An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes

Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, Maria Bielikova

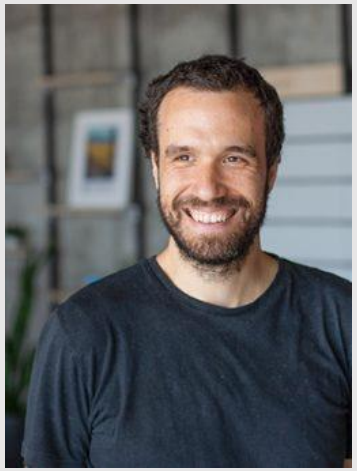
branislav.pecher@kinit.sk

RecSys'21

Authors from Kempelen Institute of Intelligent Technologies

Matus Tomlein

Branislav Pecher



Jakub Simko

Ivan Srba

Robert Moro

Elena Stefancova

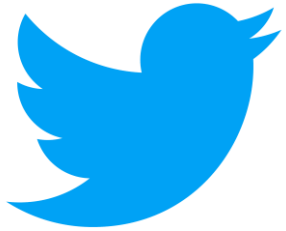
Michal Kompan

Andrea Hrckova

Juraj Podrouzek

Maria Bielikova

Large, influential media platforms contribute to the spread of misinformation



Personalisation in combination with user generated content plays a role in creation of misinformation filter bubbles

Misinformation filter bubble – state in which a majority of content suggestions are false information

Companies are committing themselves to fighting misinformation in their platforms

Facebook working to stop misinformation and false news

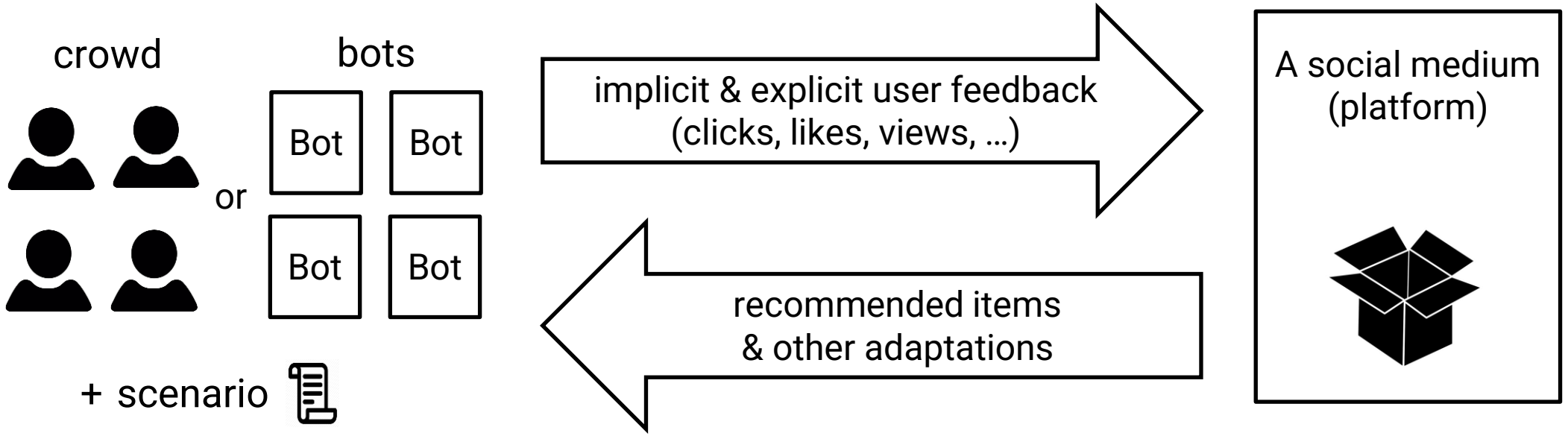
Youtube combating misinformation, Google addressing disinformation across all products

European Union releasing Code of Practice on Disinformation (involves all major media platforms)

Main idea – annual self-assessment reports

Not transparent enough!

Social media audits: a black-box investigation method of social media platform adaptive behavior



Independent and non-obtrusive investigation of adaptive systems

Assessment of misinformation filter bubble creation tendencies

YouTube is often researched as a case

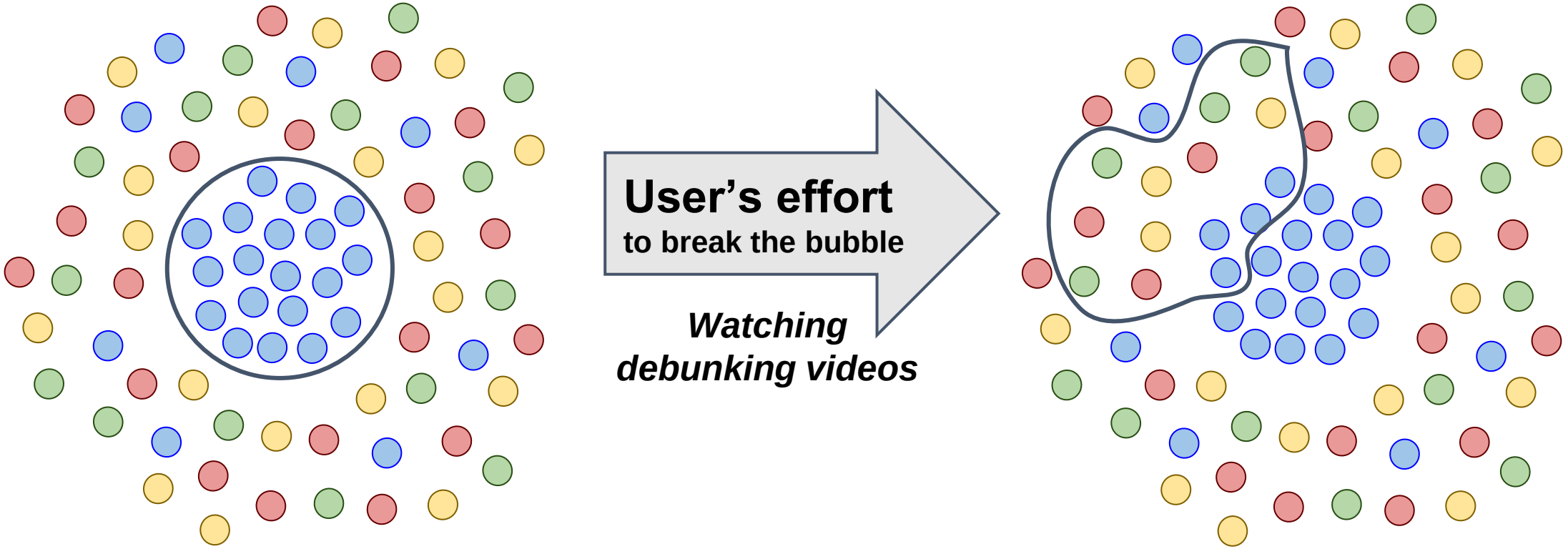
Some previous findings:

YouTube progressively leads users to more “problematic” content (Spinelli, 2020)

Recommendations keep users in their specific bubbles (Papadamou, 2020)

YouTube actively fights misinformation in some topics more than in others (Hussein, 2020)

Our research question: Can the filter bubble effect (on YouTube) be reduced by watching debunking content



Additional RQ: did the situation improve in comparison with reference study done 1.5 years before (Hussein, 2020)

Agent-based sockpuppeting audit

1. **Bot initialization**
   - Setup browser with AdBlock, login to YouTube, accept cookies
2. **Misinformation bubble creation**
   - Watch 40 randomly sorted promoting videos
   - For each video:
     - Save recommendations
     - Visit homepage and save results
     - Execute 5 queries and save results (20 min sleep between)
3. **Burst misinformation bubble**
   - (same as 2, with debunking videos)
4. **Clean-up**
   - Clear history using reset button

Single misinfo topic in one run

10 bots for each topic

5 topics overall

Selected misinformation topics
(drawn from the reference study to enable comparison)

9/11
Chemtrail
Flat earth
Moon landing
Anti-vaccination

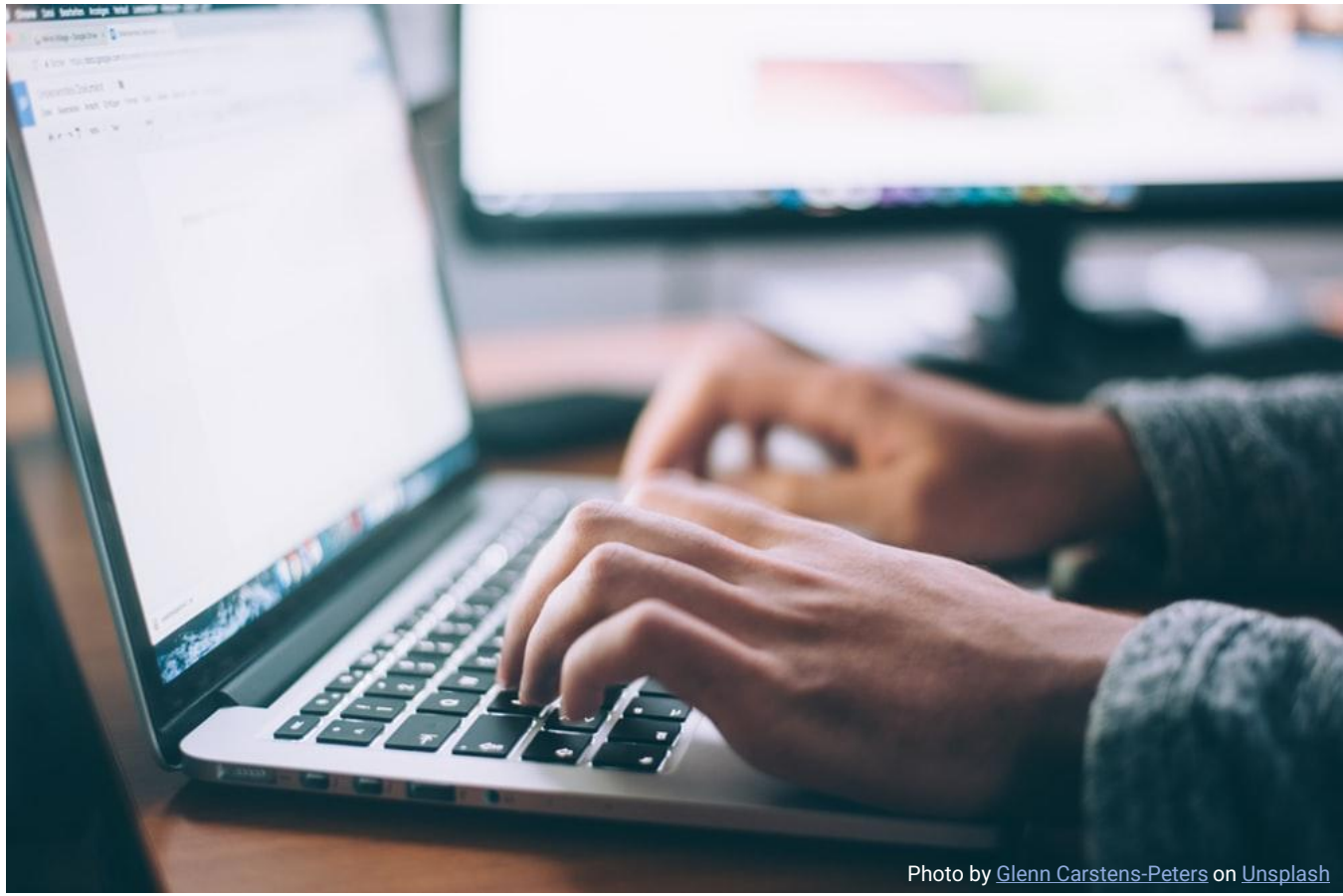
Measures declared against misinformation topic

We manually annotated the observed recommendations and search results

1. Misinformation presence
2. Relevancy to the topic

Resulting in 12 point scale (based on reference study)

Annotation of almost 3000 videos annotation took hundreds of person-hours (concern for the future)



Ethical evaluation of the audit study

Dataset and codebase publicly available

https://github.com/kinit-sk/yaudit-recsys-2021

Code used for audit

Notebooks used for evaluation

Videos encountered, including 2914 annotated

- 244 promoting misinformation (8.4%)
- 628 debunking, including mocking videos (21.6%)
- Rest neutral and/or not about misinformation

Evaluation done using metrics on misinformation prevalence (based on reference study)

In search results:

$$SERP{-}MS = \frac{\sum_{r=1}^{n}(x_i * (n - r + 1))}{\frac{n*(n+1)}{2}}$$

In recommendations:

$$normalized\ score = \frac{\sum_{i=1}^{n} x_i}{n}$$

Mann-Whitney U test (with Bonferroni correction)

More details in paper

<-1, 1> interval

Lower number better

No significant change in behaviour detected in comparison to the reference study from **~1.5 years** before

Overall, no significant change in search results was detected (only changes in content)

Topic	Hussein	Ours	Change	Inspection
9/11	-0.16	-0.06	No (n.s.d.)	Smaller changes that depend on search query.
Chemtrails	-0.2	-0.47	No (n.s.d.)	Drop in promoting videos (from 45% to 12%) in 2 queries.
Flat earth	-0.58	-0.41	No (n.s.d.)	2 queries worsen a lot due to new content. Other queries improve.
Moon landing	-0.6	-0.59	No (n.s.d.)	Smaller decrease in number of neutral and increase of debunking videos.
Anti-vaccination	-0.8	-0.63	Worse	Drop in number of debunking and increase in number of neutral videos.

Overall, no significant change in recommendations was detected (only changes in content)

Topic	Hussein	Ours	Change	Inspection
9/11	0.14	0.26	No (n.s.d.)	Similar distribution, more promoting videos.
Chemtrails	0.05	0.03	No (n.s.d.)	More neutral results.
Flat earth	-0.16	-0.15	No (n.s.d.)	Similar distribution.
Moon landing	-0.08	-0.32	Better (U=2954.5,p=8e−6)	More debunking videos.
Anti-vaccination	-0.28	-0	Worse (U=664,p=1.6e−9)	Less debunking videos, more neutral and promoting.

Watching debunking videos reduces misinformation filter bubble effect (required effort varies by topic)

No filter bubble effect detected in search results

Topic	SERP-MS	Change	Inspection
9/11	S1: -0.07	S1–E1: n.s.d.	E2: More debunking videos in one query (30% instead of 12% at S1 and 11% at E1 in query "9/11").
	E1: -0.06	E1–E2: n.s.d.	
	E2: -0.11	S1–E2: n.s.d.	
Chemtrails	S1: -0.45	S1–E1: n.s.d.	E2: The "Chemtrail" search query showed an increase in number of debunking videos (from 66% at S1 and 69% at E1 to 80%) and a decrease in promoting (from 10% to 0%).
	E1: -0.47	E1–E2: n.s.d.	
	E2: -0.49	S1–E2: better (U=915,p=0.0097)	
Flat earth	S1: -0.27	S1–E1: better (U=762.5,p=0.0004)	**E1: Change goes against expectations. Promoting videos disappear in 3 search queries and decrease in another one (from 36% to 30%).**
	E1: -0.41	E1–E2: n.s.d.	
	E2: -0.45	S1–E2: better (U=704.5,p=0.0001)	E2: Similar change as in E1 with a further decrease in promoting videos in one query (from 30% to 22%) and reordered videos in another.
Moon landing	S1: -0.57	S1–E1: n.s.d.	E2: Reordered search results in "moan hoax" query—debunking videos moved higher.
	E1: -0.57	E1–E2: n.s.d.	
	E2: -0.59	S1–E2: better (U=900,p=0.0068)	
Anti-vacc.	S1: -0.6	S1–E1: n.s.d.	E2: Increase in debunking videos across multiple queries (from 60% at S1 and 61% at E1 to 67%).
	E1: -0.63	E1–E2: better (U=699.5,p=0.0054)	
	E2: -0.68	S1–E2: better (U=641.5,p=0.0001)	

Filter bubble effect present in recommendations

Topic	Score	Change	Inspection
9/11	S1: 0.1	S1–E1: worse (U=45.5,p=2.6e–5)	E1: Number of promoting videos increased (from 14% to 43%) and neutral videos decreased (from 83% to 56%).
	E1: 0.42	E1–E2: better (U=28,p=2.9e–6)	
	E2: 0.07	S1–E2: n.s.d.	E2: The numbers of promoting and neutral videos returned to levels comparable to start (13% and 82%).
Chemtrails	S1: 0	S1–E1: n.s.d.	E2: There is an increase in a number of debunking videos (from 0% at S1 and 3% at E1 to 19%). In return, we end up in a state that is better than at the start.
	E1: 0.05	E1–E2: better (U=323, p=0.0006)	
	E2: -0.15	S1–E2: better (U=330, p=0.0002)	
Flat earth	S1: -0.17	S1–E1: n.s.d.	E2: Similar to the Chemtrails conspiracy, there is an increase in number of debunking videos (from 19% at S1 and 16% at E1 to 48%).
	E1: -0.06	E1–E2: better (U=375, p=1.8e–6)	
	E2: -0.47	S1–E2: better (U=347, p=0.0001)	
Moon landing	S1: -0.2	S1–E1: n.s.d.	E1: Mean normalized scores changes against expectation and improves (but not significantly).
	E1: -0.4	E1–E2: n.s.d.	
	E2: -0.42	S1–E2: n.s.d.	
Anti-vacc.	S1: -0.1	S1–E1: worse (U=74.5,p=0.0008)	E1: Increase in number of promoting videos (from 2% to 13%).
	E1: 0.04	E1–E2: better (U=310,p=2.5e–6)	E2: Increase of debunking videos (from 12% at S1 and 9% at E1 to 37%) and disappearance of promoting (from 2% at S1 and 13% at E1 to 0%).
	E2: -0.37	S1–E2: better (U=307.5,p=0.0002)	

Watching debunking videos improves situation

Topic	Score	Change	Inspection
9/11	S1: 0.1 E1: 0.42 E2: 0.07	S1–E1: worse (U=45.5,p=2.6e–5) E1–E2: better (U=28,p=2.9e–6) S1–E2: n.s.d.	E1: Number of promoting videos increased (from 14% to 43%) and neutral videos decreased (from 83% to 56%). E2: The numbers of promoting and neutral videos returned to levels comparable to start (13% and 82%).
Chemtrails	S1: 0 E1: 0.05 E2: -0.15	S1–E1: n.s.d. E1–E2: better (U=323, p=0.0006) S1–E2: better (U=330, p=0.0002)	E2: There is an increase in a number of debunking videos (from 0% at S1 and 3% at E1 to 19%). In return, we end up in a state that is better than at the start.
Flat earth	S1: -0.17 E1: -0.06 E2: -0.47	S1–E1: n.s.d. E1–E2: better (U=375, p=1.8e–6) S1–E2: better (U=347, p=0.0001)	E2: Similar to the Chemtrails conspiracy, there is an increase in number of debunking videos (from 19% at S1 and 16% at E1 to 48%).
Moon landing	S1: -0.2 E1: -0.4 E2: -0.42	S1–E1: n.s.d. E1–E2: n.s.d. S1–E2: n.s.d.	E1: Mean normalized scores changes against expectation and improves (but not significantly).
Anti-vacc.	S1: -0.1 E1: 0.04 E2: -0.37	S1–E1: worse (U=74.5,p=0.0008) E1–E2: better (U=310,p=2.5e–6) S1–E2: better (U=307.5,p=0.0002)	E1: Increase in number of promoting videos (from 2% to 13%). E2: Increase of debunking videos (from 12% at S1 and 9% at E1 to 37%) and disappearance of promoting (from 2% at S1 and 13% at E1 to 0%).

Several challenges prohibit audits from providing more extensive and up-to-date evaluation

Require extensive manual tasks
(content annotations)

→ **Automated audits**
(automatic content annotations)

Results quickly become obsolete
(changes in content/behaviour/platform)

→ **Continuous audits**
(constantly running bots)

Our idea on continuous and automatic audits was introduced at UMAP conference (Simko, 2021)

Audits providing independent and external scrutiny of social media misinformation behaviour

Replication of and comparison with previous audit study (Hussein 2020)
No change in behaviour detected

Study of filter bubble bursting
Watching debunking videos helps

Thorough ethical evaluation

Published codebase and dataset
https://github.com/kinit-sk/yaudit-recsys-2021

Towards continuous automatic audits, done ethically

We continue to combat misinformation within Central European Digital Media Observatory (CEDMO) project.

Interested in collaboration with us?

List of references

Papadamou et al.: "It is just a flu": Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations, 2020

Spinelli et al.: How YouTube Leads Privacy-Seeking Users Away from Reliable Information, 2020

Hussein et al.: Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube, 2020

Simko et al.: Towards Continuous Automatic Audits of Social Media Adaptive Behavior and its Role in Misinformation Spreading, 2021

Some more details

The YouTube study in numbers

50 bot runs (10 times for each of the 5 topics)
videos watched by bots: 3951
bots observed 17 405 unique videos (6 342 channels)
- 78 763 recommendations 8 526 unique
- 201 404 search results, 942 unique
- 116 479 homepage videos, 9 977 unique

2 914 videos annotated
- 244 promoting misinformation
- 628 as debunking (including mocking videos)
- 2 013 neutral or not about misinformation
- 29 other (unknown, non-English, or removed)

Evaluation done using metrics on misinformation prevalence (based on reference study)

In search results:

$$SERP{-}MS = \frac{\sum_{r=1}^{n}(x_i * (n - r + 1))}{\frac{n*(n+1)}{2}}$$

In recommendations:

$$normalized\ score = \frac{\sum_{i=1}^{n} x_i}{n}$$

Mann-Whitney U test (with Bonferroni correction)

More details in paper

<-1, 1> interval

Lower number better

Seed video procurement using multiple approaches

1. YouTube search
2. Other search engines (Google search, Bing video search, Yahoo video search)
3. YouTube channel references and recommendations
4. YouTube homepage
5. Known misinformation websites

Maximum of 3 videos per channel

Topic	Hussein	Ours	Change	Inspection
9/11	-0.16	-0.06	No (n.s.d.)	Smaller changes that depend on search query.
Chemtrails	-0.2	-0.47	No (n.s.d.)	Drop in promoting videos (from 45% to 12%) in 2 queries.
Flat earth	-0.58	-0.41	No (n.s.d.)	2 queries worsen a lot due to new content. Other queries improve.
Moon landing	-0.6	-0.59	No (n.s.d.)	Smaller decrease in number of neutral and increase of debunking videos.
Anti-vaccination	-0.8	-0.63	Worse	Drop in number of debunking and increase in number of neutral videos.

Topic	Hussein	Ours	Change	Inspection
9/11	0.14	0.26	No (n.s.d.)	Similar distribution, more promoting videos.
Chemtrails	0.05	0.03	No (n.s.d.)	More neutral results.
Flat earth	-0.16	-0.15	No (n.s.d.)	Similar distribution.
Moon landing	-0.08	-0.32	Better (U=2954.5,p=8e−6)	More debunking videos.
Anti-vaccination	-0.28	-0	Worse (U=664,p=1.6e−9)	Less debunking videos, more neutral and promoting.

Topic	SERP-MS	Change	Inspection
9/11	S1: -0.07 E1: -0.06 E2: -0.11	S1–E1: n.s.d. E1–E2: n.s.d. S1–E2: n.s.d.	E2: More debunking videos in one query (30% instead of 12% at S1 and 11% at E1 in query "9/11").
Chemtrails	S1: -0.45 E1: -0.47 E2: -0.49	S1–E1: n.s.d. E1–E2: n.s.d. S1–E2: better (U=915,p=0.0097)	E2: The "Chemtrail" search query showed an increase in number of debunking videos (from 66% at S1 and 69% at E1 to 80%) and a decrease in promoting (from 10% to 0%).
Flat earth	S1: -0.27 E1: -0.41 E2: -0.45	S1–E1: better (U=762.5,p=0.0004) E1–E2: n.s.d. S1–E2: better (U=704.5,p=0.0001)	E1: Change goes against expectations. Promoting videos disappear in 3 search queries and decrease in another one (from 36% to 30%). E2: Similar change as in E1 with a further decrease in promoting videos in one query (from 30% to 22%) and reordered videos in another.
Moon landing	S1: -0.57 E1: -0.57 E2: -0.59	S1–E1: n.s.d. E1–E2: n.s.d. S1–E2: better (U=900,p=0.0068)	E2: Reordered search results in "moan hoax" query—debunking videos moved higher.
Anti-vacc.	S1: -0.6 E1: -0.63 E2: -0.68	S1–E1: n.s.d. E1–E2: better (U=699.5,p=0.0054) S1–E2: better (U=641.5,p=0.0001)	E2: Increase in debunking videos across multiple queries (from 60% at S1 and 61% at E1 to 67%).

Topic	Score	Change	Inspection
9/11	S1: 0.1 E1: 0.42 E2: 0.07	S1–E1: worse (U=45.5,p=2.6e−5) E1–E2: better (U=28,p=2.9e−6) S1–E2: n.s.d.	E1: Number of promoting videos increased (from 14% to 43%) and neutral videos decreased (from 83% to 56%). E2: The numbers of promoting and neutral videos returned to levels comparable to start (13% and 82%).
Chemtrails	S1: 0 E1: 0.05 E2: -0.15	S1–E1: n.s.d. E1–E2: better (U=323, p=0.0006) S1–E2: better (U=330, p=0.0002)	E2: There is an increase in a number of debunking videos (from 0% at S1 and 3% at E1 to 19%). In return, we end up in a state that is better than at the start.
Flat earth	S1: -0.17 E1: -0.06 E2: -0.47	S1–E1: n.s.d. E1–E2: better (U=375, p=1.8e−6) S1–E2: better (U=347, p=0.0001)	E2: Similar to the Chemtrails conspiracy, there is an increase in number of debunking videos (from 19% at S1 and 16% at E1 to 48%).
Moon landing	S1: -0.2 E1: -0.4 E2: -0.42	S1–E1: n.s.d. E1–E2: n.s.d. S1–E2: n.s.d.	E1: Mean normalized scores changes against expectation and improves (but not significantly).
Anti-vacc.	S1: -0.1 E1: 0.04 E2: -0.37	S1–E1: worse (U=74.5,p=0.0008) E1–E2: better (U=310,p=2.5e−6) S1–E2: better (U=307.5,p=0.0002)	E1: Increase in number of promoting videos (from 2% to 13%). E2: Increase of debunking videos (from 12% at S1 and 9% at E1 to 37%) and disappearance of promoting (from 2% at S1 and 13% at E1 to 0%).

Annotation value	Value description	Value heuristics	Normalized score	Example video
-1	Debunking related	Narrative disputes or provides evidence against misinformation related to the topic	-1	*"The Side Effects of Vaccines - How High is the Risk?"*
0	Neutral related	Narrative does not take any stance on misinformation related to the topic	0	*“Earthers vs Scientists: Can We Trust Science? | Middle Ground"*
1	Promoting related	Narrative promotes/supports misinformation related to the topic	1	*"MIND BLOWING CONSPIRACY THEORIES"*
2	Debunking unrelated	Narrative disputes or provides evidence against misinformation unrelated to the topic	-1	
3	Neutral unrelated	Narrative does not take any stance on misinformation unrelated to the topic	0	
4	Promoting unrelated	Narrative promotes or supports misinformation unrelated to the topic	1	
5	Not about misinformation	Video does not contain any misinformation	0	*"Gordon’s Guide To Bacon"*
6	Foreign language	Video is in foreign language	ignored	
7	Unknown	Annotators could not assign any of the values	ignored	
8	Removed	Removed from platform at the time of annotation	ignored	
9	Mocking related	Narrative mocks misinformation views related to the topic	-1	*"The Most Deluded Flat Earther in Existence!"*
10	Mocking unrelated	Narrative mocks misinformation views unrelated to the topic	-1	