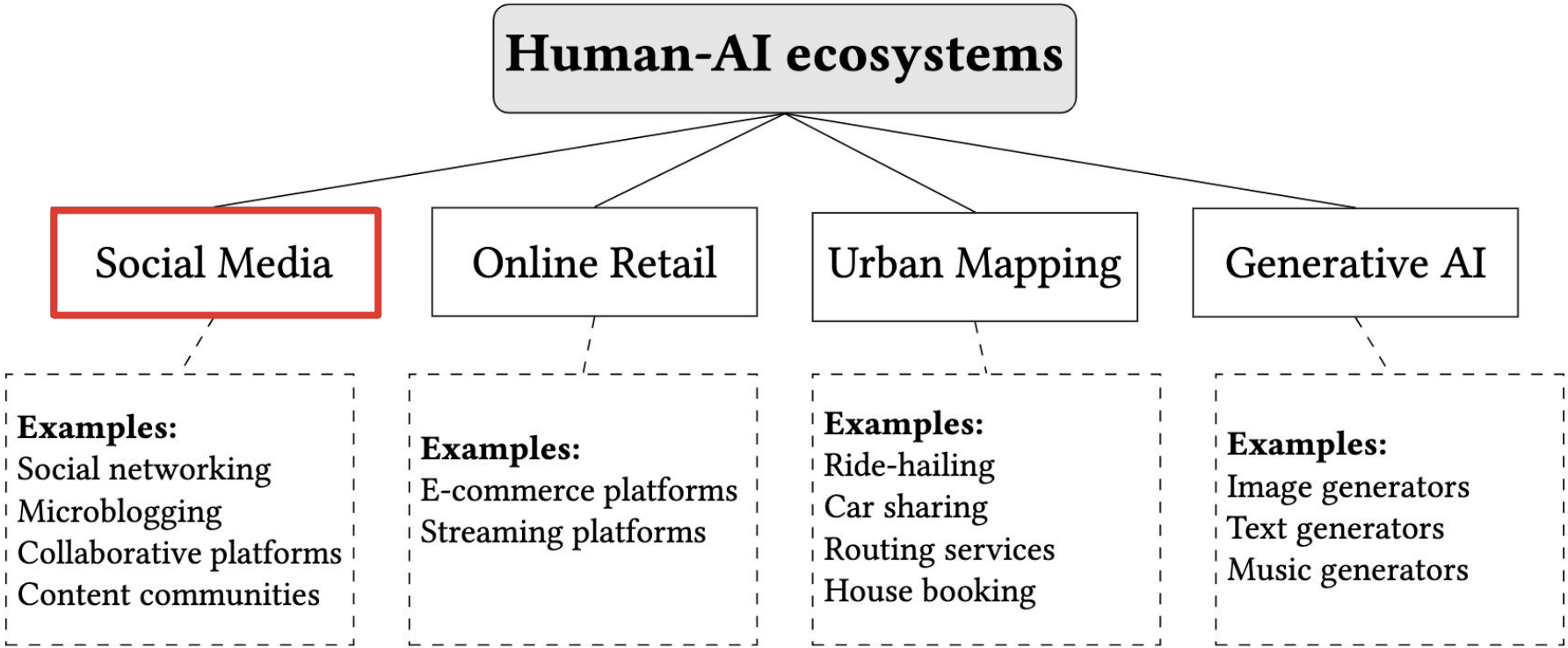

Human-AI ecosystems
Social Media
Online Retail
Urban Mapping
Generative AI
Examples:
Social networking
Microblogging
Collaborative platforms
Content communities
Examples:
E-commerce platforms
Streaming platforms
Examples:
Ride-hailing
Car sharing
Routing services
House booking
Examples:
Image generators
Text generators
Music generators

SOCIAL MEDIA

What do *recommenders* suggest to users in social media?

content and **social links**

Designated VLOPs

https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses#ecl-inpage-Infinite

updated to February 6th, 2025

facebook

Snapchat

TikTok

Pinterest

LinkedIn

YouTube

Discussion

What is the business model of social media platforms? What do they optimise for?

- **User Engagement**: More time spent on the platform = more data and ads
- **Data Collection**: Personal data is used to refine algorithms and target ads
- **Growth**: Platforms grow their user base to increase value for advertisers

> The thought process that went into building these applications, Facebook being the first of them, ... was all about: **How do we consume as much of your time and conscious attention as possible?**...And that means that we need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post or whatever. And that's going to get you to contribute more content, and that's going to get you ... more likes and comments. *It's a social-validation* ***feedback loop*** ... exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology.

Sean Parker, first president of Facebook, 2017

https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792

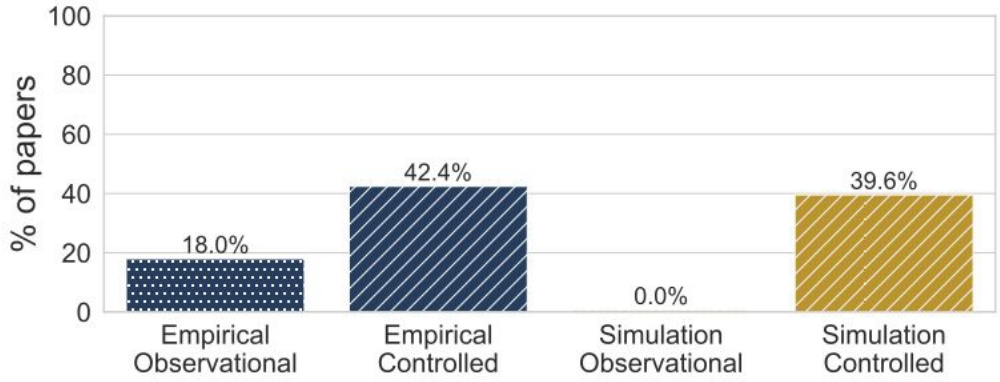
- XCheck **whitelists VIP users**, allowing them to post rule-violating material (e.g., harassment or incitement to violence)
- Instagram is **toxic for teen girls**, increasing anxiety and depression
- The new algorithm introduced in 2018 **made people angrier**
- Facebook **suppresses political movements** it deems dangerous

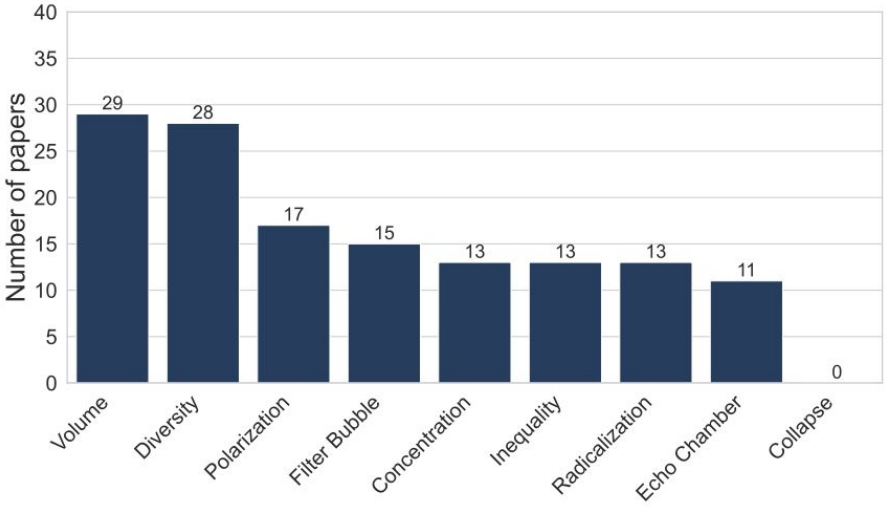
"You see a theme in all these documents that Facebook and its top executives know what their problems are, but in many instances, can't, or won't address them sometimes because it fears hurting the business or growth."

The Facebook Files Podcast, The Wall Street Journal

Experiments and outcomes

L. Pappalardo et al. A survey on the impact of AI-based recommenders on human behaviours, 2024, https://doi.org/10.48550/arXiv.2407.01630





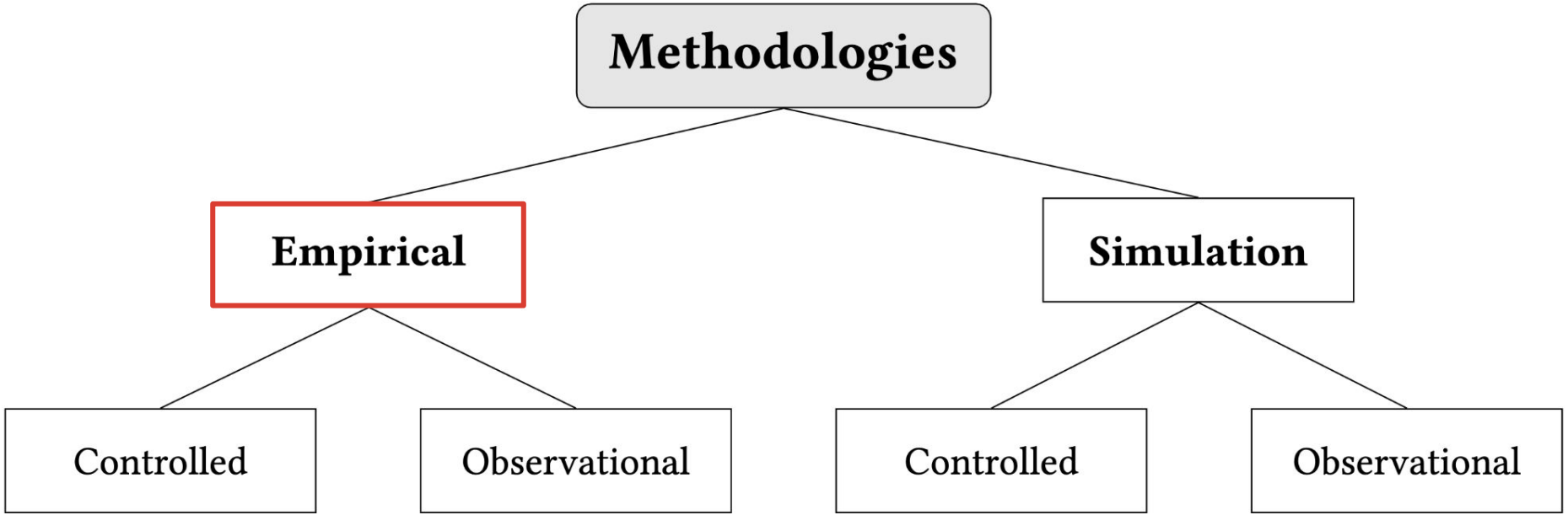
Level	Outcome	Description	Ecosystems
Individual	Diversity	Variety of users' behaviour, items consumed and users followed	SM, OR, UM
	Filter Bubble	Conformation of items or contents with own preferences or beliefs	SM, OR
	Radicalization	Items or individual attributes going towards an extreme	SM
	Volume	Quantity value of some users' attribute	SM, OR, UM
Item	Diversity	Variety of users that consume the item	SM, OR, GAI
	Volume	Quantity value of some items' attribute	SM, OR, UM
~~**Model**~~	~~Collapse~~	~~AI model degradation over time~~	~~GAI~~
Systemic	Concentration	Close gathering of people or things	SM, OR, UM
	Diversity	Aggregate diversity of users or items	SM, OR, UM
	Echo Chamber	Environment reinforcing opinions or item choices within a group	SM, OR, UM
	Inequality	Uneven distribution of resources/opportunities among group members	SM, OR, UM
	Polarization	Sharp separation of users/items into groups based on some attributes	SM
	Volume	Aggregate volume of users' or items' attributes	SM, OR, UM

Social Media		Empirical		Simulation	
		Observational	Controlled	Observational	Controlled
Individual	Filter Bubble	[14, 22, 25, 30, 33, 65, 72, 76, 80, 92, 93, 162]	[17]	[140, 142]	
	Radicalization	[13, 25, 72, 79, 80, 83, 94, 139, 160]	[113]	[85, 140, 142, 157]	
Model	Collapse				
Systemic	Concentration	[76, 87, 94, 153]	[16, 113]	[41, 51, 52, 56, 134, 135, 137]	
	Echo Chamber	[14, 25]	[126]	[32, 34, 128, 129, 135, 137, 157, 159]	
	Inequality	[76, 87, 94, 145, 153]	[16, 82, 113]	[51, 52, 56, 135, 137]	
	Polarization	[33, 65, 162]	[67, 68, 102, 106, 126]	[32, 41, 128, 129, 134, 135, 137, 138, 150, 159]	
Individual Item Systemic	Diversity	individual: [14, 22, 25, 30, 33, 65, 72, 92, 162], item: [14, 19, 122, 169]	individual: [17], item: [102, 113, 126],	individual: [41, 62, 138, 140, 157], systemic: [51, 52, 56, 85, 135, 137, 142]	
	Volume	individual: [83], item: [13, 19, 22, 25, 72, 76, 87, 94, 145, 153, 160, 169]	individual: [16, 17, 67, 68, 102], item: [16, 67, 68, 82, 126], systemic: [163]	individual: [51, 52, 56, 135, 137, 139], systemic: [62]	

Selected studies:

- [82] Huszár et al. 2021
- [80] Hosseinmardi et al. 2023

Examples on Social Media



[82] Huszár et al. 2021

[80] Hosseinmardi et al. 2023

Amplification of politics on Twitter

Huszár et al., PNAS 2021, https://doi.org/10.1073/pnas.2025334119

Type:	**Empirical controlled**
VLOP:	**Twitter**
Outcomes:	**volume (increase)**
	inequality (increase)

POLITICS

When Twitter users hear out the other side, they become more polarized

Echo chambers aren't what's polarizing America.

by **Ezra Klein**

Oct 18, 2018, 2:30 PM GMT+2

Amplification of politics on Twitter

Huszár et al., PNAS 2021 **Empirical controlled**

- Original Twitter's recommender: users obtain content from accounts they followed in **reverse chronological order**
- In 2016, a ***content-based filtering*** recommender was introduced:
  - users see tweets deemed relevant
    (both older ones and from accounts they do not follow)

> Here's how it works. You flip on the feature in your settings; then when you open Twitter after being away for a while, the Tweets you're most likely to care about will appear at the top of your timeline – still recent and in reverse chronological order. The rest of the Tweets will be displayed right underneath, also in reverse chronological order, as always. At any point, just pull-to-refresh to see all new Tweets at the top in the live, up-to-the-second experience you already know and love.

https://blog.x.com/en_us/a/2016/never-miss-important-tweets-from-people-you-follow

Amplification of politics on Twitter

Huszár et al., PNAS 2021

content-based filtering, aka
“personalized Home Timeline”

Does Twitter’s recommender systematically prioritize certain political content * by giving them greater visibility in users' feeds and recommendations?

* such as left vs. right, center vs. extremes, specific parties, or news sources with particular ideological leanings

Experimental Setup

Huszár et al., PNAS 2021

The “Timelines Quality Holdback Experiment”:

- **Control group**: **1%** of global users (randomly chosen) excluded from the personalized Home Timeline
  - which still displays tweets in reverse chronological order
- **Treatment group**: **4%** of users (randomly chosen) that experience the personalised Home Timeline
- This assignment is maintained over the lifespan of accounts

Tens of millions of users (including dormant accounts and bots)

Limitations of the study

Huszár et al., PNAS 2021

The experiment violates SUTVA (Stable Unit Treatment Value Assumption):

- the control group is not isolated from indirect effects of personalization
- the experiment cannot provide unbiased estimates of causal quantities

The study just present findings based on simple comparison of measurements between the treatment and control groups

Measuring amplification

Huszár et al., PNAS 2021

The **reach** of a set *T* of tweets in a set *U* of users is the total number of users in *U* who *encountered** a tweet from *T*

Example:

- *T* can be the set of tweets from politicians of Socialist Party in France
- *U* can be the set of French Twitter users in the control group
- the reach of *T* is how many of French users in the control group encounter tweets from the politicians in the Socialist Party

* A tweet is encountered by a user A when 50% of the UI element containing the tweet is continuously visible on the users' device for 500ms

Measuring amplification

Huszár et al., PNAS 2021

The **amplification ratio** of a set T of tweets is defined as:

$$\frac{\text{reach of T in the treatment group}}{\text{reach of T in the control group}}$$

The ratio is normalized so that:

- 0%: equal proportional reach in treatment and control groups
- 50%: the treatment group is 50% more likely to encounter a tweet

Large amplification ratios indicate the recommender assigns higher relevance to tweets in *T*, which appear more often than in a reverse chronological ordering

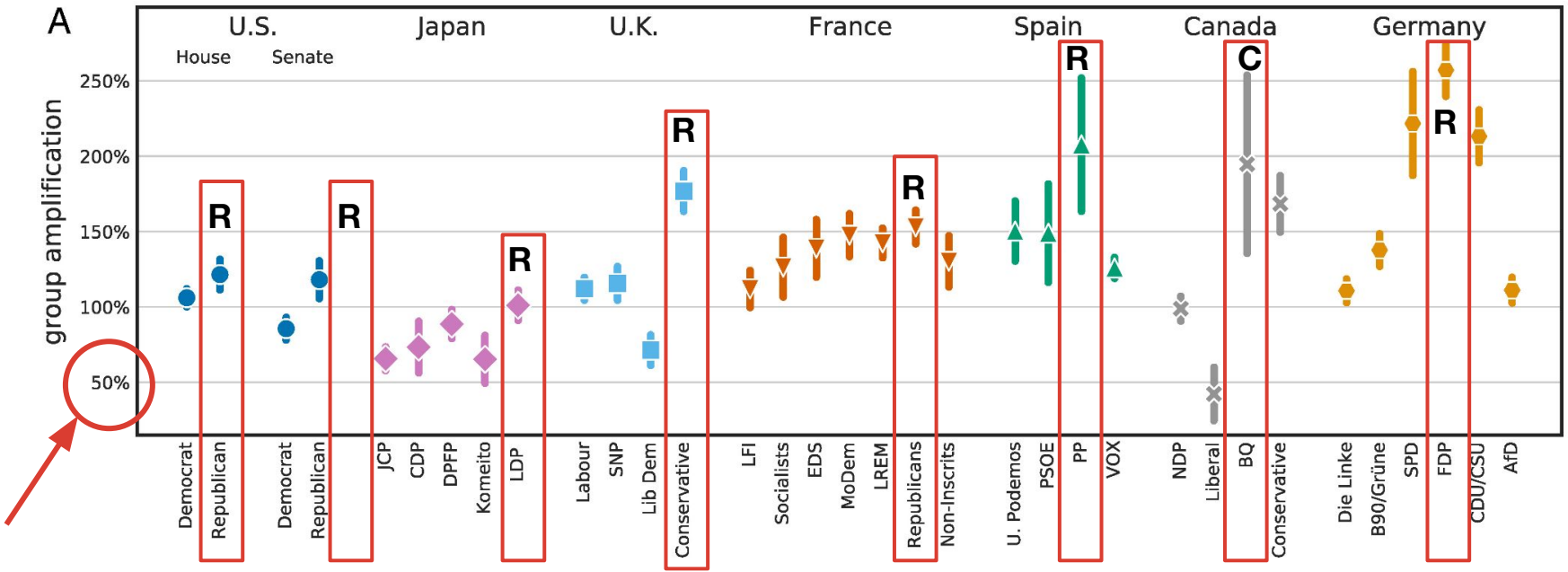
Experimental Setup

Huszár et al., PNAS 2021

- **3,643** Twitter accounts related to *currently serving* **legislators**
  - US, Canada, Japan, UK, France, Germany, Spain (>100k users in the control group)

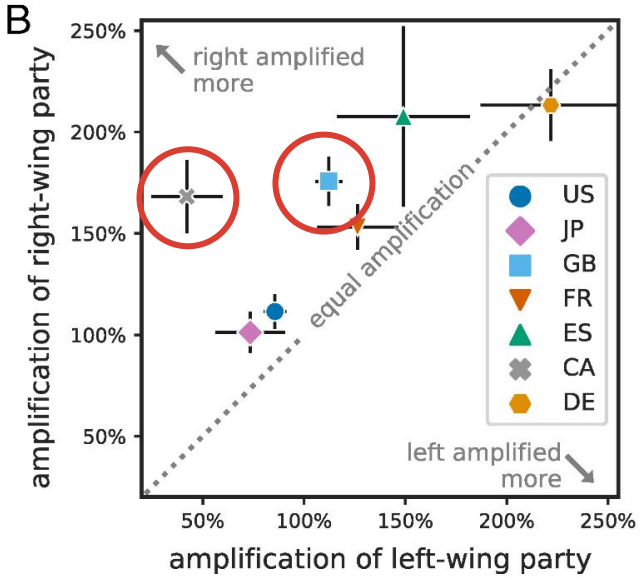
- all tweets, replies and quote tweets are considered
- the **reach** of tweets is computed in the respective country only

Group amplification: All tweets of legislators' accounts of a party



- **Amplification > 50%**
- in some cases > 200%
  - tweets exposed to an audience 3 times larger than that reached with the reverse chronological recommender

Group amplification: All tweets of legislators' accounts of a party



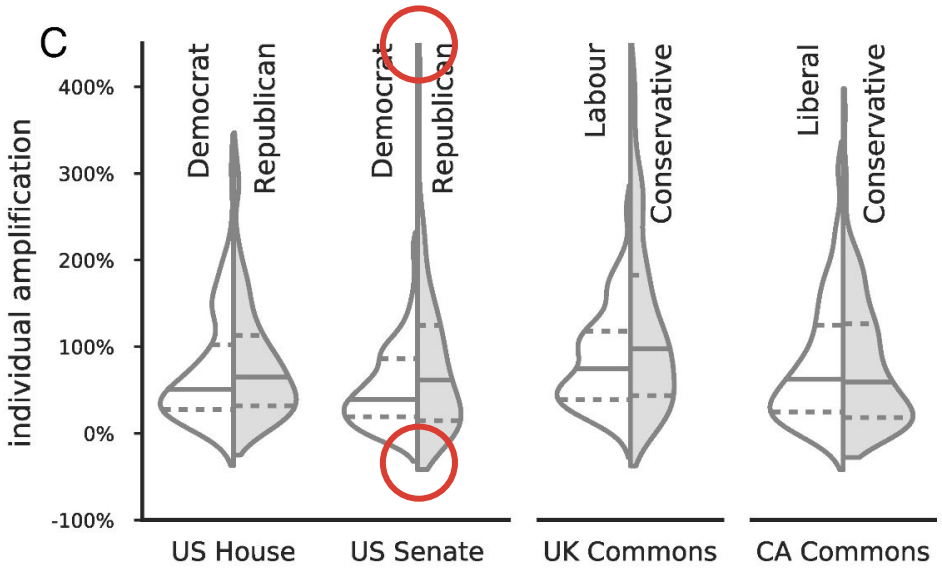
- The ***largest mainstream*** (center-)left and (center-)right parties are selected
- Statistical significant difference **favouring tweets from the political right wing** (except for Germany)

- **Canada**: Left 43% vs Right 167%
- **UK**: Left 112% vs Right 176%

Individual amplification: tweets of individual politicians

Amplification varies:

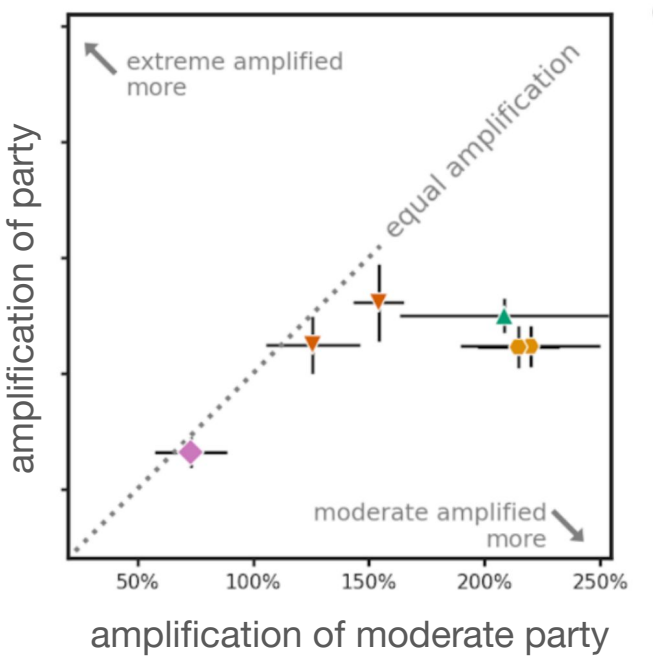
- Some politicians' amplification is up to **400%**
- for others, it is below **0%**



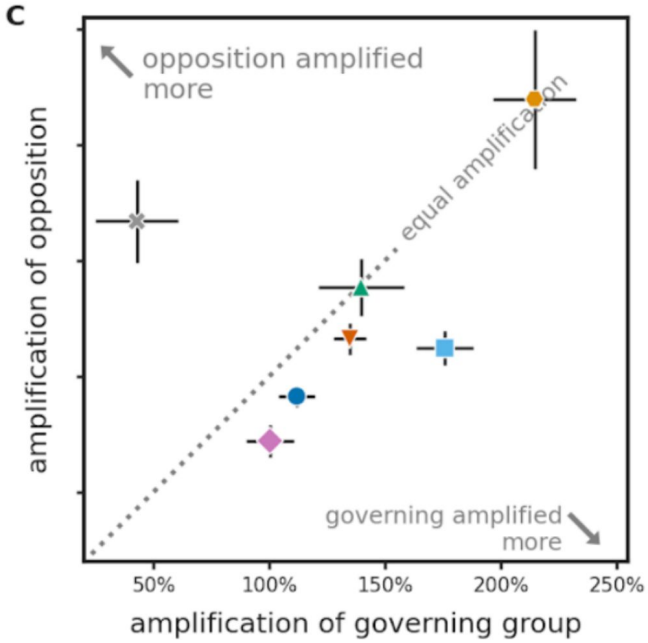
When comparing individual amplification between parties:

- **no significant association** between an individual's party affiliation and amplification

Extreme vs **moderate** parties



moderate parties are favoured
over far-left and far-right ones



Often (but not always)
governing parties are favoured

In summary

Huszár et al., PNAS 2021

Empirical controlled

- **at the party level**
  mainstream right-wing parties benefit more from the personalised Home Timeline than left-wing counterparts
- **at the individual level**
  no association between amplification and part membership
- **extreme vs moderate**
  the personalised Home Timeline does not favour extreme ideologies more than mainstream (moderate) ones

Discussion

Why right-wing tweets are amplified more?

Different political parties pursue different strategies on Twitter:

- J. H Parmelee and S. L. Bichard, **Politics and the Twitter Revolution: How Tweets Influence the Relationship between Political Leaders and the Public** (Lexington, 2011)
- D. Freelon, A. Marwick, D. Kreiss, **False equivalencies: Online activism from left to right**. Science 369 (2020)

Discussion

Huszár et al., PNAS 2021

What additional factors, beyond polarization, could be explored in this analysis?

- Misinformation
- Manipulation
- Hate speech
- Abusive content

Causally estimating the effect of YouTube's recommender system using counterfactual bots

Hosseinmardi et al., PNAS 2024, https://doi.org/10.1073/pnas.2313377121

Type:	**Empirical observational**
VLOP:	**YouTube**
Outcomes:	**filter bubble** **radicalization**

The New York Times

The Making of a YouTube Radical

By KEVIN ROOSE June 8, 2019

Opinion

YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

The New York Times

Does YouTube direct users to problematic content?

Causally estimating the effect of YouTube's recommender

Hosseinmardi et al., PNAS 2024

Empirical observational

Panel studies track clicks of users over time, but not recommendations

- What would a user have watched without recommendations?
- Is user's behavior influenced by the algorithm or their own preferences?

Audit studies record recommendations from the platform, but cannot estimate causal effects

- What a user might have chosen without algorithmic influence?
- Causal effects vary by user type (moderate vs. extreme)

Causally estimating the effect of YouTube's recommender

Hosseinmardi et al., PNAS 2024

- logged-in, programmatic users trained on a **real** user's historical trajectory
- empirical data of **desktop** behaviour by 87k users (Oct 2021 - Dec 2022)

An approach that employs "counterfactual bots" to estimate the effect of algorithmic recommendations independent of user intentions.

Experimental Setup

Hosseinmardi et al., PNAS 2024

- Experiments use **4,583 users** (those who watched >140 videos)
- From each user, **120-video-long trajectories** are extracted, starting at a random point within their watch history
  - 24,871 unique user histories in total
- An algo assigns **partisan scores** to videos based on channel labels
- Histories are clustered into **8 news consumption archetypes**, ranging from far-left to far-right
  - far-right clusters were further divided into three sub-clusters

Experiment 1

Hosseinmardi et al., PNAS 2024

125 focal users (with stratified sampling):

- centrist Ψ^{C} = 32 histories
- far-right-low: Ψ^{fR}_{low} = 35 histories
- far-right-medium: Ψ^{fR}_{med} = 41 histories
- far-right-high: Ψ^{fR}_{high} = 17 histories

- centrist (66%),
- far-right (1.12%)
  - oversampled for statistical robustness

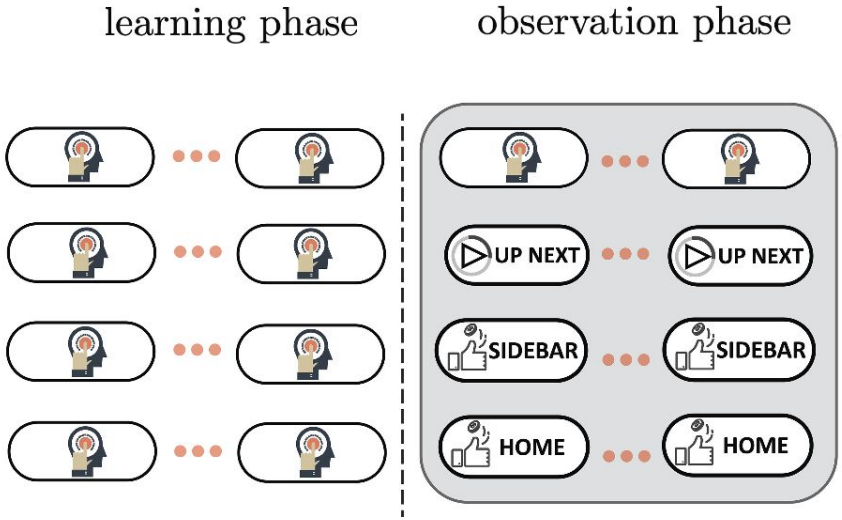
Experiment 1

Hosseinmardi et al., PNAS 2024

1) **Learning phase**: 4 bots follow the same sequence of **60** videos
   - indistinguishable "preferences"

2) **Observation phase**:

- *control*: 1 bot follows the focal user's trajectory (**60** videos)
- *"counterfactual" treatment*: 3 bots follow a predefined rule (up-next, random sidebar, random home)



Measures: causal effect for different types of users and users consuming bursts of far-right videos

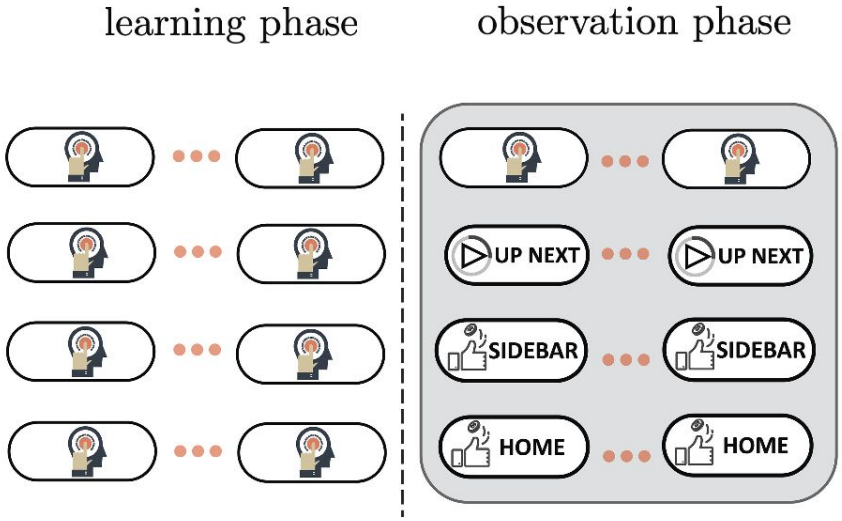
Experiment 1

Hosseinmardi et al., PNAS 2024

2) **Observation phase**:

three rules for bots:

1. **up-next** selects the first video from the sidebar (deterministic)
2. **random sidebar** randomly selects one of the top 30 videos in the sidebar
3. **random home** randomly selects a video from the top 15 videos on the homepage



Measures: causal effect for different types of users and users consuming bursts of far-right videos

Experiments

Hosseinmardi et al., PNAS 2024

Empirical observational

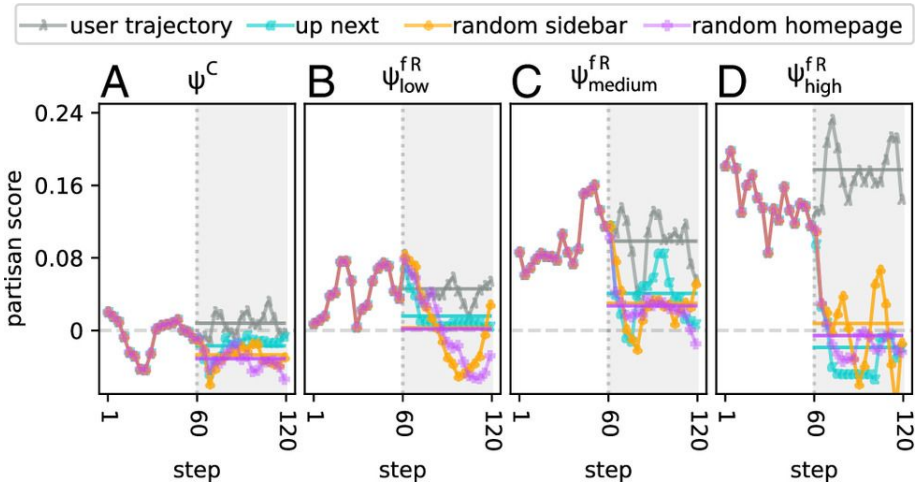
These experimental setups has three advantages:

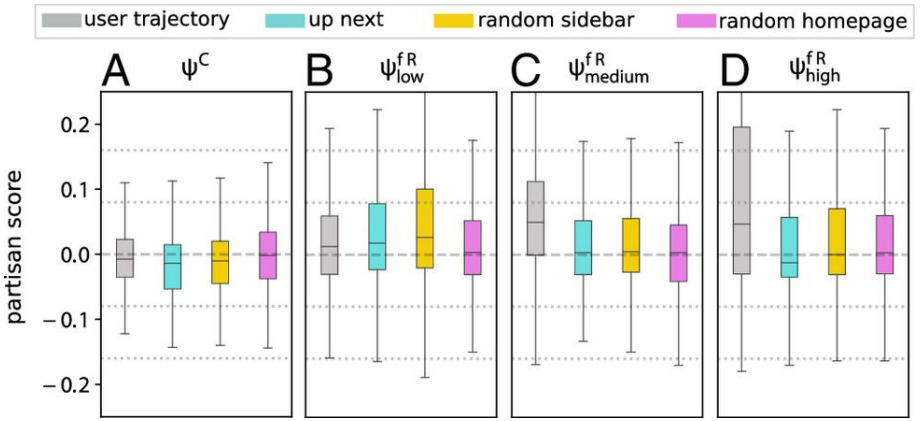
1. it **eliminates the preference** of observed consumption
2. since bots are trained on historical user data, the results describe **effects on real users**, not hypothetical ones
3. being the dataset of users large, they can follow on **those consuming the largest amount of problematic content**

Results 1: different types of users

Observation phase:

- Control bots (grey) stay on a similar trajectory
- Counterfactual bots (coloured):
  - diverge onto different paths
  - shift toward less partisan content
- Effect strongest for the far-right-high cluster Ψ^{fR}_{high}
- homepage > up-next > sidebar





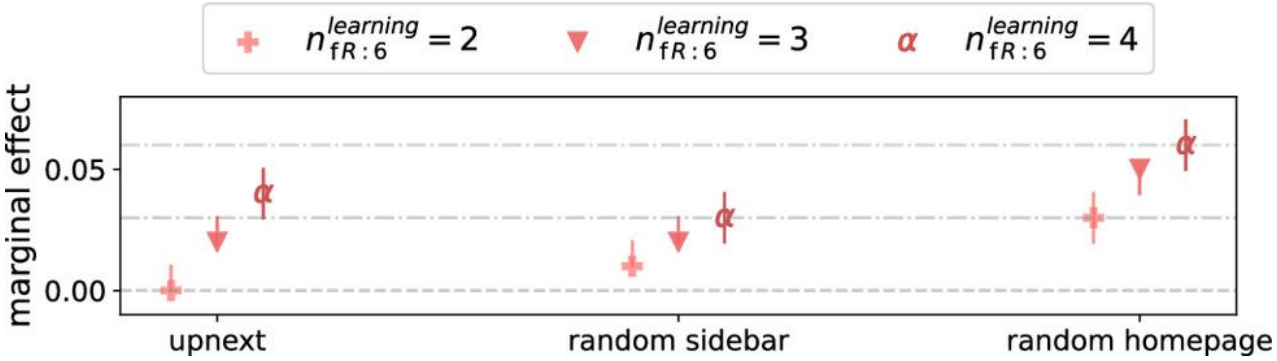
Results 1: bursts of extreme content

Users with bursts of C, R, or fR videos in the **last 6 videos** of the learning phase

$$\hat{y}_t^{\text{pref.}} = y_t^{\text{control}} - y_t^{\text{algo}}$$

→ difference in partisanship betwen control bots and counterfactual bots

$$\hat{y}_t^{\text{pref.}} = \alpha + \beta_1 n_{C:6}^{\text{learning}} + \beta_2 n_{R:6}^{\text{learning}} + \beta_3 n_{fR:6}^{\text{learning}}$$



recommendations following bursts offer more moderating effects

Experiment 2

Hosseinmardi et al., PNAS 2024

64 focal users (with stratified sampling):

- far-right-medium: $\Psi^{fR}_{\text{med}} = 27$ histories
- far-right-high: $\Psi^{fR}_{\text{high}} = 17$ histories

Each counterfactual bot is supplied by a randomly selected history from Ψ^{C}

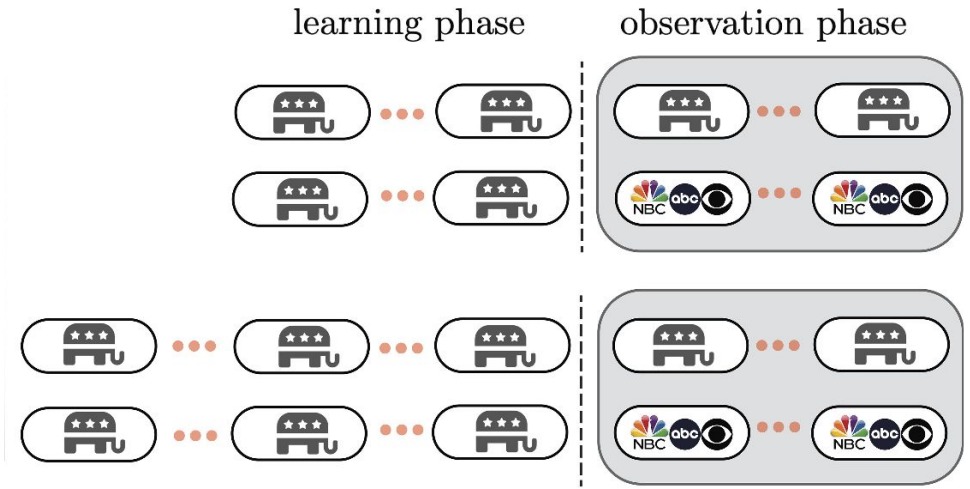
- experiment for each user **replicated 3 times**

Experiment 2

Hosseinmardi et al., PNAS 2024

Learning phase: bots trained on a **far-right user**

- half short (30 videos)
- half long (120 videos)



Observation phase:

- 1 bot watches the same trajectory
- 1 bot switches to moderate videos
- sidebar & homepage recs are tracked

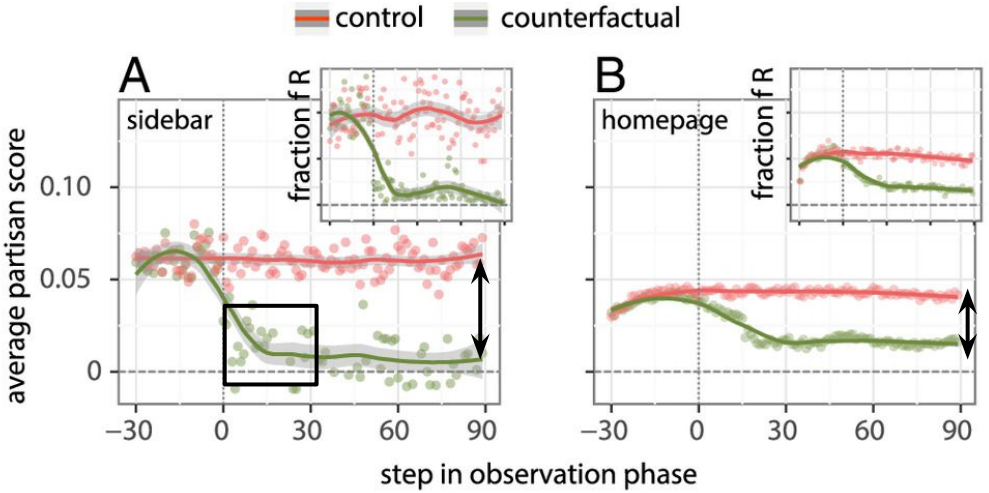
Measures: forgetting times of users with short (30) and long (120) histories

Results 2: forgetting time

Average partisanship of sidebar and homepage recommendations

Sidebar: large and rapid decrease in partisanship

- within **30 videos**, recommendations become similar to those of moderate users



Homepage: less marked decrease in partisanship than sidebar recommendations

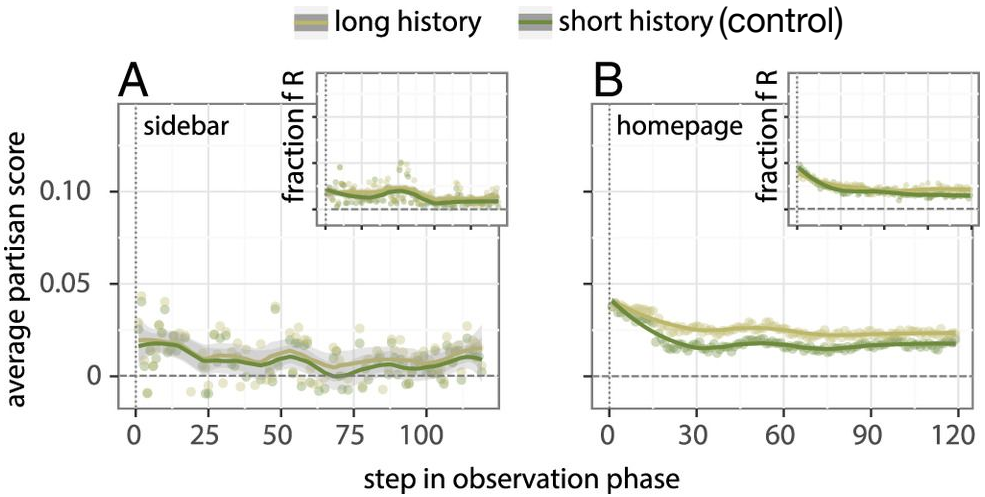
- on average, fR videos disappear between 30 and 40 videos
- a small fraction of fR videos continue to appear

Results 2: forgetting time and history length

- **Control bot**:
  watches 150 videos (30+120)
- **Counterfactual bot**:
  watches 240 videos
  (120+120)

Sidebar: both short and long paths exhibit a **similar drop rate** converging towards 0 fR videos

Homepage: long history paths exhibits a gradual decrease that persists until the end



- fR videos drops along the trajectory, where from step 70 they diverge slightly

In summary…

Hosseinmardi et al., PNAS 2024

Empirical observational

1. **Bots receive and consume less** partisan content than real users (especially heavy partisan consumers)
2. Users consuming bursts of highly partisan content engage with **more partisan content** than bots
3. Switching from far-right to moderate news removes far-right recommendations from the sidebar **within 30 videos** (but lingers longer on the homepage)
4. Longer histories of far-right consumption **extend homepage recommendation persistence** but do not affect sidebar "forgetting" time

Recommendations moderate user experiences (especially extreme users)

Discussion

Hosseinmardi et al., PNAS 2024

What are possible limitations of this study?

References

Articles (useful for the project):

- Huszár et al., **Algorithmic amplification of politics on Twitter**, PNAS 2021, https://doi.org/10.1073/pnas.2025334119
- Hosseinmardi et al., **Causally estimating the effect of YouTube's recommender system using counterfactual bots**, PNAS 2023, https://doi.org/10.1073/pnas.2313377121
- L. Pappalardo et al. **A survey on the impact of AI-based recommenders on human behaviours: methodologies, outcomes and future directions**, 2024, https://doi.org/10.48550/arXiv.2407.01630
    - Section 3 Social Media Ecosystem

References

Articles (useful for the project):

- Gauthier, G., Hodler, R., Widmer, P. et al. **The political effects of X's feed algorithm**. Nature 652, 416–423 (2026). https://doi.org/10.1038/s41586-026-10098-2
- Andrew M. Guess et al., **How do social media feed algorithms affect attitudes and behavior in an election campaign?** Science381,398-404(2023).DOI:10.1126/science.abp9364

Books, articles, podcasts

- Never miss important tweets from people you follow, https://blog.x.com/en_us/a/2016/never-miss-important-tweets-from-people-you-follow
- The Making of a YouTube radical, New York Times, https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html
- **The Facebook Files**, The Wall Street Journal, https://www.wsj.com/articles/the-facebook-files-11631713039
- Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations, PNAS 2023, https://doi.org/10.1073/pnas.2213020120
- Your Filter Bubble is Destroying Democracy, https://www.wired.com/2016/11/filter-bubble-destroying-democracy/
- Sean Parker unloads on Facebook: "God only knows what it's doing to our children's brains", https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792

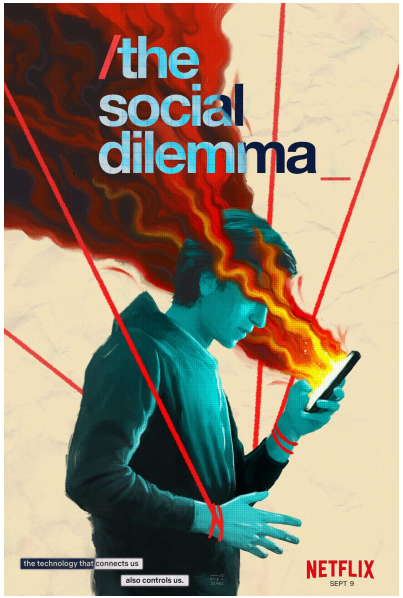
Books, articles, podcasts

Intellectually stimulating:

- Eli Pariser, "The Filter Bubble: What the Internet Is Hiding from You", Penguin, 2025
- Albert-László Barabási, “Linked: The New Science of Networks”, Perseus Books Group, 2002

Movies

The Social Dilemma
2020



The Social Network
2010

15 million merits
2011

Ready Player One
2018