Knight–Georgetown Institute

Better Feeds: Algorithms That Put People First

A How-To Guide for Platforms and Policymakers

March 2025

Knight Georgetown Institute

Session Logistics

How-to guide is available at:
https://kgi.georgetown.edu/research-and-commentary/better-feeds/

The meeting is being recorded. The recording and slides will be available at:
https://kgi.georgetown.edu/events/better-feeds-report-launch-webinar/

Please add questions to the Q&A panel throughout the presentation!

If you need help, use the chat or send email:
kgi-media@georgetown.edu.

Agenda

1. Policy landscape
2. Recommender systems 101
3. *Better Feeds* policy guidance
4. Q&A

About KGI

The Knight-Georgetown Institute (KGI) is a new center at Georgetown University dedicated to connecting independent research with technology policy and design.

As part of its research translation efforts, KGI convenes expert working groups that bring together relevant experts from across academia, industry, civil society, journalism, and practitioner communities to summarize knowledge and articulate policy options.

KGI Expert Working Group on Recommender Systems

Alex Moehring
Purdue University

Alissa Cooper
Knight-Georgetown Institute

Arvind Narayanan
Princeton University

Aviv Ovadya
AI & Democracy Foundation

Elissa Redmiles
Georgetown University

Jeff Allen
Integrity Institute

Jonathan Stray
University of California, Berkeley

Julia Kamin
Prosocial Design Network

Leif Sigerson
Integrity Institute

Luke Thorburn
King's College London

Matt Motyl
Psychology of Technology Institute, University of Southern California

Motahhare Eslami
Carnegie Mellon University

Nadine Farid Johnson
Knight First Amendment Institute at Columbia University

Nathaniel Lubin
Berkman Klein Center, Harvard University

Ravi Iyer
University of Southern California, Neely Center

Zander Arnao
Knight-Georgetown Institute

Policy Landscape

US Policy Efforts to Address Algorithmic Harms

35

States that introduced bills in 2023-2024 aiming to address algorithmic harms related to social media

75+

Number of state bills introduced in 2023-24 meant to address social media algorithms

500+

Number of lawsuits brought on behalf of children, families, school districts, municipalities, and Attorneys General alleging algorithmic harms

EU Digital Services Act

Article 38

Recommender systems

In addition to the requirements set out in Article 27, providers of very large online platforms and of very large online search engines that use recommender systems shall provide at least one option for each of their recommender systems which is not based on profiling as defined in Article 4, point (4), of Regulation (EU) 2016/679.

Plus transparency, accountability, and risk assessment provisions.

The False Choice

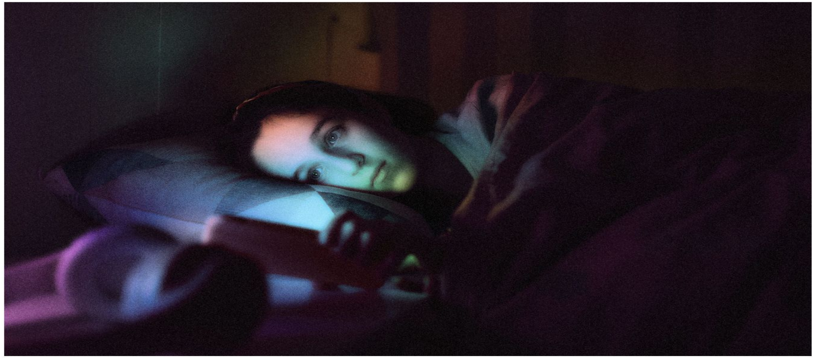
Oklahoma Social Media Transparency Act of 2023

A social media platform must “allow a user to opt out of post-prioritization and shadow banning algorithm categories to allow sequential or chronological posts and content.”

06-07-2024 | DESIGN

We’re about to glimpse life on the other side of algorithms

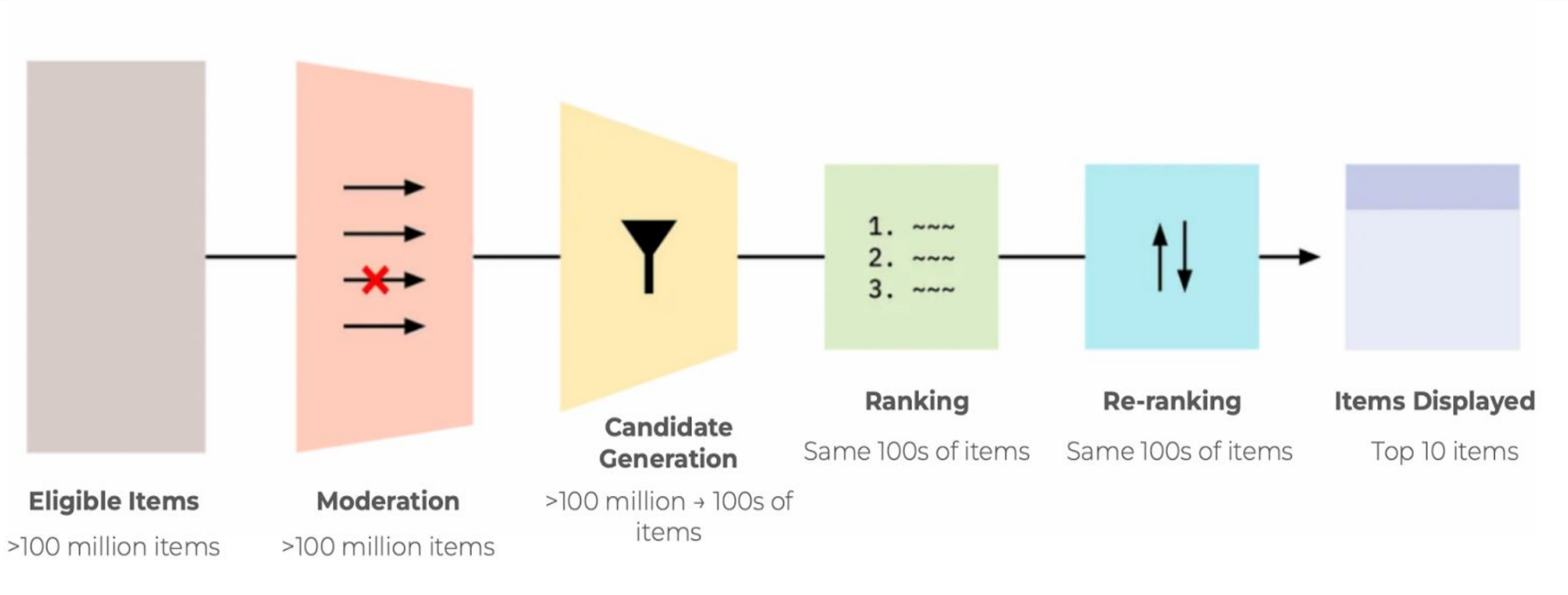
For the past decade, social media companies have used algorithms to puppeteer our digital lives. That was always the wrong idea. Now the government is giving us the chance to opt out. Will we take it?



[Source Photo: Israel Sebastian/Getty Images]

Understanding Recommender System Design and How to Make It Better

Recommender Systems 101



Engagement, defined

Engagement (noun)

en·gage·ment

Actions taken by users on recommended items, such as clicks, likes, comments, reposts, watch time, dwell time, upvote, downvote, and many others.

Item (noun)

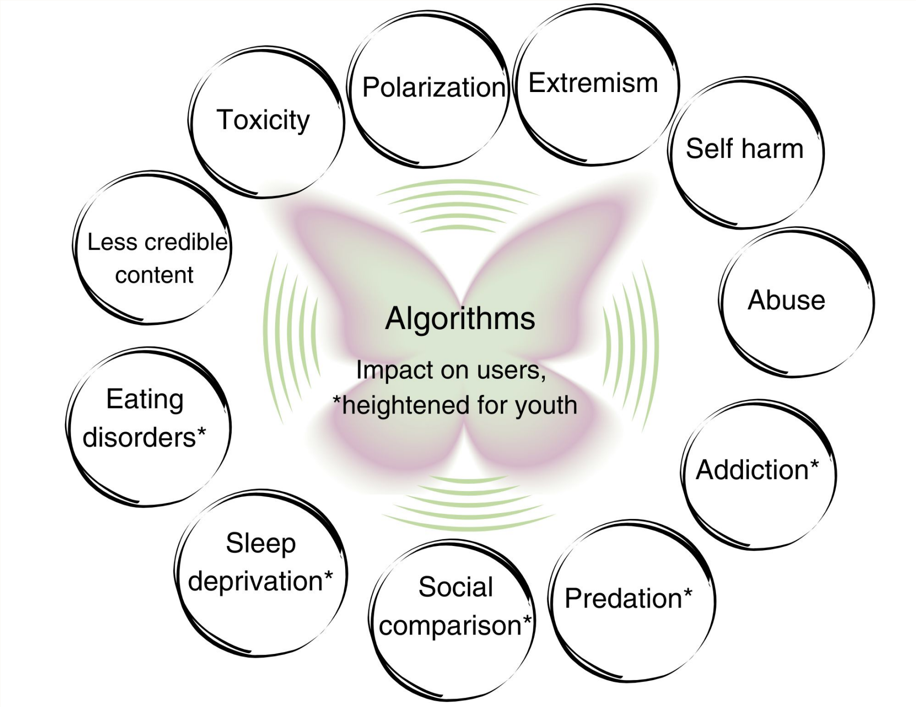
it·em

An element eligible for display by a recommender system. Items can include individual pieces of content, accounts, groups, pages, channels, products, or ads.

The Potato Chip Problem

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. “The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization.” February 23, 2022. https://arxiv.org/abs/2202.11776v3.

Harms Associated with Algorithms



A Path Forward

Long-Term User Value

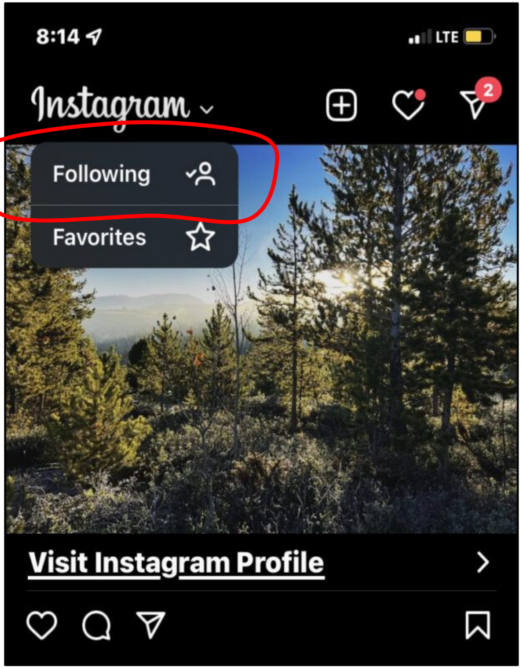
Outcomes aligned with the deliberate, forward-looking preferences and aspirations of users.

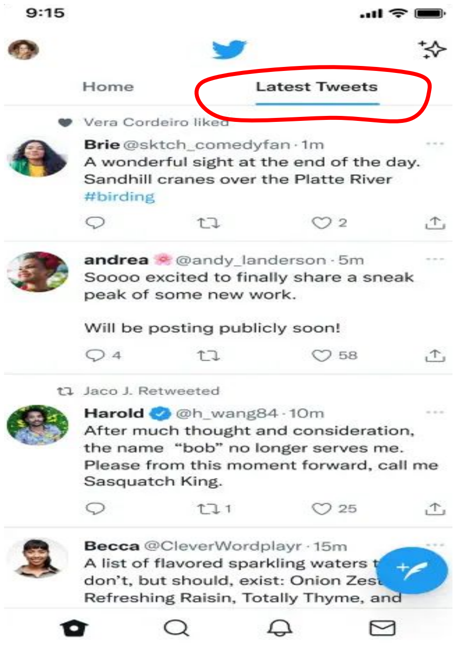
Designs consistent with this approach may:

- ask users directly to state their explicit preferences;
- rely on surveys, quality indicators selected by the user, or predictions of each;
- rely on signals that are deliberative, clear, or onerous;
- or a combination thereof.

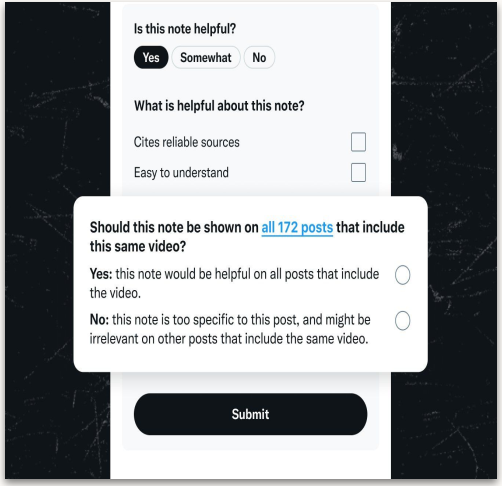
Alternatives to Maximizing Engagement

Chronological and Non-Personalized Feeds

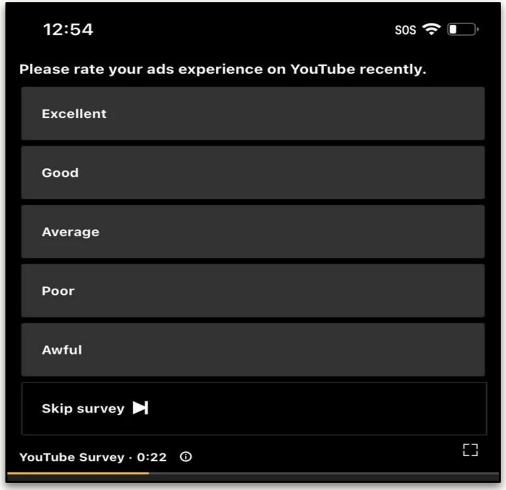




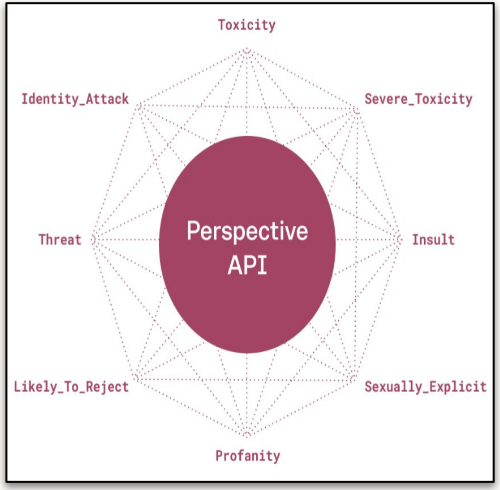
Better Approaches



Bridging

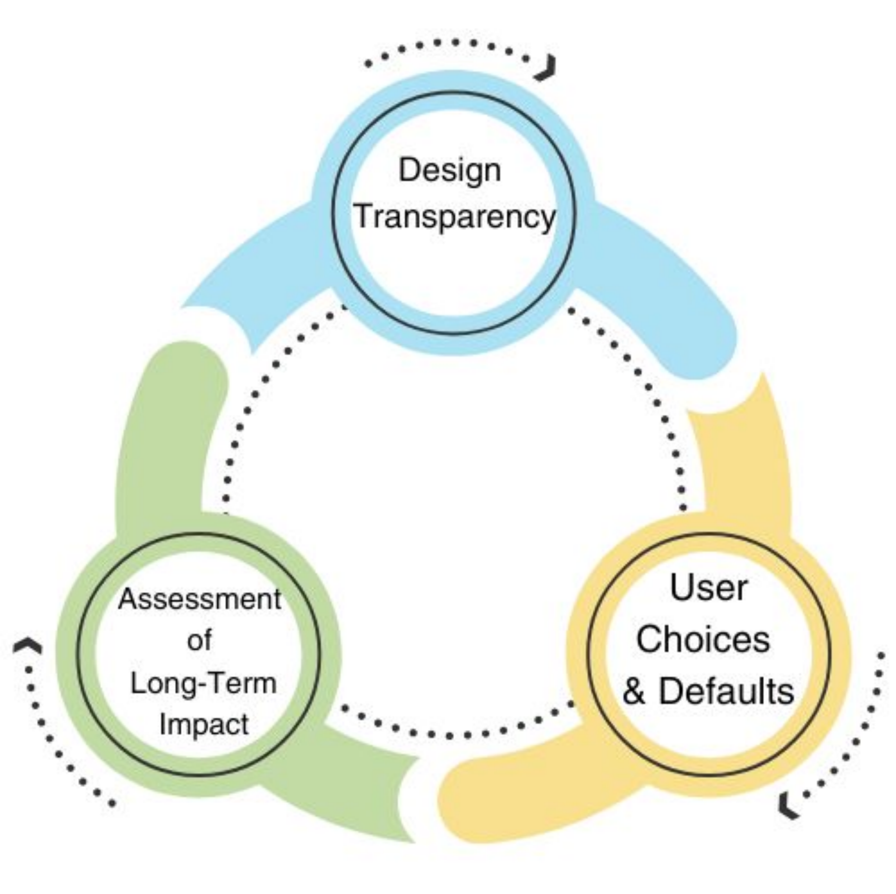


Surveys

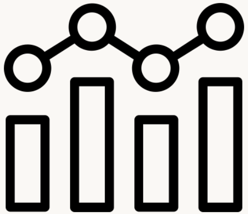


Quality

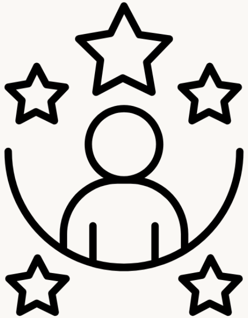
Core Policy Guidance


Design Transparency
User Choices & Defaults
Assessment of Long-Term Impact

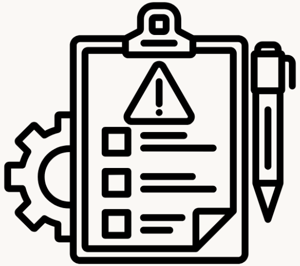
Design Transparency



Input data sources and weights



Metrics used to measure long-term user value



Metrics used to evaluate teams responsible for recommender systems

Existing Design Transparency

How likely you are to spend more than 10 seconds on the first post

Signals influencing this prediction include:

- How much time you've spent viewing the author's posts
- How many videos you've viewed
- How many times you've viewed the author's posts
- Which device platform you're using to view content, such as the web, Android or iOS

Instagram Feed AI System, https://transparency.meta.com/features/explaining-ranking/ig-feed/

Input Data Sources and Weights

All sources of raw information used in ranking

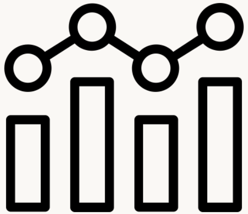
Including:

- item content and metadata
- engagement history
- user survey data
- quality feedback from users
- annotations from raters
- user settings
- profile and social graph data
- context data (day, time, location)

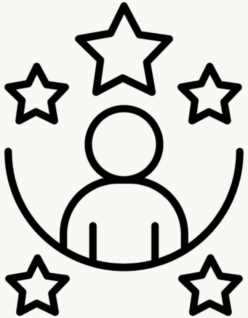
Values and their weights

- Weights reveal which values have greater or lesser impact on ranking.
- Report the complete list of values and their weights for the system as a whole.
- Report the quartile of each weight.

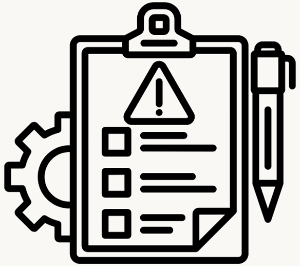
Design Transparency



Input data sources and weights

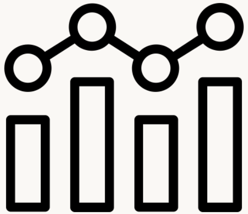


Metrics used to measure long-term user value

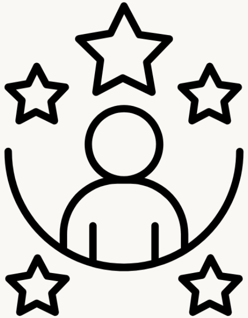


Metrics used to evaluate teams responsible for recommender systems

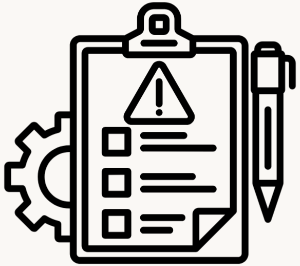
Design Transparency



Input data sources and weights



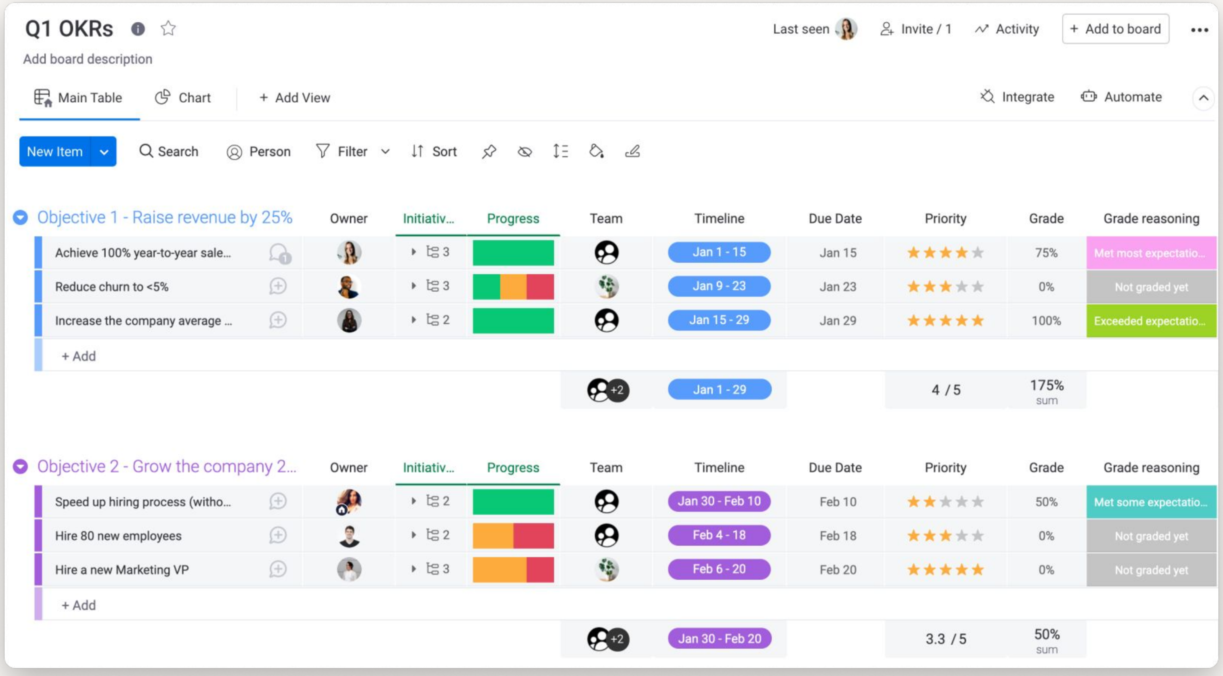
Metrics used to measure long-term user value



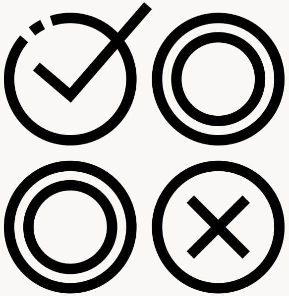
Metrics used to evaluate teams responsible for recommender systems

Metrics Used to Evaluate Teams

Objectives and Key Results (OKRs)



User Choices and Defaults



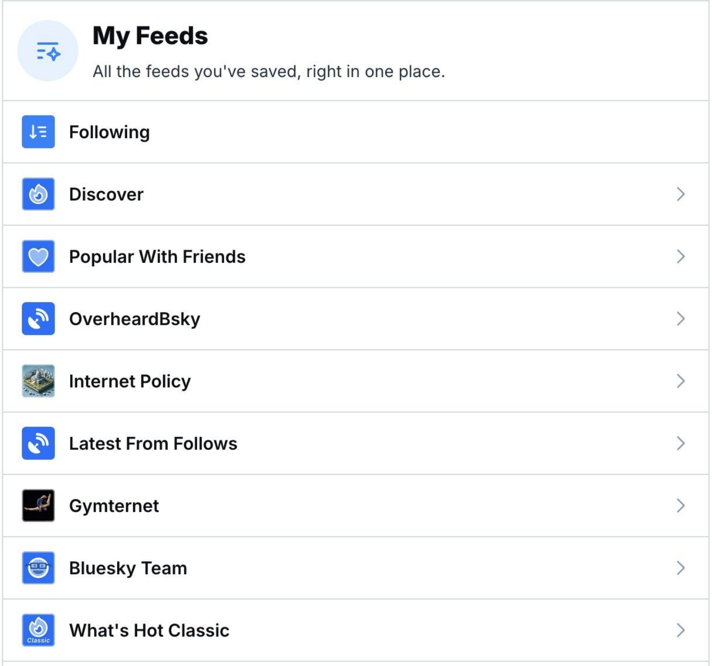
Easily accessible choice of recommender systems, at least one optimized to support long-term value

Created by aristeles
from Noun Project

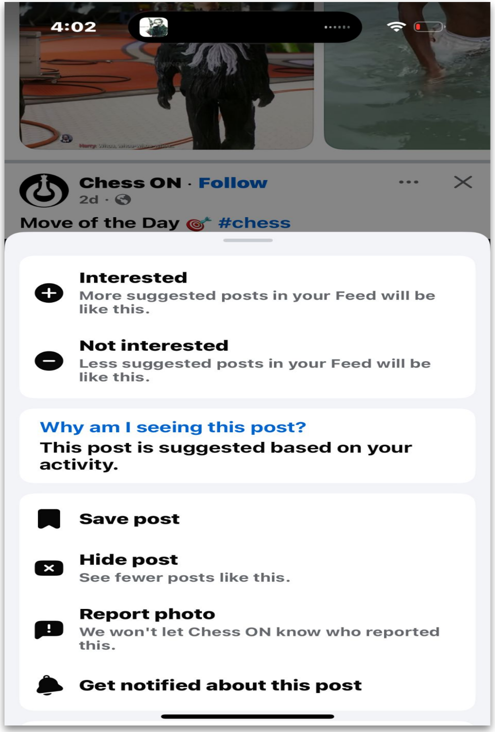
Honor users' preferences concerning recommended or blocked items

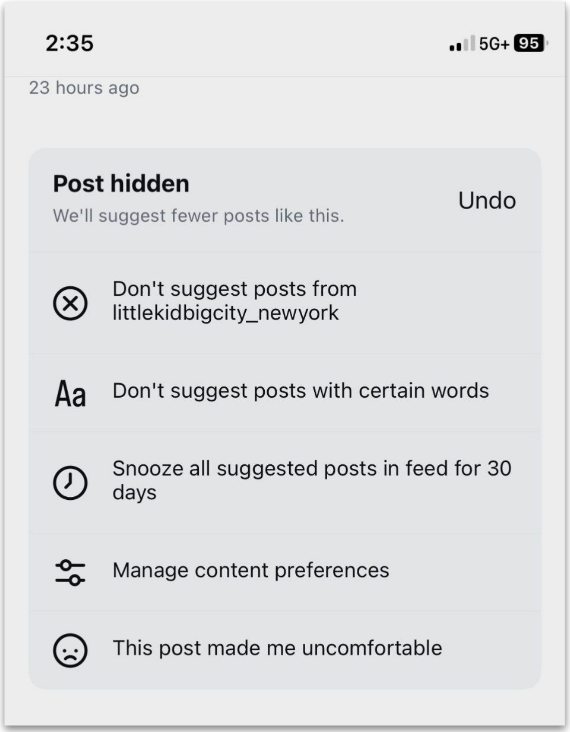
Set minors' recommender systems to be optimized to support long-term value by default

Easily Accessible Choice

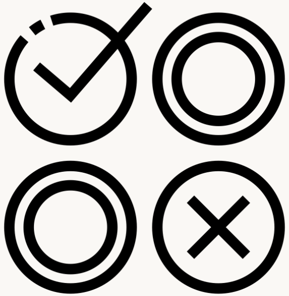


Honor Users' Preferences





User Choices and Defaults



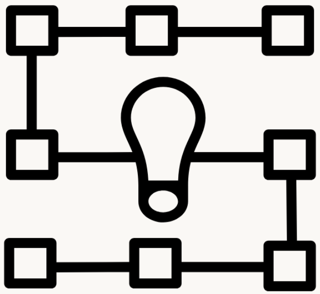
Easily accessible choice of recommender systems, at least one optimized to support long-term value

Created by aristeles
from Noun Project

Honor users' preferences concerning recommended or blocked items

Set minors' recommender systems to be optimized to support long-term value by default

Assessments of Long-Term Impact

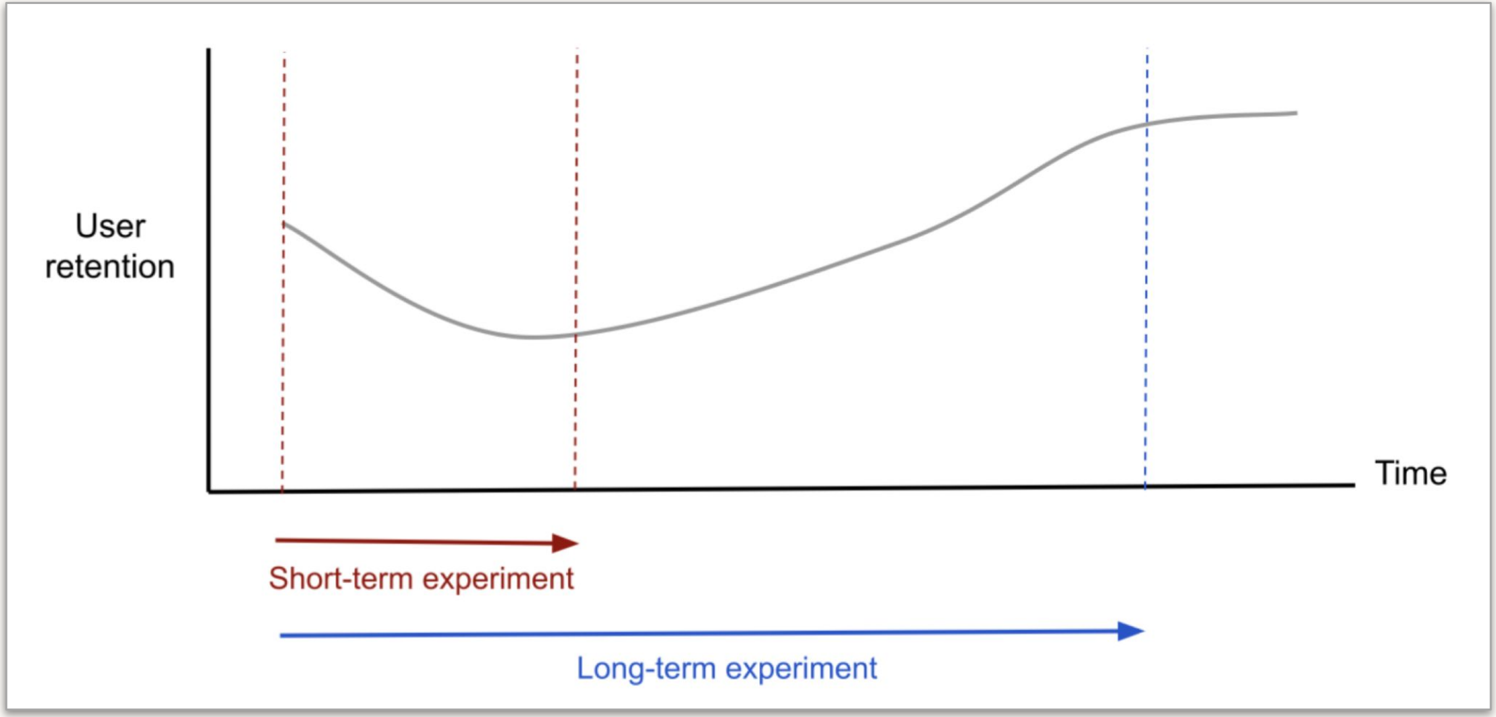


Run long-term (12-month or longer) holdout experiments on a continuous basis

Report the aggregate, anonymized results of the holdout experiments publicly

Subject to an audit by an independent third party

Short-Term vs. Long-Term Experiments



Long-Term Holdout Groups

How holdout groups drive sustainable growth

Pinterest Engineering · Follow
Published in Pinterest Engineering Blog · 3 min read · Feb 13, 2015

40 1

John Egan | Pinterest tech lead, Growth

When it comes to growth, one potential pitfall is over optimizing for short-term wins. Growth teams operate at a pretty fast pace, and our team is no exception. We're always running dozens of experiments at any given time, and once we find something that works, we ship it and move on to the next experiment. However, sometimes it's important to take a step back and validate that a new tweak or feature really delivers long-term sustainable growth and isn't just a short-term win that users will get tired of after prolonged exposure. In this post I'll cover how we optimize for long-term sustainable growth.

Exposure to Marginally Abusive Content on Twitter

Proceedings of the Seventeenth International AAAI Conference on Web and Social Media (ICWSM 2023), Forthcoming

10 Pages • Posted: 16 Aug 2022

Jack Bandy
Northwestern University

Tomo Lazovich
Northeastern University - Northeastern University, School of Law, Students; U.S. Census Bureau

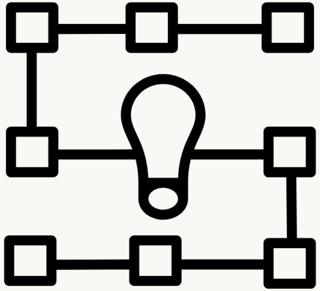
Universal Holdout Groups at Disney Streaming

Tian Yang · Follow
Published in disney-streaming · 7 min read · Oct 22, 2021

310 4

At Disney Streaming, we strive to make quality decisions about which features to ship based on the results of rigorous A/B experiments, or online randomized control trials.

Assessments of Long-Term Impact



Run long-term (12-month or longer) holdout experiments on a continuous basis

Report the aggregate, anonymized results of the holdout experiments publicly

Subject to an audit by an independent third party

Additional Global Policy Guidance

Public Content Disclosures

Continuously publish sample of popular content and representative sample of typical user session.

Strong Defaults

Optimize default recommender system to support long-term user value.

Report Aggregate Harms

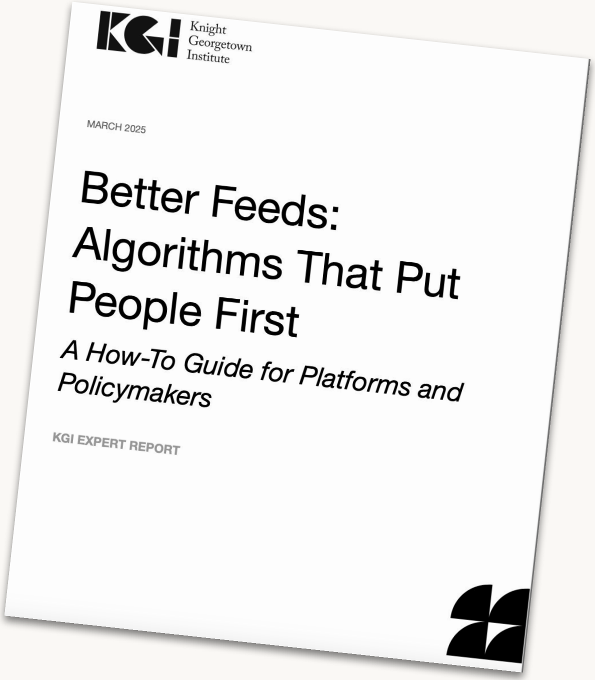
Measure and report aggregate harms to at-risk populations.

What's Next?

☑️ Reach out! We are happy to engage with policymakers and product teams.
Email alissa.cooper@georgetown.edu

☑️ KGI is developing:

- Modular language to inform legislation
- Mapping of guidelines onto DSA implementation
- Collection of examples where the guidelines are implemented in practice



Q
A

Thank *You*