One such group is children, frequent users of platforms driven by recommender systems. Children differ from adults in preferences and can be particularly vulnerable to certain content, raising questions about the harm recommender systems may pose.

This PhD project advocates for *child-aware* recommender systems: systems that explicitly account for children as part of their users, recognizing their distinct needs, vulnerabilities, and rights. In pursuit of this goal, we investigate how well current recommender systems serve children, auditing algorithmic strategies from two complementary perspectives: The 'traditional' perspective focuses on whether recommendations align with children's preferences. The perspective of 'non-maleficence' assesses suitability of content recommended, evaluating whether it respects children's vulnerabilities to potentially harmful material. To do so, we audit current recommender systems according to both perspectives—not only in the short term, but also in the long term through simulation studies. Beyond auditing, we explore strategies and design directions for making recommender systems more responsible. Outcomes from this work should inform both academic and practitioner communities about the gaps in current systems and lay the groundwork for more equitable, safe, and meaningful recommendations for children.

CCS Concepts

• **Information systems** → **Recommender systems**; • **Social and professional topics** → **Children**.

Keywords

Recommender Systems, Children, Harms, Simulations

ACM Reference Format:
Robin Ungruh. 2025. Are Recommender Systems Serving Children? Toward Child-Aware Design and Evaluation. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25), September 22–26, 2025, Prague, Czech Republic.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3705328.3748752

ACM ISBN 979-8-4007-1364-4/25/09
https://doi.org/10.1145/3705328.3748752

1 Introduction

The traditional goal of recommendations is to provide different users with suggestions tailored to their preferences [56]. Thus, a recommender should have knowledge about *every* user [33]. However, users are diverse, and not every user is equally represented by the data used to build recommender systems (RS), and differences in preferences might be overlooked in classical design and evaluations. This unequal representation might lead to problems since RS particularly learn the preferences of homogeneous groups and majorities [12, 17, 40, 44, 66], resulting in systems less suited for minority groups. Recent research efforts in ethical recommender system design critique this bias, emphasizing consumer fairness—the principle that all users should be served equitably, regardless of individual characteristics or group membership [1, 76].

One minority group that remains mostly overlooked is children. Children are prominent users of online platforms [48, 51, 54, 72], where they frequently encounter content curated by RS [70]. Despite the need to study children's interactions with information access systems thoroughly [38], current research on the interplay between children and RS is sparse, evidencing a critical gap in understanding how these systems behave when children are the users. To bridge this gap, it is crucial to recognize that children cannot simply be treated as smaller or less experienced adults [8] that have the same experiences and needs. Children are in developmental phases [2, 6, 31], with different interests and preferences for media [45, 62] and varying skills that influence how they perceive, interact, and are impacted by online media [41]. Further, children are a vulnerable user group—the media they consume can profoundly impact their development and perception of the world [2, 7].

Motivated by these distinct characteristics, we aim to investigate the extent to which RS designed for a broad user base serve children, and how potential drawbacks can be addressed effectively, by involving knowledge about children explicitly in the design and evaluation of RS. To achieve this, we do not focus on RS specifically tailored to children, as—in reality—children interact frequently with platforms that are for a general audience [42], and thus encounter RS that aim to tailor recommendations to a broad range of users. Instead, we advocate for **child-aware** RS that acknowledge that children are a part of their user base, which must be considered throughout the design and evaluation of these. Central to this goal is the question: what does it mean to truly *serve* children in the context of RS? Motivated by children's needs and developmental context, we approach this through two key perspectives:

(1) The "classical" perspective includes concepts such as performance and relevance. Those traditionally assess whether an item is considered successful [5], i.e., whether users like

the recommended items. Focusing on these aspects specifically in the context of children acknowledges their specific preferences and consumption patterns.

(2) The principle of *non-maleficence* emphasizes the importance of avoiding harm; considering children's vulnerabilities and that a large number of items present on online platforms can be harmful to children [63]. This principle thus includes the need to recognize and mitigate risks such as the reinforcement of stereotypes or exposure to inappropriate content, which can have long-lasting effects on children.

To advance knowledge of RS and children under consideration of the two perspectives, we define three research objectives (ROs).

RO1: Characterizing recommenders' alignment with children's preferences. RO1 aims to characterize differences between children and other user groups. While children are typically the statistical minority in datasets used for RS research, young adults dominate the data. This majority group, which we refer to as *mainstream users*, may be the driver of recommendations, potentially affecting the suggestions children are presented with. In pursuit of this RO, we address three research questions (RQs):

RQ1.1: *To what degree do children's* ***interactions and consumption patterns*** *differ from those of more prominent mainstream adults?* This *RQ* drives research focus on assessing the extent to which children are *not* part of the mainstream. As we assume that children have a unique role in RS environments, we characterize the items consumed by users to identify and model the distinct consumption patterns of children.

RQ1.2: *To what extent are* ***recommendations for children*** *affected by more prominent mainstream user groups?* With this *RQ*, we investigate to what degree RS are actually able to create high-quality recommendations for children despite the prominence of mainstream users. To do so, we probe whether differences in consumption patterns and preferences manifest in recommendations children receive, i.e., whether they receive recommendations that reflect their own preferences or those of the mainstream.

RQ1.3: *How can we model children's interactions with RS in* ***simulation studies*** *to enable meaningful and realistic long-term insights?* In order to not only capture short-term performance of RS for children, but also examine their long-term impact on children and their long-lasting alignment with their preferences, we turn to simulation-based RS studies, as collecting real-world interaction data from children poses ethical and practical concerns. Simulations can enable controlled experimentation of long-term effects of algorithmic strategies where little user data is available or user access is limited [32]. However, the realism and ecological validity of simulation studies for children remain uncertain. Unlike adults, children are in active developmental stages—cognitively, emotionally, and socially—which likely affects how they engage with systems over time. As a result, simulations based on static or adult-centric choice models may not capture the dynamics of children's evolving preferences or vulnerabilities. To make simulations a viable tool to assess children in RS environments, we answer this *RQ* to evaluate and adapt how these choice models represent child behavior.

RO2: Auditing recommenders' non-maleficence. This RO is geared to evaluate RS from the perspective of non-maleficence. As children can be considered particularly vulnerable users [43], we argue that adhering to this principle should be a core priority for RS—motivating us to audit whether current systems align with it.

RQ2.1*: What are the* ***characteristics of items that can be deemed harmful*** *for certain children when recommended?* This *RQ* aims to define what it means for an item recommended to children to be non-maleficent, i.e., non-harmful. This requires nuanced understanding of children's developmental needs, individual differences in vulnerabilities, and the content that may pose risks. Importantly, this line of inquiry seeks to lays the groundwork for more systematic approaches to identifying such risks, supporting efforts to operationalize harm-related concepts in the context of recommender systems.

RQ2.2: *To what extent do RS present children with* ***recommendations that violate the principle of non-maleficence****?* To assess whether recommendations present children to potentially harmful items, we assess the recommendations children receive in systems not specifically designed for them. As children are frequent users of such systems, this allows us to gauge the degree to which children can be exposed to unfitting or maleficent content. For this, we aim to operationalize risk assessments of RS from a child-centered perspective to make risks more transparent and actionable.

RO3: Designing 'child-aware' recommender systems that truly serve children. Based on insights from the previous objectives, research efforts for *R03* lead to the design of RS that aim to create recommendations that truly serve children. They must explicitly recognize children as part of the user base. We argue that *child-aware recommender systems* should generate recommendations that are developmentally appropriate and responsive to individual needs. They must be fitting, both from the 'traditional' perspective but also in terms of non-maleficence, not only in the short term but also over time.

RQ3.1: *Which* ***algorithmic solutions*** *that optimize toward traditional performance as well as non-maleficence can achieve recommendations that are truly 'fitting' for children?* To design 'child-aware' methods, we explore suitable solutions. This will be done by either adapting existing strategies or developing new multi-objective recommendation solutions.

RQ3.2: *How can the* ***trade-off between traditional performance and non-maleficence*** *be evaluated?* In pursuit of the RO, developed strategies need to be assessed from both perspectives. This requires a holistic grasp of children's needs and preferences, but also vulnerabilities. Further, to be viable in real-world settings, these systems must also maintain high utility for adults, ensuring that improvements for children do not come at the expense of broader user satisfaction. Thus, we develop a thorough evaluation approach that considers users affected and determines the fit of items with respect to our perspectives.

2 Related Work

Here, we lay out works relevant to our research project.

Recommenders & Children. Despite children being frequent users of platforms driven by RS, they are rarely considered in the design and evaluation of these systems. As noted by Ekstrand [15] and Gómez Gutiérrez et al. [26], key challenges in studying RS for children include limited data availability, constrained opportunities for user studies, and a lack of clearly defined assessment methods. Previous studies mainly focus on highlighting children's distinct preferences [45, 50, 59, 62], and designing RS specifically designed for children to suggest learning reading or learning material [e.g., 37, 49, 53, 58, 65]. These studies highlight the potential of personalizing items to children that match their unique needs, skills, or preferences [26, 46]. However, even the design of RS for children in a specific domain, such as education, faces a multitude of complexities that need to be considered [47]. Largely overlooked is the perspective of how children are treated when they are just *one* user group out of many that use the same system.

Harmful Recommendations. One frequent concern when it comes to children's interactions with RS is the suitability of items for children [15, 26, 46, 47]. Tang and Winoto [64] argue that appropriateness should be a relevant consideration of recommender systems that goes beyond simple accuracy or liking of a user. However, many prominent platforms that utilize RS show that inappropriate items often appear on those [24, 52, 55]. For children, who are a particularly vulnerable population [43] and susceptible to being influenced by consumed media [3], such content can be distressing and potentially harmful. This underscores the urgent need for responsible recommender systems that can recognize vulnerability in users like children and actively avoid recommending harmful content [11, 25]. Addressing these risks requires moving beyond technical optimization, toward ethical considerations that center the well-being of vulnerable users like children.

Recommender System Simulations. Simulation-based RS studies allow the analysis of what users would be exposed to when interacting with a system repeatedly [16, 29]. Various effects of RS that only become apparent when a user interacts continuously with RS can be uncovered with those (e.g., bias amplification [20–22, 44] or preference homogenization [10]), providing implications about the societal impacts RS may have when deployed and interacted with over extended time. Previous studies highlight how different users are treated differently by RS in repeated interactions [21, 44]. The power to make implications about **long-term interactions** with RS [14, 17, 77] is particularly relevant in cases where no real users can be observed [32], as in the case of child-focused studies.

Despite the potential to obtain insights with high ecological validity from simulation-based studies, recently, concerns about the realism of simulation settings have arisen. Specifically, choice models tend to simplify interactions with recommended items, and they assume the same model for each user [e.g., 19, 22, 34, 39]. However, realistic choice models are imperative to capture real-life impacts [9], particularly when studying differences between users.

3 Methods

Children are a protected user group, and research involving them requires careful ethical deliberation [13, 18, 36]. Due to their vulnerability, children are subject to specific safeguards and privacy regulations. As such, our work primarily relies on publicly available datasets that contain anonymized data from self-declared child users, collected following the terms of service of platforms. Additionally, we draw on datasets where we can infer associations with children's interests based on item characteristics. To ensure safety and ethical integrity, we adopt offline evaluation methods, allowing us to generate initial insights and improve RS for children without directly exposing them to potentially inappropriate content.

Our methods will mainly include empirical explorations that investigate how current recommender systems fare for children, with a focus on both preference alignment and non-maleficence. To this end, we quantify the fit of items to children's preferences and sensitivities (*RQ1.1* and *RQ2.1*) by analyzing item characteristics in commonly used recommender system datasets. This involves identifying features that may signal preference relevance or potential harm. We empirically evaluate how well these characteristics are reflected in actual recommendations by simulating recommendations and interactions, allowing us to assess both short-term alignment and long-term effects (*RQ1.2* and *RQ2.2*). We will explore strategies to adapt or design RS that explicitly account for children's needs (*RO3*). This includes developing interventions and training objectives that incorporate child-specific constraints, preferences, or safety considerations. Further, we empirically evaluate the outcomes of such design methods in a holistic manner by considering whether *all* goals of the methods are achieved for *all* users.

Datasets. Generally, we focus on datasets from entertainment-related domains. As for those, a certain degree of autonomy of users can be expected that may not be available when it comes to domains such as e-commerce or tourism, where parents would have a more prominent role in children's decisions. However, publicly available datasets that include demographic information are rare—even more so when it comes to the inclusion of child-related data. One of the most prominent datasets in RS research, Movielens-1m [28], includes a limited number of 222 users under 18, with no specifics about the more concrete age of these users. Other datasets, such as Bookcrossing [79] or MyAnimeList[1], provide more detailed information about users' ages, but the unavailability of timestamps leads to uncertainty about the reliability of age information at the time of the interactions. The most detailed information can be found in LFM-2b [60]—yet, this is not publicly available anymore—and MLHD [73], both including listening events with songs. To expand our analysis to other domains beyond music, we aim to leverage datasets without demographic information, but in which we can infer preferences that may align with child interests. For instance, in the Goodreads dataset [74, 75], genre information hints toward the target audience, i.e., `Children's` and `Young Adult`.

Offline Evaluations. We utilize these datasets in a variety of empirical offline evaluations. Through data analyses, we evaluate children's preferences, oftentimes using genre consumption as a proxy. To assess appropriateness of items, we rely on a variety of measures, including those based on natural language processing

[1]https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset

[e.g., 4, 23]. Auditing of RS performance and maleficence will be evaluated with classical offline RS experiments [27]—focused on differences between users of varying ages—but also with simulation-based studies [16, 29]. These can provide insights into long-term dynamics between users and RS where no real users can be observed [32]. As direct access to users or sufficient data is limited when it comes to children [15, 26], simulations have the opportunity to gain nuanced insights into the interplay between RS and children.

Child-aware Recommender Systems Design. To design strategies for *RO3*, various methods will be explored. One key component requires the detection of appropriateness levels of children. As RS typically do not have explicit information about users' demographics or specific needs or vulnerabilities, we will explore methods to implicitly assess these. Insights from these can then be further leveraged to improve recommendations for children. Approaches for this can range from simple reranking approaches [35, 61], which could be used to adapt already trained RS, to more advanced approaches such as multi-objective optimization that considers both perspectives during training [57, 78].

4 Progress and Results

Here, we lay out research efforts conducted in pursuit of our ROs as well as plans for future work for *RO* completion. An overview of the current progress throughout the years, as well as planning for the remainder of the PhD trajectory, can be seen in Table 1.

Characterizing recommenders' alignment with children's preferences. Throughout years 1 and 2, we conducted two studies to address *RQ1.1* and *RQ1.2*. In [69], we explore preferences of users of varying ages, assessed through previous genre consumption, in a movie and music-related dataset. We analyzed whether salient differences in consumption patterns between children and mainstream adults are captured by a variety of recommendation algorithms despite children being underrepresented in the datasets used. Our analysis on LFM-2b highlighted that while children generally receive recommendations overall well-aligned with their preferences, RS show inconsistent behavior when removing mainstream users' influence on the training of an RS, i.e., when solely training on children's data. To further explore these effects, we expanded this research across more datasets in a reproducibility study [71] (under review). Here, we find that differences in preferences between children and mainstream users persist in the music, movie, and book domains. Moreover, whether a recommender is able to create recommendations well-aligned with a child's preferences depends on the dataset studied. On those where an recommendation algorithm (RA) performs well for children, we generally find that they are interested in overall popular items. The results from these two studies highlight that (1) children are *not* mainstream users, but have prominently different genre preferences than the mainstream (*RQ1.1*), and (2) that recommenders may fail when it comes to children if their preferences are not fulfilled by popularity-biased recommendations (*RQ1.2*). This highlights the importance of considering that RS that are expected to perform well for their user base may fail to cater to some of their underrepresented users, such as children—even in a classical sense of performance.

In year 3, we aim to further assess how well recommendations align with children's preferences. In particular, given children's active development, we consider whether RS can provide high-quality recommendations for children **in the long term**. For this, we utilize simulation-based RS studies. To gather realistic insights from these studies, a realistic simulation of users' interactions with these systems is imperative to maintain high ecological validity of findings. However, explorations in year 2 show a crucial issue of simulation studies focused on children: *Choice models*—models that simulate the users' interactions with recommendations—typically assume the same behavior of users when presented with recommendations (homogeneous choice models). Concerns arise that these cannot capture individual user behavior, particularly those of children who may deviate from 'typical' behavior. To highlight this issue addressed by *RQ1.3*, we show that homogeneous choice models in a simulation study fail to capture the individual behavior of users realistically [68]. We are in the midst of another study that highlights how well **children's interactions** with RS are captured by choice models. We investigate whether short-term explorations of users' choice behavior can be utilized to create long-term valid choice models for users of different ages. As children undergo developmental changes throughout childhood, long-term insights from simulation studies about the interplay between RS and children may not be reliable. In the latter half of year 2, we will explore methods to better capture children's interactions with RS to create simulation studies that can provide thorough insights into the interplay between children and RS.

Auditing recommenders' non-maleficence. To contribute research outlined in *RO2*, we conducted initial studies. In year 1, we analyzed a common corpus of song lyrics with different lenses to probe which lyrics can be deemed harmful (*RQ2.1*) [30]. As expressions that could be classified as unsuitable for certain children were found in almost every lyric, we discuss what it means to be truly 'fitting' for a child. Nuanced differences between different ages, but also different cultural backgrounds or experiences, may influence what would be a 'fit' for a child. Our reflections prompt that creating recommendations aimed to be non-maleficent is not a simple content-detection and filtering task but instead requires a nuanced understanding of children's needs, experiences, and perceptions.

In another study conducted in year 1 [67], we probe items based on another type of harm, namely stereotypes. We investigate the prominence of different types of stereotypes in items in large movie and book item corpora and apply common RS. Analyzing the presence of stereotypes within recommendations made to children allowed us to gauge how frequently children may be exposed to stereotypical items due to RS underlying the platforms used by them. This prompts us to reflect on the efforts necessary to mitigate the effects of such content that can profoundly impact children's self-perception, but also their picture of other social groups.

Particularly throughout the latter half of year 2 and year 3, we aim to focus on the operationalization of risk assessments from a child-centered perspective. Our goal is to systematically identify which types of recommendations may pose potential risks or be deemed harmful to children. This operationalization not only provides a foundation for evaluating RS and RS-driven platforms, but also offers a structured approach for assessing the effectiveness of

Table 1: Current progress on the research objectives. A '*P*' in a cell indicates that research efforts are planned for the corresponding year with respect to the research question. The column 'In Progress/Planning' specifies the nature of ongoing or upcoming explorations.

Research Objective	Research Question	Year 1	2	3	4	In Progress/In Planning
RO1: Auditing Preference Alignment	RQ1.1: Characterizing preference differences	[69]	[71]	—	—	—
	RQ1.2: Assessing RS performance for children	[69]	[71]	*P*	—	Long-term Explorations
	RQ1.3: Modelling long-term behavior	—	[68]	*P*	—	Develop long-term valid models for children
RO2: Auditing Non-Maleficence	RQ2.1: Characterizing harmful items	[30, 67]	*P*	*P*	—	Expansions to different types of harms
	RQ2.2: Documenting harmful recommendations	[67]	*P*	*P*	—	Expansions to different types of harms and differences between users
RO3: Designing holistic strategies	RQ3.1: Creating recommendation approaches	—	—	*P*	*P*	Design reranking and recommendation approaches
	RQ3.2: Evaluating holistic recommendation quality	—	—	—	*P*	Evaluate trade-offs between perspectives

mitigation strategies aimed at reducing harm.

Design recommendations approaches to achieve 'child-aware' recommender systems that truly serve children. To strive toward *RO3*, we will leverage findings from previous studies to design and evaluate recommendation strategies that attempt better to serve children **holistically**, taking into account the two previously considered perspectives. We plan to begin concrete explorations in year 3, to produce actionable outcomes in year 4. Understanding for which children current RS fail to create high-quality recommendations (*RO1*) and which recommendations can be harmful (*RO2*) provides a broad picture about *where* RS currently fail. We will explore methods that target these failures by reranking, adapting, or designing new recommendation approaches that explicitly aim to serve children's preferences better, but also acknowledge their vulnerabilities. Importantly, such strategies can only be implemented successfully if they account for the broader user base. Improving and safeguarding recommendations for children must not come at the expense of recommendation quality for adults. To this end, we will explore approaches that automatically detect content appropriateness and consider techniques such as multi-objective learning to balance needs and constraints of different user groups.

5 Concluding Remarks

This research project establishes a foundational understanding of what it means to truly serve children in RS—beyond accuracy—by framing recommendation quality around both preference alignment and non-maleficence. It introduces the notion of child-aware recommender systems, placing responsibility on RS to be aware of their user base and, consequently, their impact on society [cf. 35]. This paradigm treats children not as edge cases, but as a meaningful part of the user base with distinct needs, vulnerabilities, and rights. Through this lens, we develop evaluation strategies, probe current RS, and design strategies that lay the groundwork for ethically sound and developmentally appropriate RS for children.

Acknowledgments

I want to thank my supervisor, Sole Pera, for her invaluable guidance and support throughout my PhD journey. I am also grateful to my recent collaborators, Alejandro Bellogín, Dominik Kowald, and Delfina Martinez Pandiani, whose insights and contributions have not only shaped this research but have also played a significant role in my personal and academic development.

References

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30 (2020), 127–158.
[2] Albert Bandura. 2009. Social cognitive theory of mass communication. In *Media effects*. Routledge, 110–140.
[3] Albert Bandura and Richard H Walters. 1977. *Social learning theory*. Vol. 1. Englewood cliffs Prentice Hall.
[4] Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa-Anke, and Leonardo Neves. 2020. TweetEval:Unified Benchmark and Comparative Evaluation for Tweet Classification. In *Findings of EMNLP*.
[5] Christine Bauer, Eva Zangerle, and Alan Said. 2024. Exploring the landscape of recommender systems evaluation: Practices and perspectives. *ACM Transactions on Recommender Systems* 2, 1 (2024), 1–31.
[6] Laura Berk. 2015. *Child development*. Pearson Higher Education AU.
[7] Rebecca S Bigler and Lynn S Liben. 2007. Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current directions in psychological science* 16, 3 (2007), 162–166.
[8] Dania Bilal. 2010. The mediated information needs of children on the Autism Spectrum Disorder (ASD). In *Proceedings of the 31st ACM SIGIR Workshop on Accessible Search Systems, Geneva, Switzerland*. ACM Geneva, 42–49.
[9] Allison JB Chaney. 2021. Recommendation system simulations: A discussion of two key challenges. *SimuRec: Workshop on Synthetic Data and Simulation Methods for Recommender Systems Research, in conjunction with 15th ACM Conference on Recommender Systems (RecSys 2021), Available at https://doi.org/10.48550/arXiv.2109.02475* (2021).
[10] Allison JB Chaney, Brandon M Stewart, and Barbara E Engelhardt. 2018. How algorithmic confounding in recommendation systems increases homogeneity

and decreases utility. In *Proceedings of the 12th ACM conference on recommender systems*. 224–232.

[11] Vasiliki Charisi, Stephane Chaudron, Rosanna Di Gioia, Riina Vuorikari, Marina Escobar Planas, MARTIN Jose Ignacio Sanchez, and Emilia Gomez Gutierrez. 2022. *Artificial intelligence and the rights of the child: Towards an integrated agenda for research and policy.* Technical Report. Joint Research Centre (Seville site).

[12] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems* 41, 3 (2023), 1–39.

[13] Pia Christensen and Allison James. 2017. *Research with children.* Taylor & Francis.

[14] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, David Sculley, and Yoni Halpern. 2020. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 525–534.

[15] Michael Ekstrand. 2017. Challenges in evaluating recommendations for children. In *International Workshop on Children & Recommender Systems. Available at: https://md.ekstrandom.net/pubs/kidrec-eval-challenges.pdf*.

[16] Michael D Ekstrand. 2021. Multiversal Simulacra: Understanding Hypotheticals and Possible Worlds Through Simulation. *SimuRec: Workshop on Synthetic Data and Simulation Methods for Recommender Systems Research, in conjunction with 15th ACM Conference on Recommender Systems (RecSys 2021), Available at https://doi.org/10.48550/arXiv.2110.00811* (2021).

[17] Francesco Fabbri, Maria Luisa Croci, Francesco Bonchi, and Carlos Castillo. 2022. Exposure inequality in people recommender systems: the long-term effects. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 194–204.

[18] Nahid Ferdousi. 2015. Children as research subjects: The ethical issues. *Bangladesh Journal of Bioethics* 6, 1 (2015), 6–10.

[19] Andres Ferraro, Dmitry Bogdanov, Xavier Serra, and Jason Yoon. 2019. Artist and style exposure bias in collaborative filtering based music recommendations. *Workshop on Designing Human-Centric MIR Systems, in conjunction with 20th International Society for Music Information Retrieval Conference (ISMIR 2019), Available at: https://doi.org/10.48550/arXiv.1911.04827* (2019).

[20] Andres Ferraro, Michael D Ekstrand, and Christine Bauer. 2024. It's Not You, It's Me: The Impact of Choice Models and Ranking Strategies on Gender Imbalance in Music Recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 884–889.

[21] Andres Ferraro, Dietmar Jannach, and Xavier Serra. 2020. Exploring longitudinal effects of session-based recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 474–479.

[22] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the loop: Gender imbalance in music recommenders. In *Proceedings of the 2021 conference on human information interaction and retrieval*. 249–254.

[23] Yacine Gaci, Boualem Benatallah, Fabio Casati, and Khalid Benabdeslem. 2022. Masked Language Models as Stereotype Detectors?. In *EDBT 2022*.

[24] Myrsini Gkolemi, Panagiotis Papadopoulos, Evangelos Markatos, and Nicolas Kourtellis. 2022. YouTubers Not MadeForKids: Detecting channels sharing inappropriate videos targeting children. In *Proceedings of the 14th ACM Web Science Conference 2022*. 370–381.

[25] Jennifer Golbeck. 2025. Recommender System-Induced Eating Disorder Relapse: Harmful Content and the Challenges of Responsible Recommendation. *ACM Transactions on Intelligent Systems and Technology* 16, 1 (2025), 1–13.

[26] Emilia Gómez Gutiérrez, Vicky Charisi, and Stephane Chaudron. 2021. Evaluating recommender systems with and for children: towards a multi-perspective framework. In *CEUR Workshop Proceedings. 2021; 2955*. CEUR Workshop Proceedings.

[27] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2012. Evaluating recommender systems. In *Recommender systems handbook*. Springer, 547–601.

[28] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.

[29] Naieme Hazrati and Francesco Ricci. 2022. Simulating users' interactions with recommender systems. In *Adjunct proceedings of the 30th acm conference on user modeling, adaptation and personalization*. 95–98.

[30] Jasper Heijne, Robin Ungruh, and Maria Soledad Pera. 2025. Blurred Lines: Understanding The Fit of Song Lyrics in Music Catalogs That Can Reach Children Through Recommendations. *2nd workshop on Information Retrieval for Understudied Users (IR4U2), co-located with ACM SIGIR 2025* (2025).

[31] Suzanne Hidi and K Ann Renninger. 2006. The four-phase model of interest development. *Educational psychologist* 41, 2 (2006), 111–127.

[32] Chih-Wei Hsu, Martin Mladenov, Ofer Meshi, James Pine, Hubert Pham, Shane Li, Xujian Liang, Anton Polishko, Li Yang, Ben Scheetz, et al. 2024. Minimizing live experiments in recommender systems: User simulation to evaluate preference elicitation policies. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2925–2929.

[33] D Jannach. 2010. *Recommender Systems: An Introduction*. Cambridge University Press.

[34] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction* 25 (2015), 427–491.

[35] Raoul Kalisvaart, Masoud Mansoury, Alan Hanjalic, and Elvin Isufi. 2025. Towards Carbon Footprint-Aware Recommender Systems for Greener Item Recommendation. *ACM Transactions on Recommender Systems* (2025).

[36] Loretta M Kopelman. 2000. Children as research subjects: a dilemma. *The Journal of medicine and philosophy* 25, 6 (2000), 745–764.

[37] Natalia Kucirkova. 2019. The Learning Value of Personalization in Children's Reading Recommendation Systems: What Can We Learn From Constructionism? *International Journal of Mobile and Blended Learning (IJMBL)* 11, 4 (2019), 80–95.

[38] Monica Landoni, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 2024. Good for Children, Good for All?. In *European Conference on Information Retrieval*. Springer, 302–313.

[39] Oleg Lesota, Jonas Geiger, Max Walder, Dominik Kowald, and Markus Schedl. 2024. Oh, Behave! Country Representation Dynamics Created by Feedback Loops in Music Recommender Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 1022–1027.

[40] Kun Lin, Nasim Sonboli, Bamshad Mobasher, and Robin Burke. 2019. Crank up the volume: preference bias amplification in collaborative recommendation. *arXiv preprint arXiv:1909.06362* (2019).

[41] Sonia Livingstone, Giovanna Mascheroni, and Mariya Stoilova. 2023. The outcomes of gaining digital skills for young people's lives and wellbeing: A systematic evidence review. *New media & society* 25, 5 (2023), 1176–1202.

[42] Sonia Livingstone, Kjartan Ólafsson, and Elisabeth Staksrud. 2011. Social networking, age and privacy. (2011).

[43] Gianclaudio Malgieri and Jędrzej Niklas. 2020. Vulnerable data subjects. *Computer Law & Security Review* 37 (2020), 105415.

[44] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 2145–2148.

[45] Ashlee Milton, Levesson Batista, Garrett Allen, Siqi Gao, Yiu-Kai D Ng, and Maria Soledad Pera. 2020. "Don't judge a book by its cover": Exploring book traits children favor. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 669–674.

[46] Ashlee Milton, Emiliana Murgia, Monica Landoni, Theo Huibers, and Maria Soledad Pera. 2019. Here, there, and everywhere: Building a scaffolding for children's learning through recommendations. (2019).

[47] Emiliana Murgia, Monica Landoni, Theo Huibers, Jerry Alan Fails, and Maria Soledad Pera. 2019. The seven layers of complexity of recommender systems for children in educational contexts. (2019).

[48] National Center for Education Statistics. 2023. Children's Internet Access at Home. Condition of Education. https://nces.ed.gov/programs/coe/indicator/cch Retrieved April 9, 2024.

[49] Yiu-Kai Ng. 2017. MovRec: a personalized movie recommendation system for children based on online movie features. *International Journal of Web Information Systems* 13, 4 (2017), 445–470.

[50] Yiu-Kai Ng. 2023. Read to grow: exploring metadata of books to make intriguing book recommendations for teenage readers. *Knowledge and Information Systems* (2023), 1–26.

[51] UK OfCom. 2023. Children and parents: Media use and attitudes report 2023. *London: Office of Communications London* (2023).

[52] Kostantinos Papadamou, Antonis Papasavva, Savvas Zannettou, Jeremy Blackburn, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Michael Sirivianos. 2020. Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 522–533.

[53] Maria Soledad Pera and Yiu-Kai Ng. 2014. Automating readers' advisory to make book recommendations for k-12 readers. In *Proceedings of the 8th ACM Conference on Recommender Systems*. 9–16.

[54] Pew Research Center. 2024. Teens and internet, device access fact sheet. https://www.pewresearch.org/internet/fact-sheet/teens-and-internet-device-access-fact-sheet/ Retrieved April 9, 2024.

[55] Jenny Radesky, Enrica Bridgewater, Shira Black, August O'Neil, Yilin Sun, Alexandria Schaller, Heidi M Weeks, and Scott W Campbell. 2024. Algorithmic Content Recommendations on a Video-Sharing Platform Used by Children. *JAMA Network Open* 7, 5 (2024), e2413855–e2413855.

[56] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2021. Recommender systems: Techniques, applications, and challenges. *Recommender systems handbook* (2021), 1–35.

[57] Mario Rodriguez, Christian Posse, and Ethan Zhang. 2012. Multiple objective optimization in recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems*. 11–18.

[58] Almudena Ruiz-Iniesta, Luis Melgar, Alejandro Baldominos, and David Quintana. 2018. Improving children's experience on a mobile EdTech platform through a recommender system. *Mobile Information Systems* 2018, 1 (2018), 1374017.

[59] Markus Schedl and Christine Bauer. 2017. Online music listening culture of kids and adolescents: Listening analysis and music recommendation tailored to the young. In *1st International Workshop on Children and Recommender Systems, in conjunction with 11th ACM Conference on Recommender Systems (RecSys 2017)- Available at: https://doi.org/10.48550/arXiv.1912.11564*.
[60] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekabsaz. 2022. LFM-2b: A dataset of enriched music listening events for recommender systems research and fairness analysis. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 337–341.
[61] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic multi-aspect fairness through personalized re-ranking. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 239–247.
[62] Lawrence Spear, Ashlee Milton, Garrett Allen, Amifa Raj, Michael Green, Michael D Ekstrand, and Maria Soledad Pera. 2021. Baby shark to barracuda: Analyzing children's music listening behavior. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 639–644.
[63] Mariya Stoilova, Sonia Livingstone, Rana Khazbak, et al. 2021. Investigating Risks and Opportunities for Children in a Digital World: A rapid review of the evidence on children's internet use and outcomes. (2021).
[64] Tiffany Ya Tang and Pinata Winoto. 2016. I should not recommend it to you even if you will like it: the ethics of recommender systems. *New Review of Hypermedia and Multimedia* 22, 1-2 (2016), 111–138.
[65] Konstantinos Tsiakas, Emilia Barakova, Javed Vassilis Khan, and Panos Markopoulos. 2020. 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*. 1–6.
[66] Virginia Tsintzou, Evaggelia Pitoura, and Panayiotis Tsaparas. 2018. Bias disparity in recommendation systems. *arXiv preprint arXiv:1811.01461* (2018).
[67] Robin Ungruh, Murtadha Al Nahadi, and Maria Soledad Pera. 2025. Mirror, Mirror: Exploring Stereotype Presence Among Top-N Recommendations That May Reach Children. *ACM Transactions on Recommender Systems* (2025).
[68] Robin Ungruh, Alejandro Bellogín, and Maria Soledad Pera. 2025. From Monolith to Mosaic: Uncovering Behavioral Differences for Choice Models in Recommender Systems Simulations. In *Proceedings of the 48th international ACM SIGIR conference on research and development in information retrieval*.
[69] Robin Ungruh, Alejandro Bellogín, and Maria Soledad Pera. 2025. The Impact of Mainstream-Driven Algorithms on Recommendations for Children. In *European Conference on Information Retrieval*. Springer, 67–84.
[70] Robin Ungruh and Maria Soledad Pera. 2024. Ah, that's the great puzzle: On the Quest of a Holistic Understanding of the Harms of Recommender Systems on Children. *DCDW co-located with ACM IDC 2024 - Available at: https://doi.org/10.48550/arXiv.2405.02050* (2024).
[71] Robin Ungruh et al. 2025. The Impact of Mainstream-Driven Algorithms on Recommendations for Children: A Reproducibility Study. (Under Review)).
[72] UNICEF. 2020. How many children and young people have internet access at home? (2020).
[73] Gabriel Vigliensoni and Ichiro Fujinaga. 2017. The music listening histories dataset.. In *ISMIR*. 96–102.
[74] Mengting Wan and Julian McAuley. 2018. Item recommendation on monotonic behavior chains. In *Proceedings of the 12th ACM conference on recommender systems*. 86–94.
[75] Mengting Wan, Rishabh Misra, Ndapa Nakashole, and Julian McAuley. 2019. Fine-grained spoiler detection from large-scale review corpora. *arXiv preprint arXiv:1905.13416* (2019).
[76] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–43.
[77] Jingjing Zhang, Gediminas Adomavicius, Alok Gupta, and Wolfgang Ketter. 2020. Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Information Systems Research* 31, 1 (2020), 76–101.
[78] Yong Zheng and David Xuejun Wang. 2022. A survey of recommender systems with multi-objective optimization. *Neurocomputing* 474 (2022), 141–153.
[79] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 22–32.