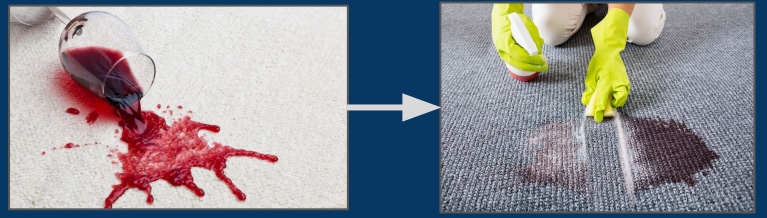
SCHOOL OF INFORMATION CENTER FOR
SOCIAL MEDIA RESPONSIBILITY
UNIVERSITY OF MICHIGAN



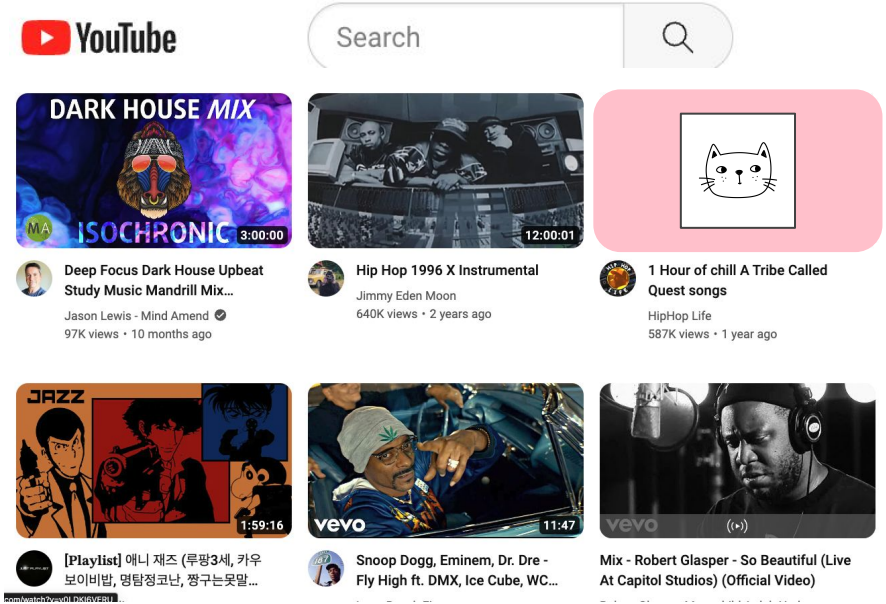
How to Train Your YouTube Recommender to Avoid Unwanted Videos

Alex Liu, Siqi Wu, Paul Resnick

ICWSM 2024
Buffalo, NY, USA

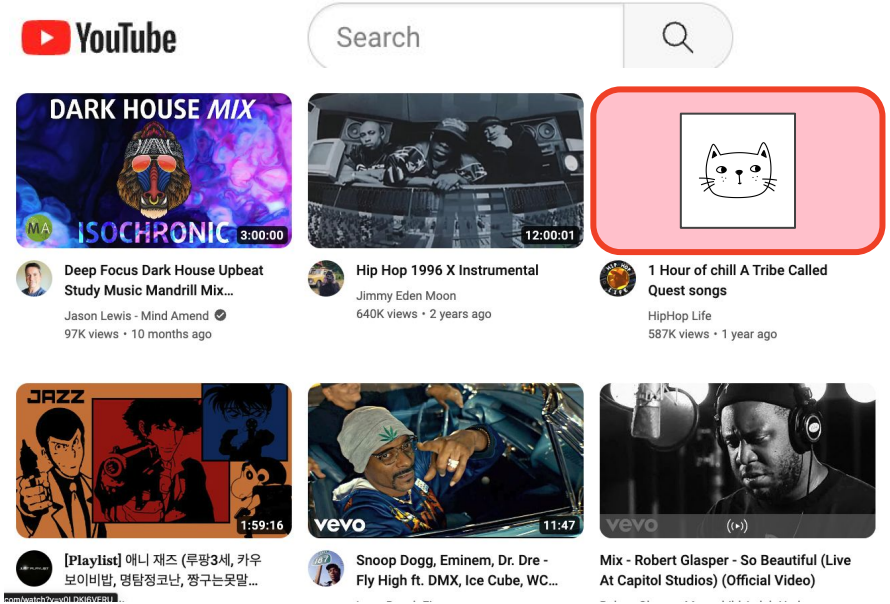
Motivation

(1) watch a few cat videos



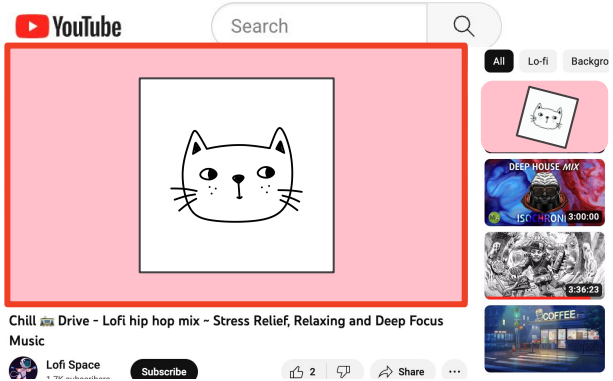
Motivation

(1) watch a few cat videos



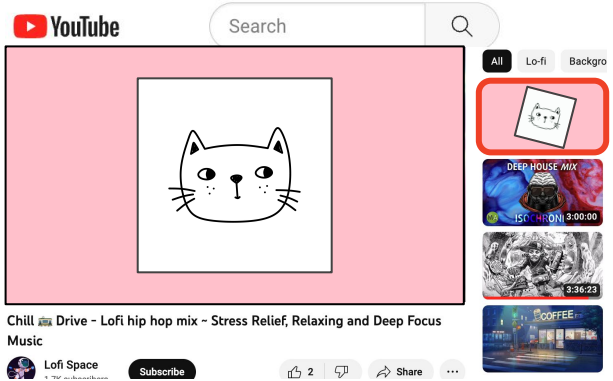
Motivation

(1) watch a few cat videos



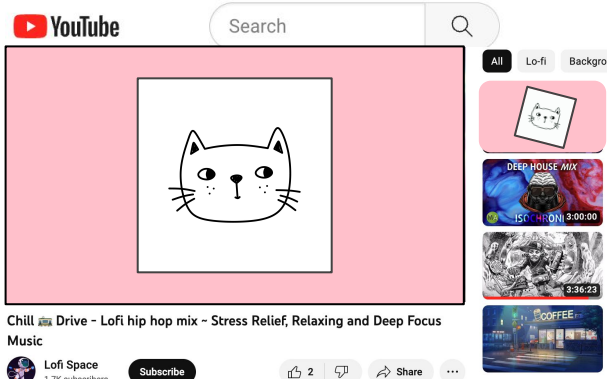
Motivation

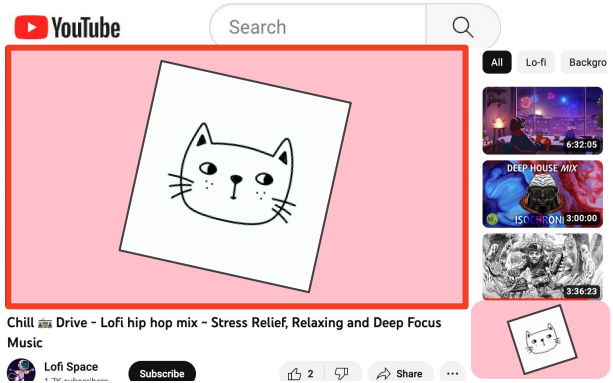
(1) watch a few cat videos



Motivation

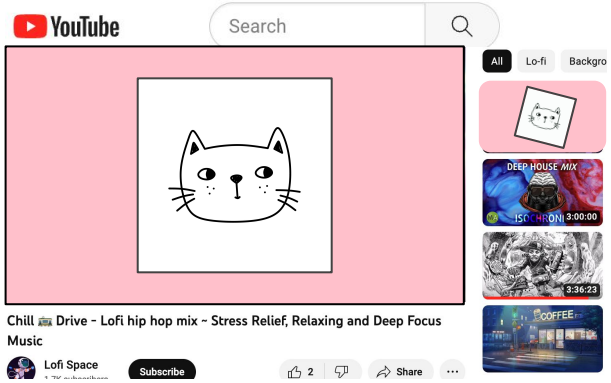
(1) watch a few cat videos

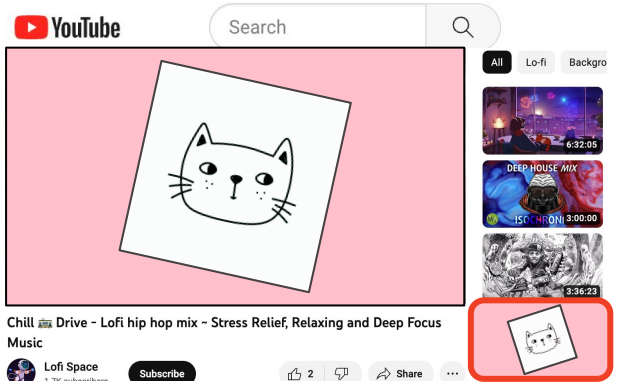




Motivation

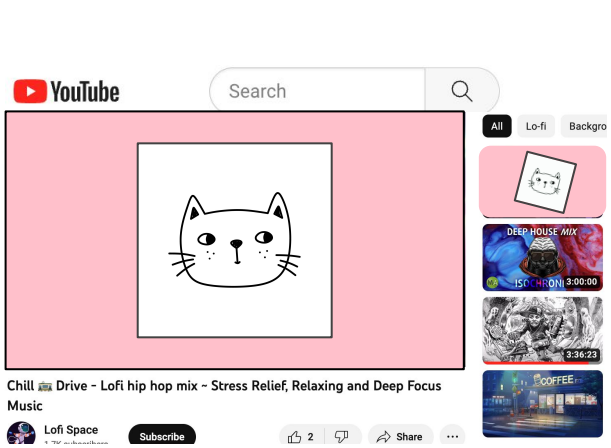
(1) watch a few cat videos

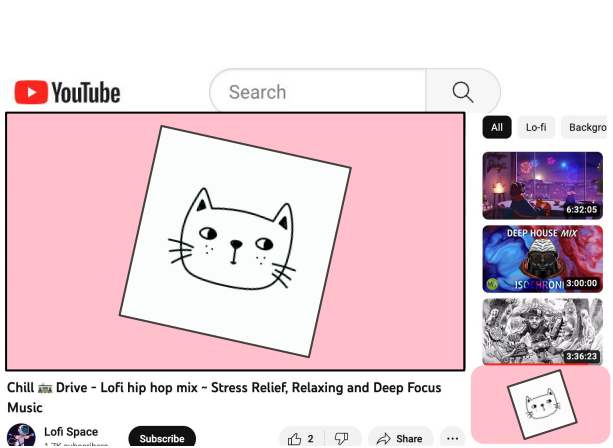


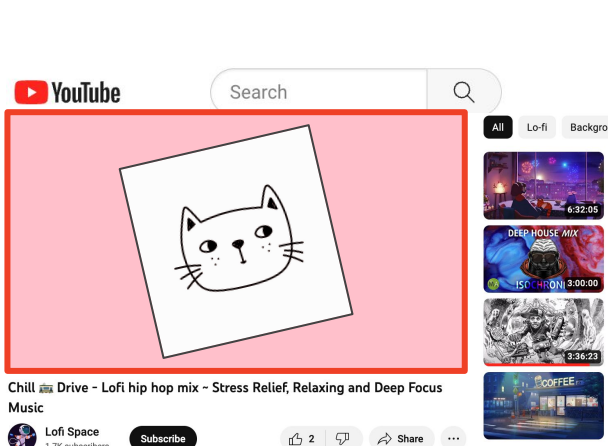


Motivation

(1) watch a few cat videos

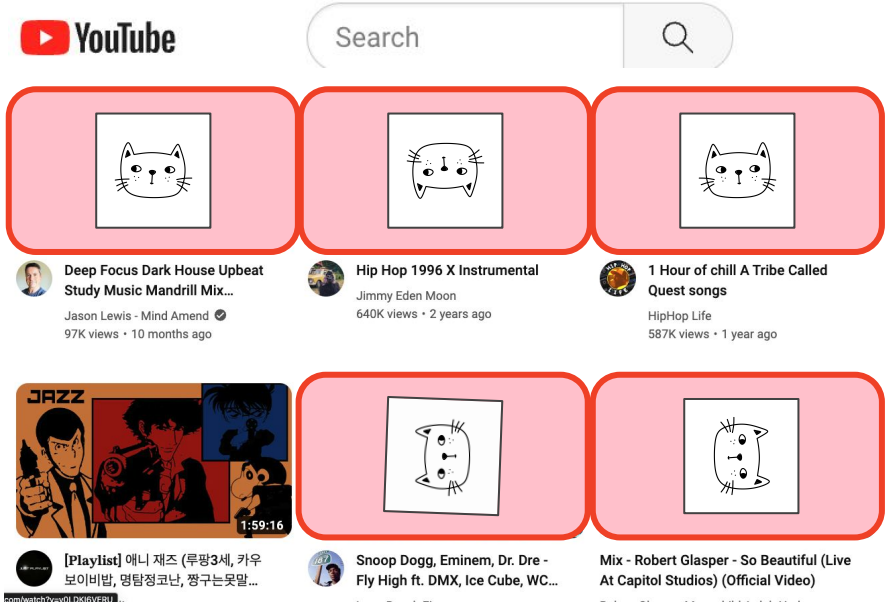






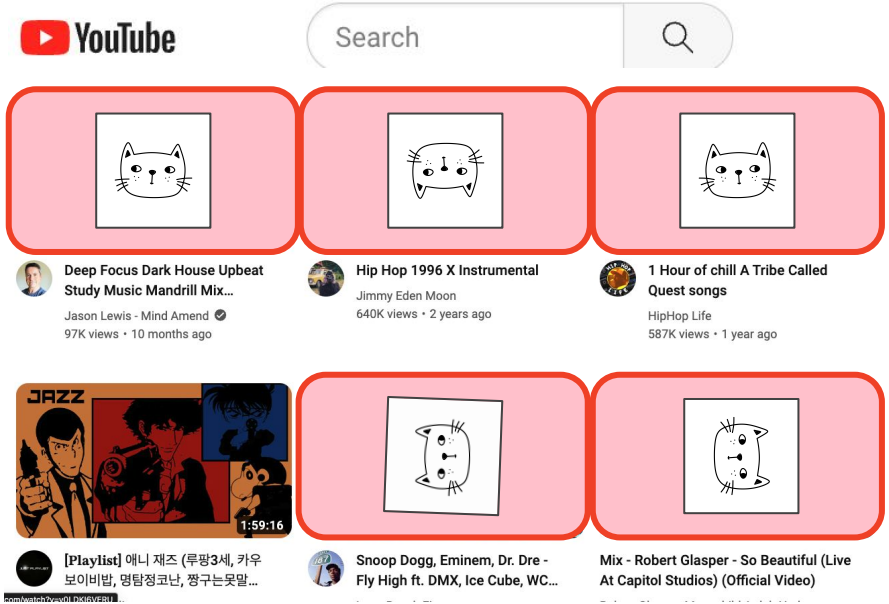
Motivation

(1) watch a few cat videos **(2) too many cat recommendations!**



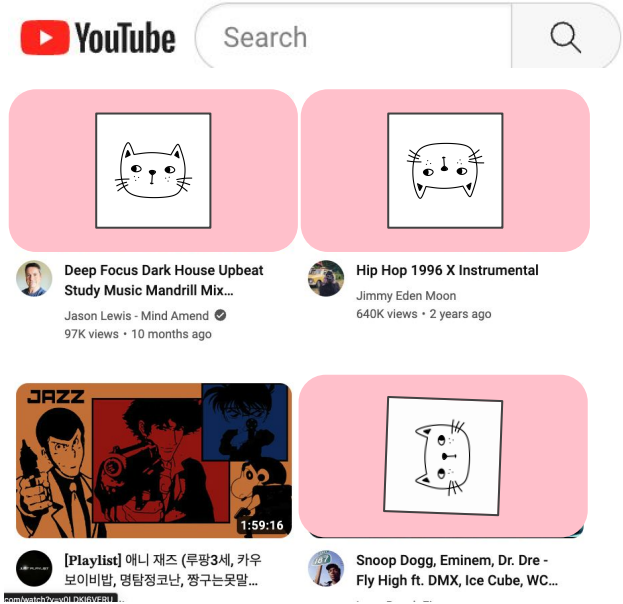
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



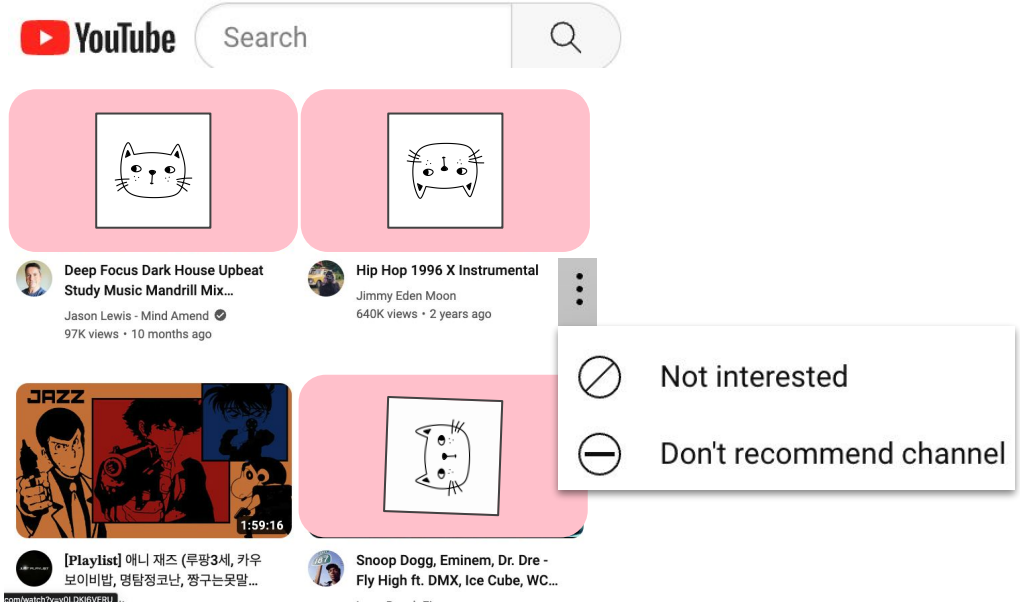
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



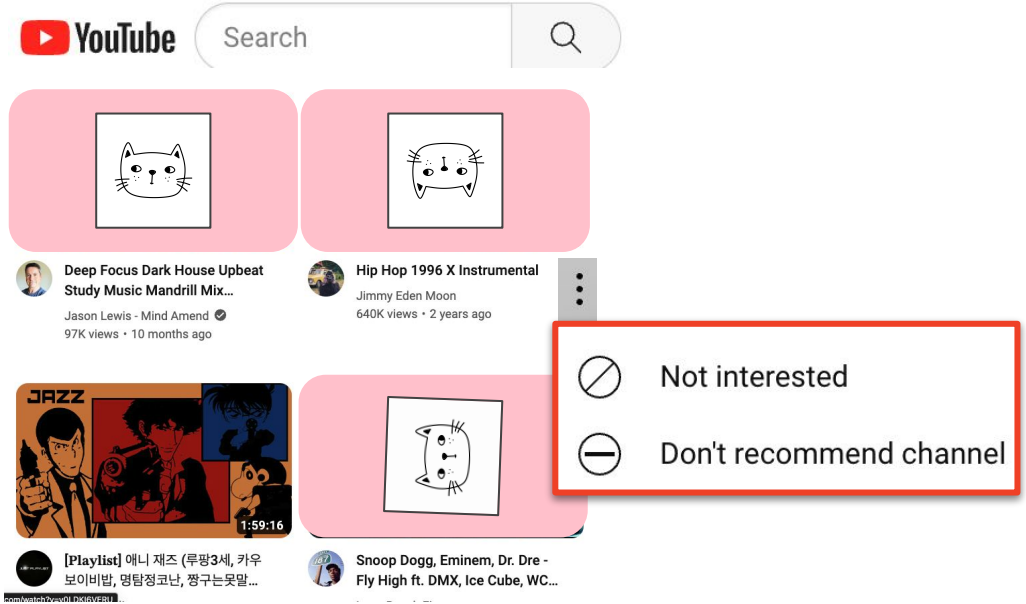
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



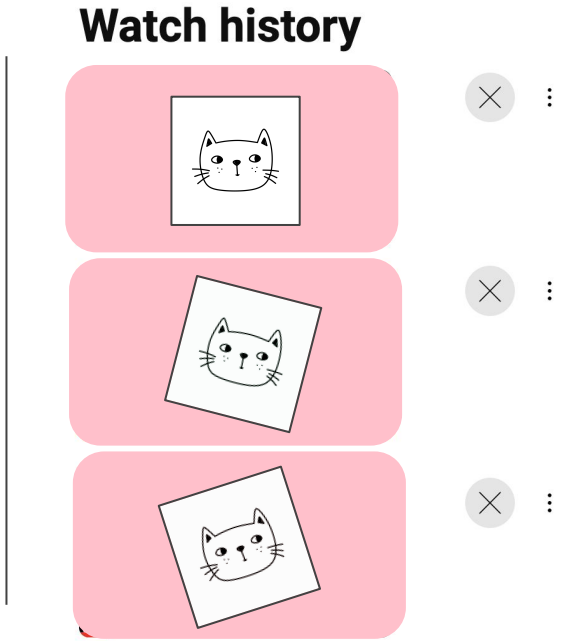
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



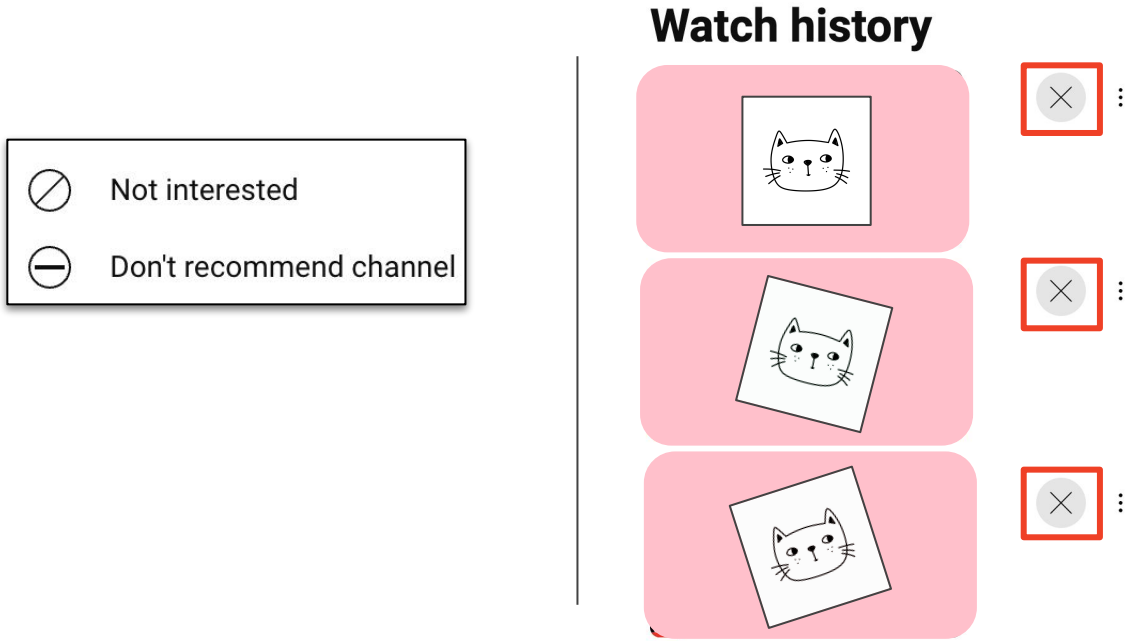
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



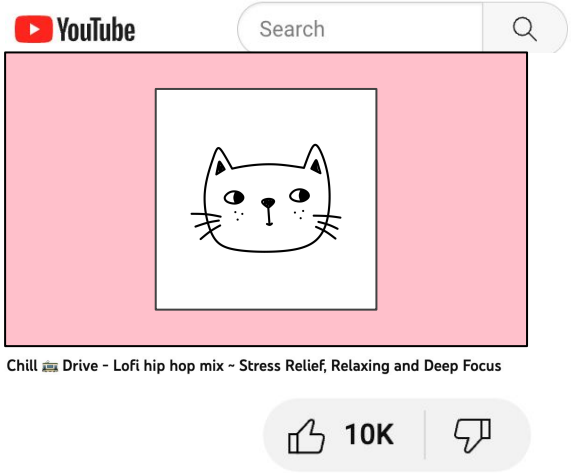
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



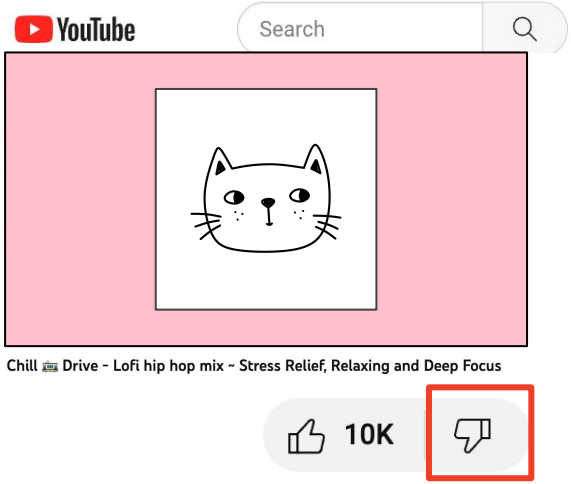
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



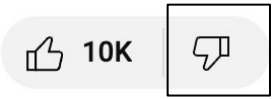
Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



Motivation

(1) watch a few cat videos (2) too many cat recommendations! **(3) how to remove?**



Motivation

- YouTube makes unwanted recommendations
  - Societally-harmful content (Hosseinmardi et al. 2021, 2024; Lewis 2018; Ribeiro et al. 2020).
  - YouTube Regrets (Mozilla Foundation, 2021).

Motivation

- YouTube makes unwanted recommendations
- Many platform features could help remove them
  - → **But, do they work?**

Research Questions

1. How big are YouTube **stains***?
    a. What are the differences in topics?

*stain = % recommendations of a topic, after watching many of their videos

Research Questions

1. How big are YouTube **stains***?
2. Can platform features **remove stain***?
    a. Which ones work best?

*stain = % recommendations of a topic, after watching many of their videos

Research Questions

1. How big are YouTube **stains***?
2. Can platform features **remove stain***?
3. How do users interact with these features?
    a. How many are **aware** of them?
    b. How many **use** them?
    c. How **effective** do they find them?

*stain = % recommendations of a topic, after watching many of their videos

Outline

1. Introduction
2. **Sock puppet audit**
   a. **Methods**
   b. Results
3. User survey
4. Conclusion

Sock puppet- Overview

(1) Stain phase

Sock puppet- Overview

(1) Stain phase

(2) Scrub phase

Sock puppet- Overview

(1) Stain phase

(2) Scrub phase

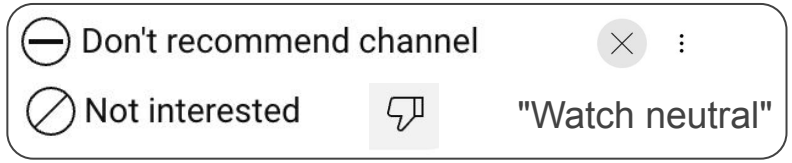
One of:

Alt-Right Antitheism

Political Left Political Right

Sock puppet- Overview

One of:



(1) Stain phase

(2) Scrub phase

One of:

Alt-Right **Antitheism**

Political Left **Political Right**

Sock puppet- Overview

Data collection

P1

(1) Stain phase

One of:

Alt-Right Antitheism

Political Left Political Right

P2

(2) Scrub phase

One of:

Don't recommend channel

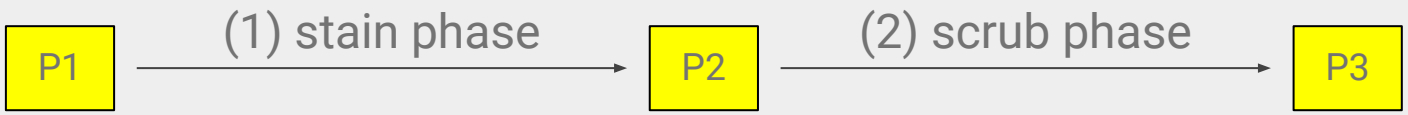
Not interested

"Watch neutral"

P3

Sock puppet- Example

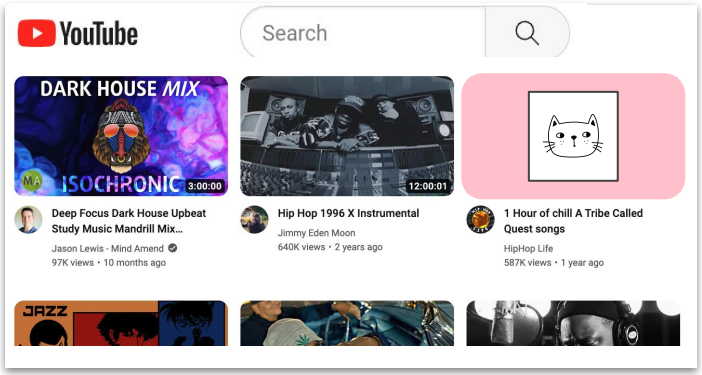
Topic: Cats
Scrub: ⊘ Not interested



Sock puppet- Example

Topic: Cats
Scrub: Not interested

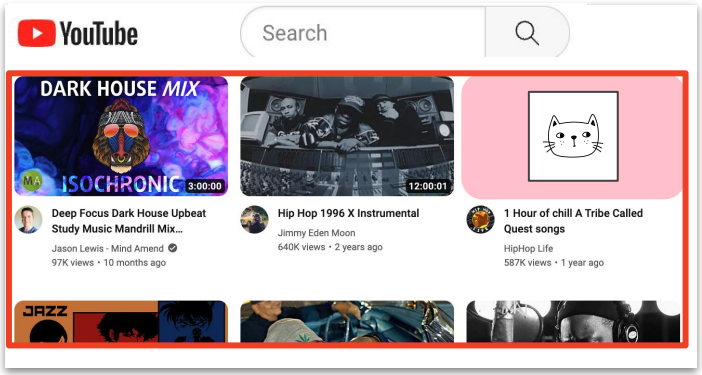
P1 → (1) stain phase → P2 → (2) scrub phase → P3



Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

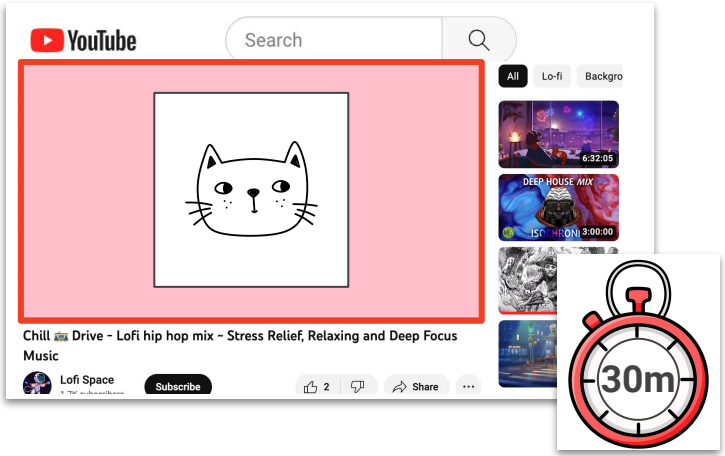


Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → **(1) stain phase** → P2 → (2) scrub phase → P3

collect homepage recommendations

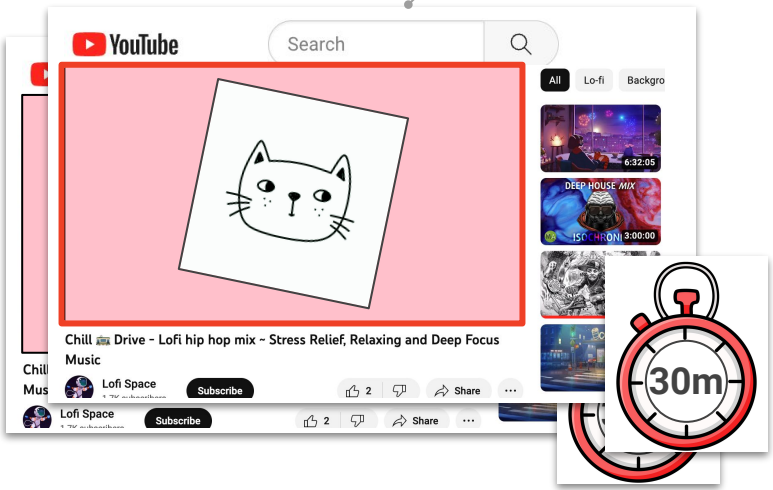


Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → **(1) stain phase** → P2 → (2) scrub phase → P3

collect homepage recommendations

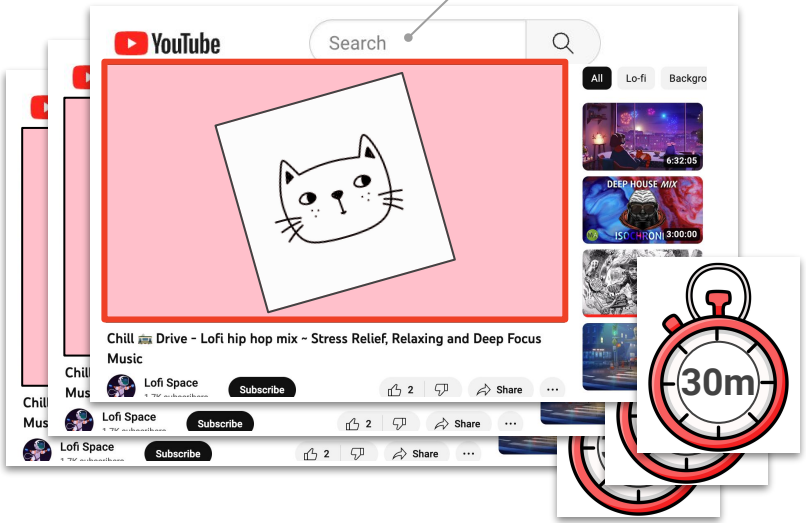


Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → **(1) stain phase** → P2 → (2) scrub phase → P3

collect homepage recommendations

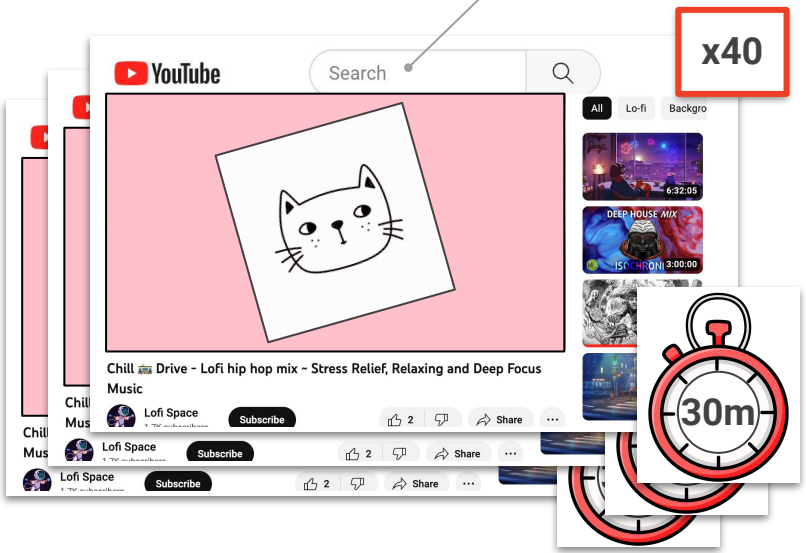


Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → **(1) stain phase** → P2 → (2) scrub phase → P3

collect homepage recommendations



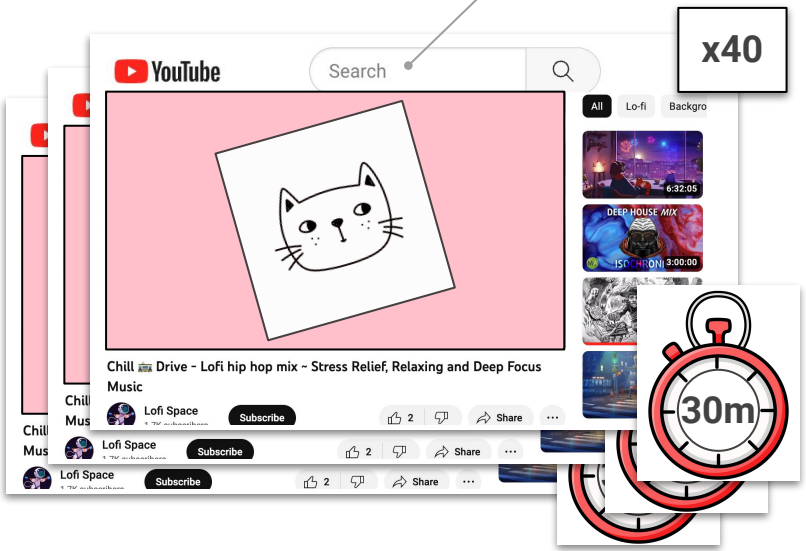
Sock puppet- Example

Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations



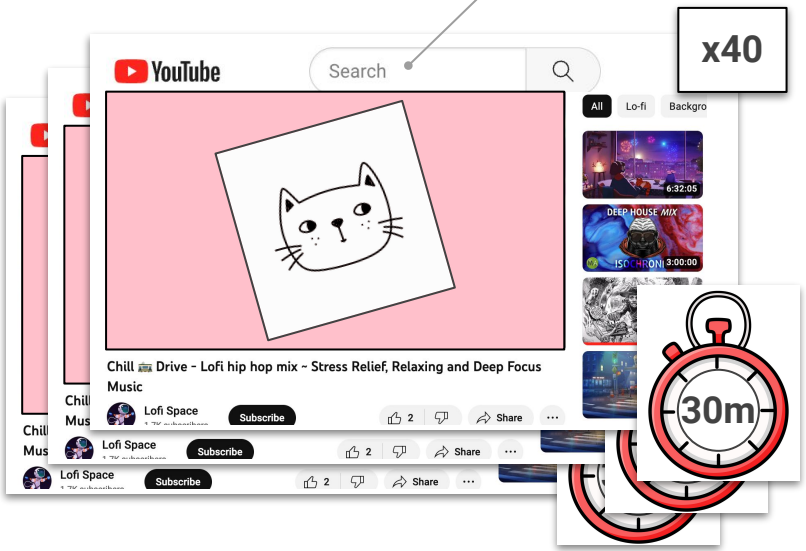
Sock puppet- Example

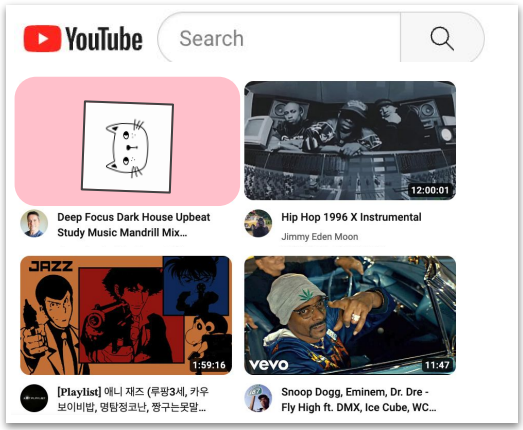
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





Sock puppet- Example

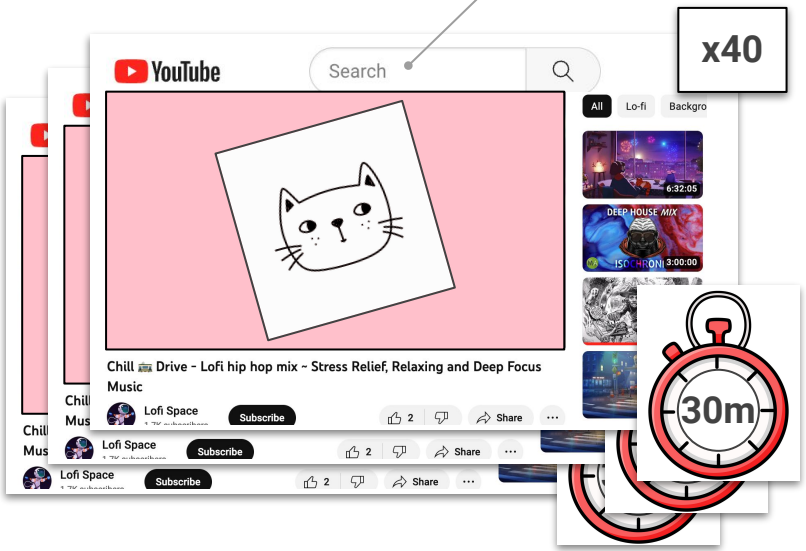
Topic: Cats

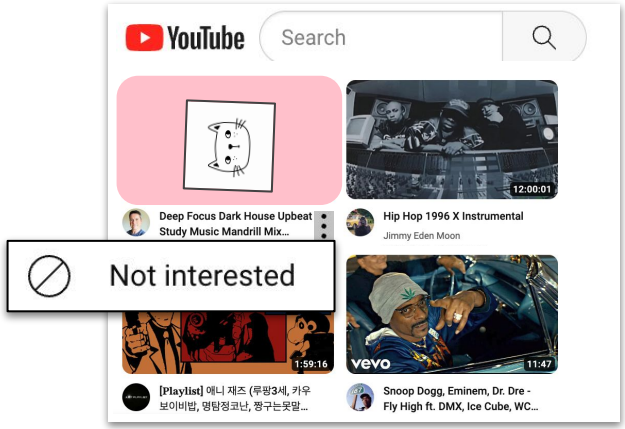
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





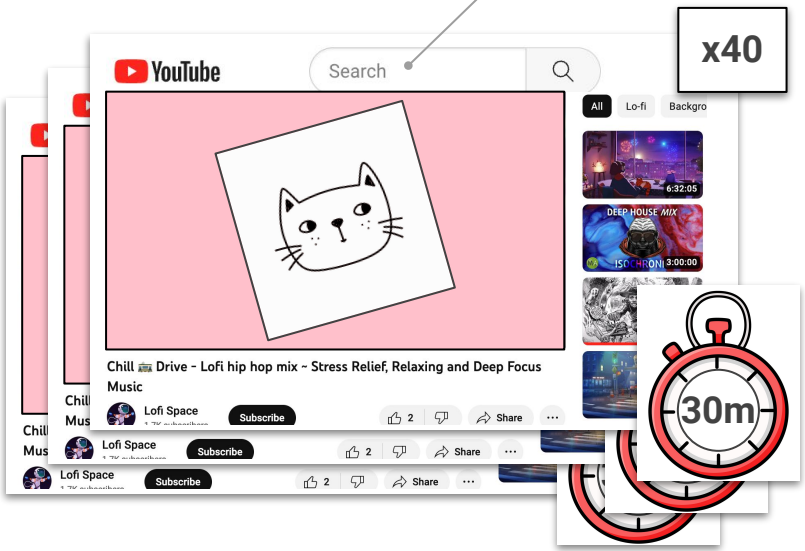
Sock puppet- Example

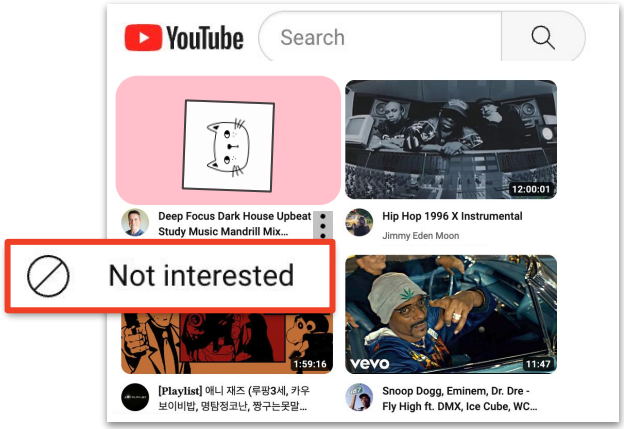
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





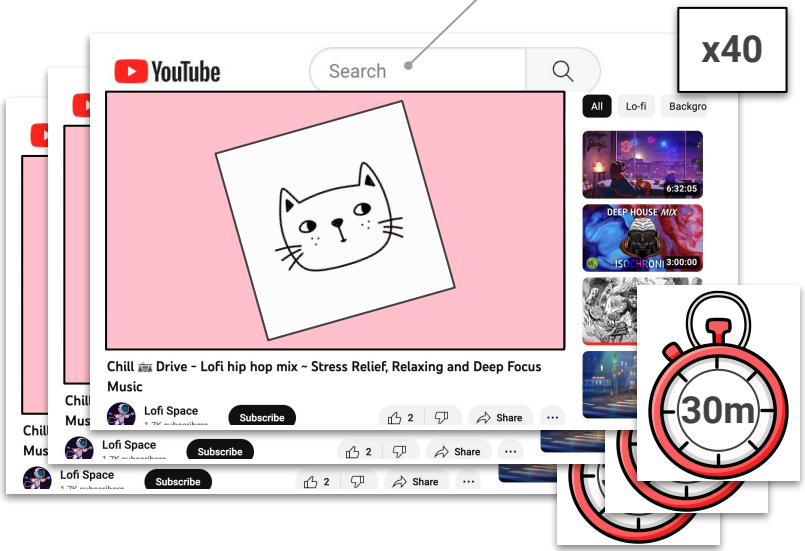
Sock puppet- Example

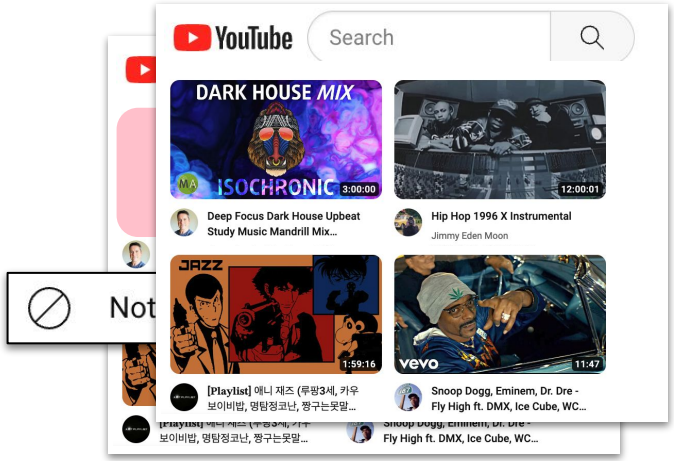
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





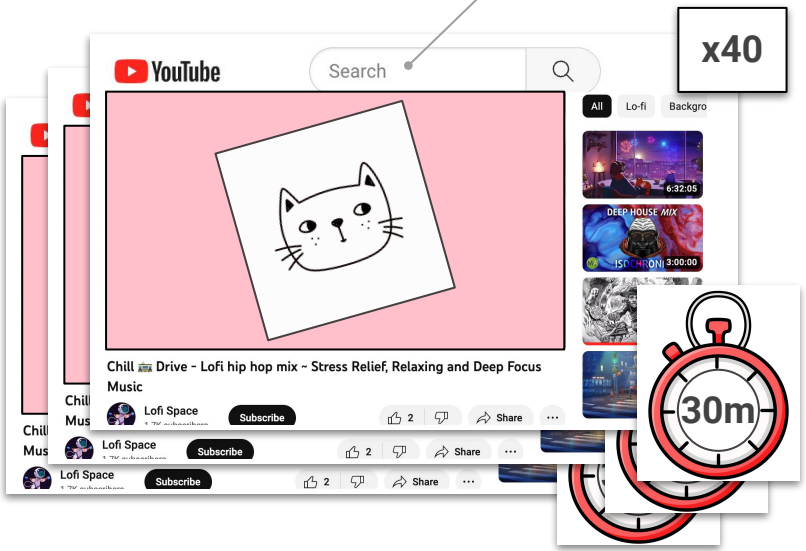
Sock puppet- Example

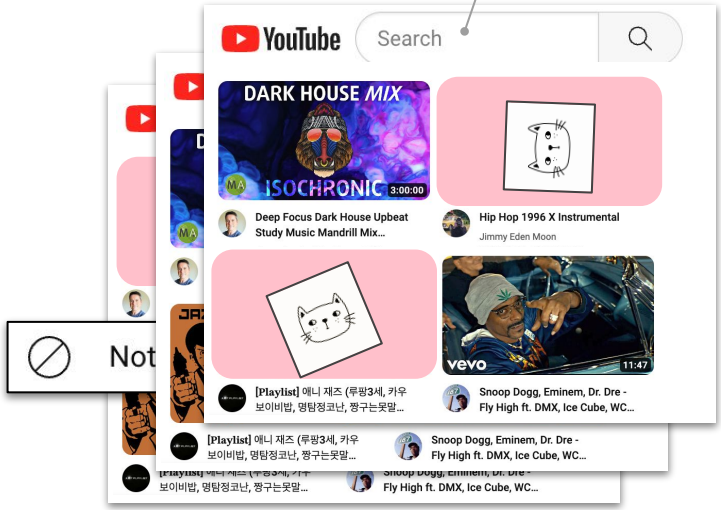
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





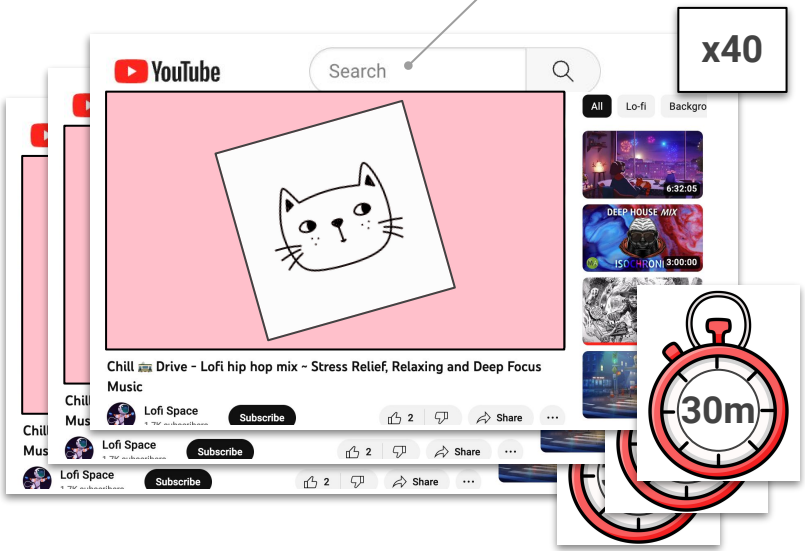
Sock puppet- Example

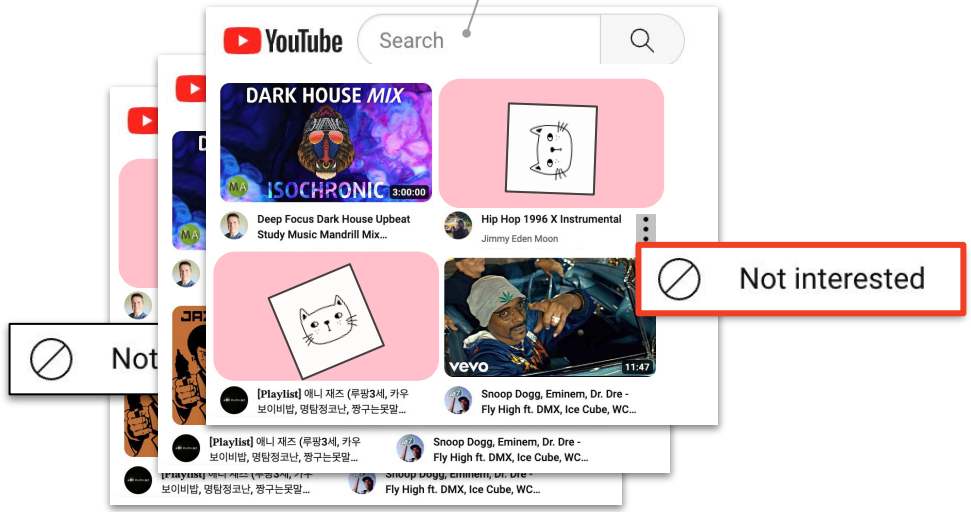
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





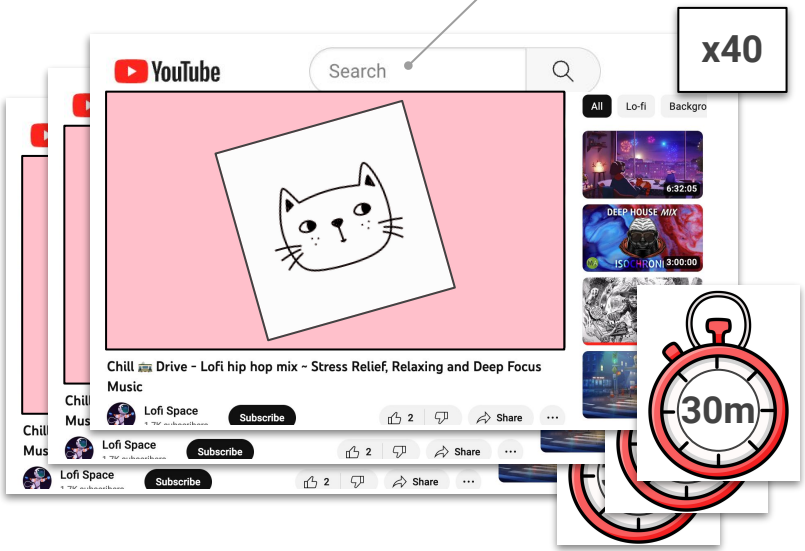
Sock puppet- Example

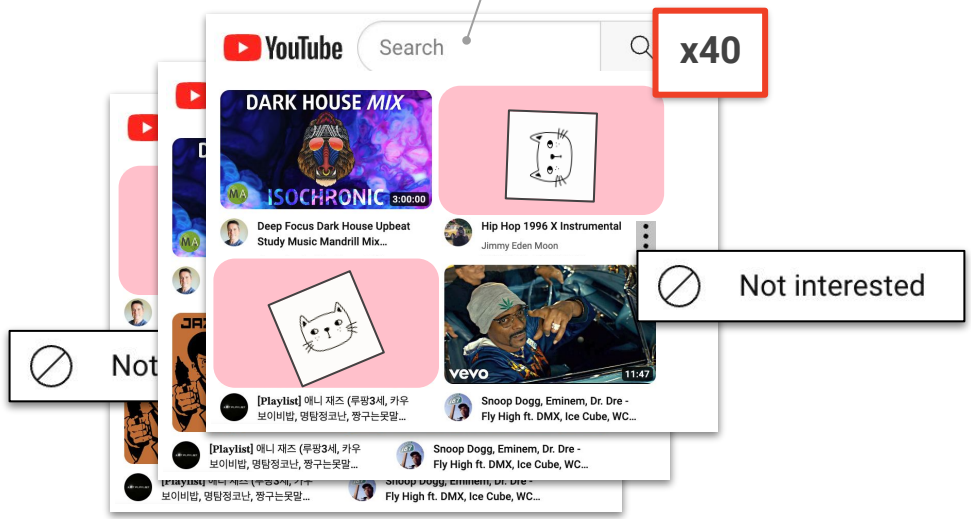
Topic: Cats
Scrub: Not interested

P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations





Sock puppet- Example

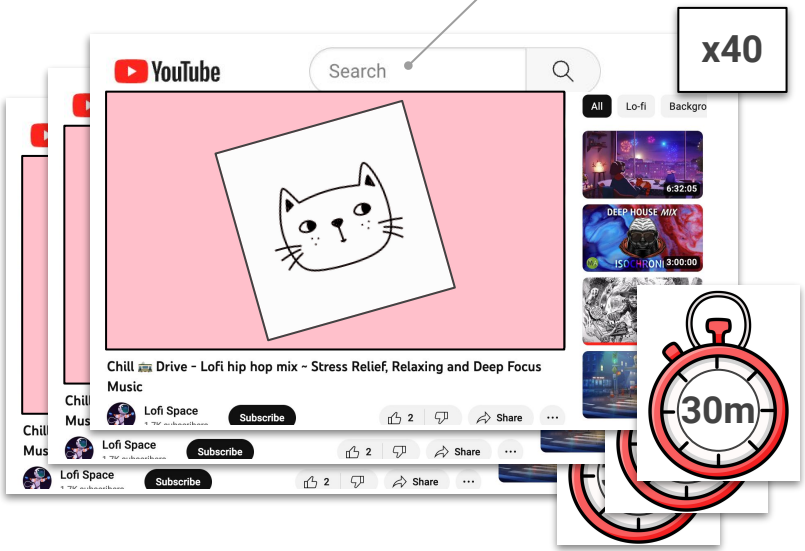
Topic: Cats
Scrub: Not interested

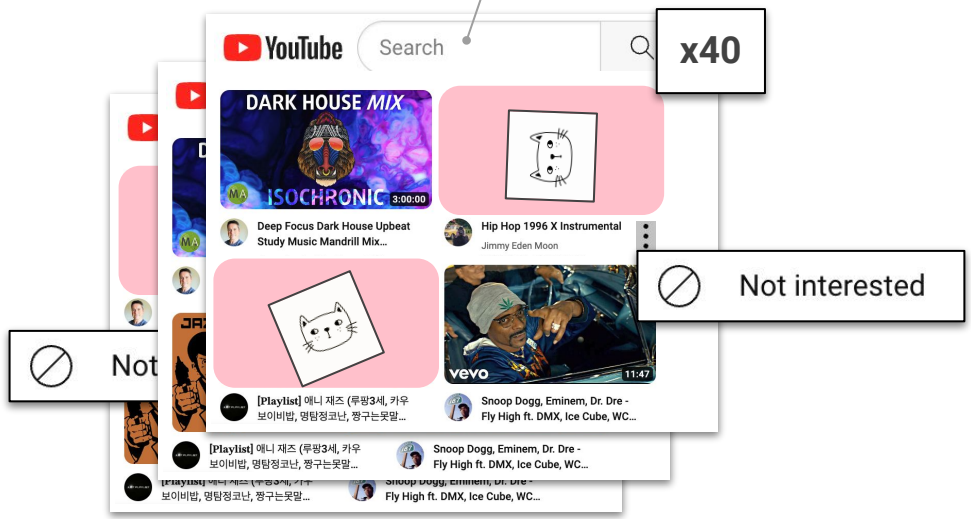
P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations

collect homepage recommendations





Sock puppet- Example

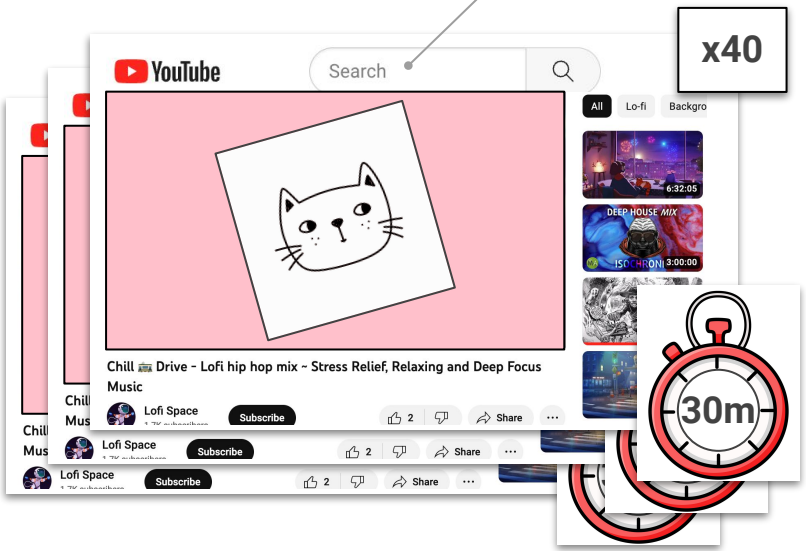
Topic: Cats
Scrub: Not interested

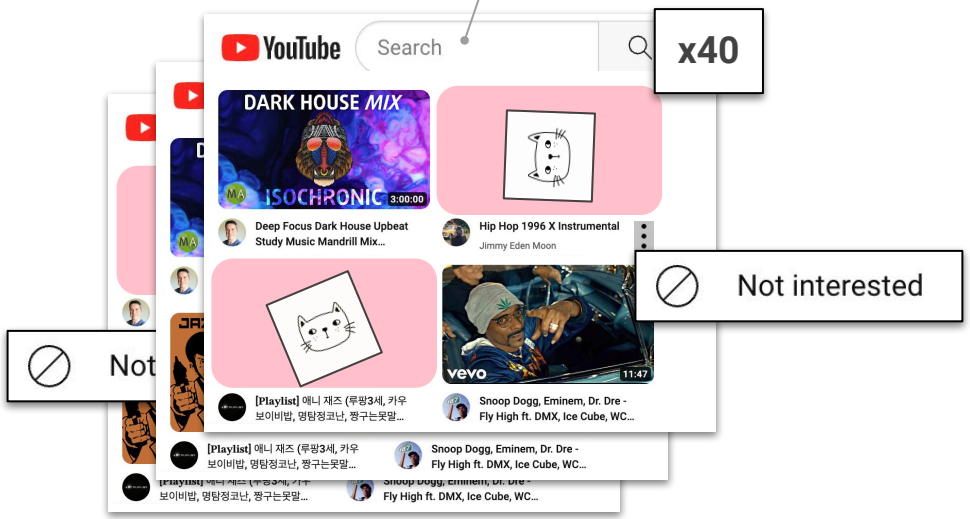
P1 → (1) stain phase → P2 → (2) scrub phase → P3

collect homepage recommendations

collect homepage recommendations

collect homepage recommendations





Sock puppet audit- implementation details

Schedule of sock puppets

- For each topic, test 7 strategies, 5 times → 35 bots in parallel

Manual labeling of recommendations

- On-topic / Off-topic
- 0.66-0.95 Cohen's Kappa for different topics

Outline

1. Introduction
2. **Sock puppet audit**
    a. Methods
    b. Results
3. User survey
4. Conclusion

Stain phase effects

- Stain happens on the homepage
  - but never reaches over half

Stain phase effects

- Stain happens on the homepage
  - but never reaches over half

	Alt-Right	
P1		P2
2%	$\overset{*}{\rightarrow}$	20%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

Stain phase effects

- Stain happens on the homepage
  - but never reaches over half

Alt-Right			*Antitheist*			*Political Left*			*Political Right*		
P1		P2	P1		P2	P1		P2	P1		P2
2%	$\overset{*}{\rightarrow}$	20%	11%	$\overset{*}{\rightarrow}$	37%	6%	$\overset{*}{\rightarrow}$	28%	9%	$\overset{*}{\rightarrow}$	29%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

Scrub phase effects

- Not interested works best

* → = significant difference (MWU test)

Paper: Table 1

Scrub phase effects

- ⊘ Not interested works best

	Alt-right	
	P2	P3
Watch neutral	12%	8%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

Paper: Table 1

Scrub phase effects

- Not interested works best

	Alt-right		*Antitheist*		*Political Left*		*Political Right*		Avg. relative change
	P2	P3	P2	P3	P2	P3	P2	P3	P2 to P3
Watch neutral	12%	8%	34%	22%	32%	$\overset{*}{\rightarrow}$ 18%	26%	$\overset{*}{\rightarrow}$ 16%	-38%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

Paper: Table 1

Scrub phase effects

- ⊘ Not interested works best

	Alt-right			*Antitheist*			*Political Left*			*Political Right*			Avg. relative change
	P2		P3	P2		P3	P2		P3	P2		P3	P2 to P3
Watch neutral	12%		8%	34%		22%	32%	$\overset{*}{\rightarrow}$	18%	26%	$\overset{*}{\rightarrow}$	16%	-38%
Delete	20%	$\overset{*}{\rightarrow}$	0%	32%		24%	30%	$\overset{*}{\rightarrow}$	0%	24%	$\overset{*}{\rightarrow}$	4%	-77%
Dislike	20%		4%	30%	$\overset{*}{\rightarrow}$	10%	24%		28%	32%	$\overset{*}{\rightarrow}$	16%	-45%
Not interested	18%	$\overset{*}{\rightarrow}$	0%	42%	$\overset{*}{\rightarrow}$	2%	28%	$\overset{*}{\rightarrow}$	10%	26%	$\overset{*}{\rightarrow}$	2%	-88%
No channel	28%	$\overset{*}{\rightarrow}$	10%	44%	$\overset{*}{\rightarrow}$	22%	30%	$\overset{*}{\rightarrow}$	16%	36%	$\overset{*}{\rightarrow}$	24%	-49%
Dislike rec.	20%	$\overset{*}{\rightarrow}$	8%	42%	$\overset{*}{\rightarrow}$	14%	32%		24%	26%		34%	-30%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

Paper: Table 1

Scrub phase effects

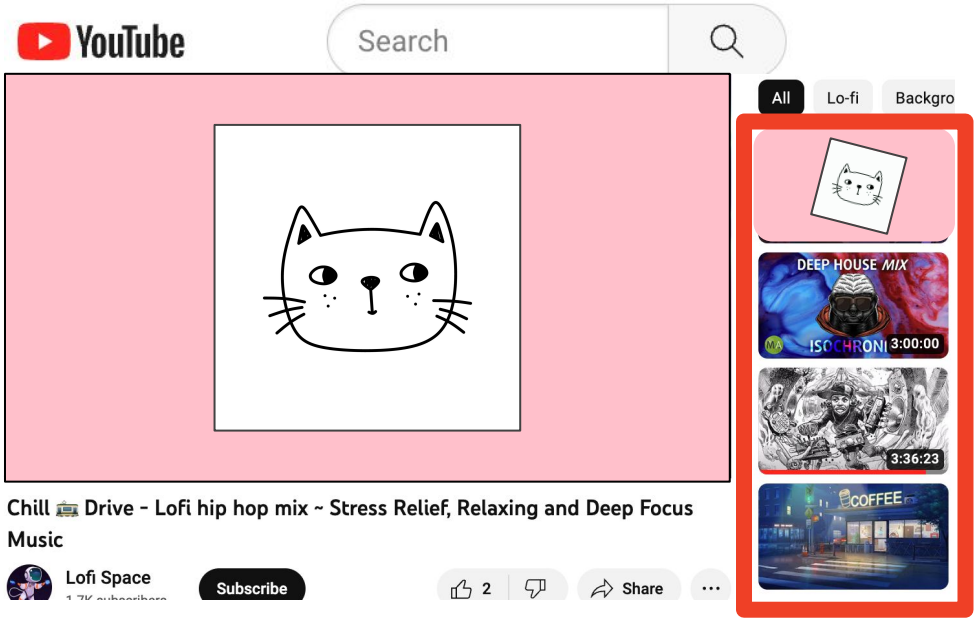
- ⊘ Not interested works best

	Alt-right			*Antitheist*			*Political Left*			*Political Right*			Avg. relative change
	P2		P3	P2		P3	P2		P3	P2		P3	P2 to P3
Watch neutral	12%		8%	34%		22%	32%	$\overset{*}{\rightarrow}$	18%	26%	$\overset{*}{\rightarrow}$	16%	-38%
Delete	20%	$\overset{*}{\rightarrow}$	0%	32%		24%	30%	$\overset{*}{\rightarrow}$	0%	24%	$\overset{*}{\rightarrow}$	4%	-77%
Dislike	20%		4%	30%	$\overset{*}{\rightarrow}$	10%	24%		28%	32%	$\overset{*}{\rightarrow}$	16%	-45%
Not interested	18%	$\overset{*}{\rightarrow}$	0%	42%	$\overset{*}{\rightarrow}$	2%	28%	$\overset{*}{\rightarrow}$	10%	26%	$\overset{*}{\rightarrow}$	2%	-88%
No channel	28%	$\overset{*}{\rightarrow}$	10%	44%	$\overset{*}{\rightarrow}$	22%	30%	$\overset{*}{\rightarrow}$	16%	36%	$\overset{*}{\rightarrow}$	24%	-49%
Dislike rec.	20%	$\overset{*}{\rightarrow}$	8%	42%	$\overset{*}{\rightarrow}$	14%	32%		24%	26%		34%	-30%

$\overset{*}{\rightarrow}$ = significant difference (MWU test)

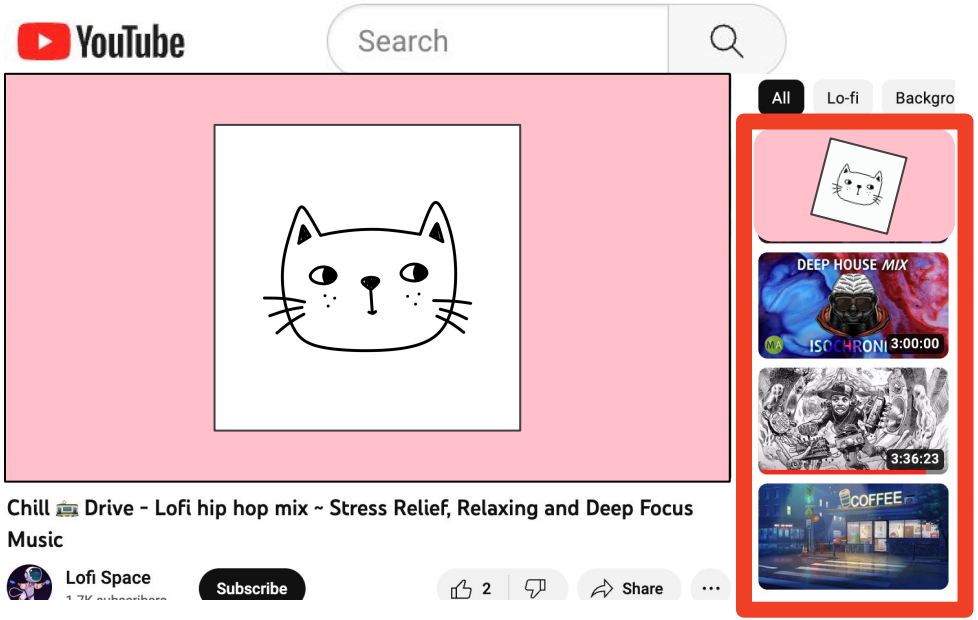
Paper: Table 1

Videopage effects



Videopage effects

- Both the stain phase and scrub phase were largely *in*effective



Outline

1. Introduction
2. Sock puppet audit
3. **User survey**
4. Conclusion

User survey

For every button:

1. Were you **aware** this button existed?
2. Have you **used** it to remove unwanted recommendations (given aware?)
3. How **effective** was it (given used)?

User survey

Not interested :

1. Were you **aware** this button existed? → **56%** ± 7%
2. Have you **used** it (given aware)? → **83%** ± 8%
3. How **effective** was it (given used)? → **3.4/5** ± 0.3

95% CI, post-stratified for target population (US adult YT users)

Summary

Results

- Partial stain occurs
- Many YouTube users (44%) are unaware of the most effective recommendation-removal strategy (

)

Summary

Results

- Partial stain occurs
- Many YouTube users (44%) are unaware of the most effective recommendation-removal strategy (

)

Implications

- Users: Use the

button
- Platform: Make effective features more prominent

Summary

Results

- Partial stain occurs
- Many YouTube users (44%) are unaware of the most effective recommendation-removal strategy (

)

Implications

- Users: Use the

button
- Platform: Make effective features more prominent

… Thank you!

Works cited and sources

- Hosseinmardi, H.; Ghasemian, A.; Clauset, A.; Mobius, M.; Rothschild, D. M.; and Watts, D. J. 2021. Examining the consumption of radical content on YouTube. PNAS.
- Hosseinmardi, H.; Ghasemian, A.; Rivera-Lanas, M.; Horta Ribeiro, M.; West, R.; and Watts, D. J. 2024. Causally estimating the effect of YouTube's recommender system using counterfactual bots. PNAS.
- Lewis, R. 2018. Alternative influence: Broadcasting the reactionary right on YouTube. Data & Society.
- McCrosky, Jesse, and Brandi Geurkink. "YouTube Regrets." Mozilla Foundation. 2021.
- Ribeiro, M. H.; Ottoni, R.; West, R.; Almeida, V. A. F.; and Meira, W. 2020. Auditing radicalization pathways on YouTube. In FAccT.

- YouTube screenshots: Retrieved from YouTube.com. 2024.
- Cat cartoon