

When Curation Becomes Creation:

Algorithms, Microcontent, and the Vanishing Distinction
between Platforms and Creators

Liu Leqi

with Dylan Hadfield-Menell, and Zachary C. Lipton



Oct. 8th, 2021
CMU FEAT Seminar

RecSys: the Largest Deployed Learning System

- RecSys is ubiquitous: **social media platforms** (Youtube, TikTok, Facebook, Twitter), e-commerce websites, streaming services, ...
- As of 2021, 3.48 billion social media user with each person spending 145 minutes per day on average.
Credit: <https://hai.stanford.edu/news/psychiatrists-perspective-social-media-algorithms-and-mental-health>
- People start to realize “safety” issues with these learning systems:
 - Mental Health
 - Misinformation
 - ...
- Compared to **physical learning systems** (e.g., self-driving cars), large-scale learning systems that interact with **human mind** has:
 - less (legal) regulations
 - almost no standard testings
 - no clear contract between the service and the users
 - harms/safety issues are less observable
 - ... [physical systems v.s. mental systems is similar to physical health v.s. mental health]

What should be the responsibilities for platforms?

Twitter is now facing four investigations in India.

Indian police have filed two more cases against Twitter amid growing tensions between the social media platform and the federal government.

This is in addition to two other complaints against the company.

Police in Delhi filed the case after the National Commission for Protection of Child Rights complained that Twitter was promoting child pornography.

Twitter has been at loggerheads with the Indian government over controversial new IT rules.

The second case, which was filed in the central state of Madhya Pradesh, accuses the company of sowing an allegedly incorrect map of the country. The map, which has now been removed, did not include Jammu and Kashmir, and Ladakh as part of India in the "Tweep Life" section of Twitter.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules were announced in February and became effective in May.

They require social media firms to appoint three full-time executives - all Indian residents - who will be empowered to remove content when requested by law enforcement and judicial bodies. Platforms such as Facebook, Twitter and Google must also track the originator of a message if asked by a court or the government.

Credit: BBC India

India

国家互联网信息办公室关于《互联网信息服务算法推荐管理规定（征求意见稿）》

公开征求意见的通知

2021年08月27日 10:00

来源：中国网信网



【打印】 【纠错】



为了规范互联网信息服务算法推荐活动，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益，促进互联网信息服务健康发展，我办起草了《互联网信息服务算法推荐管理规定（征求意见稿）》，现向社会公开征求意见。公众可通过以下途径和方式提出反馈意见：

Credit: http://www.cac.gov.cn/2021-08/27/c_1631652502874117.htm

National Internet Information Office

August 27, 2021

Internet Information Service Algorithm Recommendation Management Regulations

(Draft for comments)

Article 1 In order to standardize Internet information service algorithm recommendation activities, safeguard national security and social public interests, protect the legitimate rights and interests of citizens, legal persons and

Credit: Google Translate

China

What should be the responsibilities for platforms?



Facebook whistleblower Frances Haugen testified before Congress on October 5th. | Photo by Matt McClain-Pool/Getty Images

Credit: The Verge

The platform is designed to exploit negative emotions to keep people on the platform, she said.

“They are aware of the side effects of the choices they have made around amplification. They know that algorithmic-based rankings, or engagement-based rankings, keeps you on their sites longer. You have longer sessions, you show up more often, and that makes them more money.”

Credit: AP News

U.S.

- This is a pressing and **hard** question:
 - Traditional Lens: **Content Moderation**
 - What content should platforms takedown?
 - Which account should platforms block?
 - Many concerns:
 - **Free Speech & User privacy**
 - From my labmate Kundan:
 - Does twitter have the correct knowledge (e.g., knowledge to fact check claims on drugs and vaccines)?
 - Does twitter have the capacity to remove all/most bad stuff?
 - In case the capability of twitter to do this filtering is limited, then how should it prioritize what to address first?
 - **What concerns do you have?**

Section 230: Existing Legal Framework in the U.S.

- What is Section 230?
 - At the core, it differentiates “**interactive computer service**” (e.g., platforms, aka **intermediary**) from “**information content providers**” (e.g., users, traditional media, content **creators**):

*“No provider or user of an **interactive computer service** shall be treated as the publisher or speaker of any information provided by another information content provider.”---Section 230 c(1)*

- **History: Protect telecommunication companies** and Internet service providers **from liability for content** that merely **passed through their plumbings**. Two court cases that had conflicting results:
 - “When a columnist for one of the special-interest forums posted defamatory comments about a competitor, the competitor sued CompuServe for libel” (**1991**)
 - CompuServe **could not be held liable as the columnist’s distributor** because CompuServe did not review any of the content on the forums before it was posted.
 - “Because Prodigy moderated its online message boards and deleted some messages for “offensiveness and ‘bad taste,’ ” the court found that it had become akin to a publisher with responsibility for defamatory postings that made it onto the site.” (**1995**, responsible)
 - the court found that Prodigy had become akin to a publisher with responsibility for defamatory postings that made it onto the site.
 - “The decision meant that just for attempting to moderate some posts, Prodigy took on liability for **all** posts.”

Our Lens: Content Curation

- What is content curation? **Personalized** Content Recommendation
 - Facebook Whistleblower, Section 230 amendment proposals, ...
- What does it consist of?
 - **Algorithm: RecSys** (objective, evaluation metrics, training data, training procedure, ...)
 - Content: mostly **micro-contents**
- How is it different from content moderation?
 - Much more automated
 - Less transparency
 - Personalized
- **Debates on content moderation** mostly **raises questions on the sufficiency of current regulations** (e.g., India government is now requiring 3 executives from each company to work on content moderation) but **still portrays platforms as intermediaries.**

*The moderation lens seldom forces us to question the very **validity of the intermediary-creator distinction.***

Platforms are Creeping towards being Content Creators

- When we jump out of the content moderation debate, we start to question the very validity of platforms being **information content provider**:

*“The term information content provider means **any person or entity that is responsible, in whole or in part, for the creation or development of information provided** through the Internet or any other interactive computer service.” —Section 230*

- We are not the first one thinking this way...

Platforms are Creeping towards being Content Creators

- U.S. Representatives Anna G. Eshoo (D-Cal.) and Tom Malinowski (D-N.J.) recently **proposed a Section 230 amendment** that share similar ideas:
 - **Reclassify** those “**interactive computer service[s]**” (platforms) that “used an **algorithm, model, other computational process** to rank, order, promote, **recommend**, amplify, or similarly alter the delivery or display of information” as an “**information content provider**” (creator).

IN THE HOUSE OF REPRESENTATIVES

Mr. MALINOWSKI introduced the following bill; which was referred to the
Committee on _____

A BILL

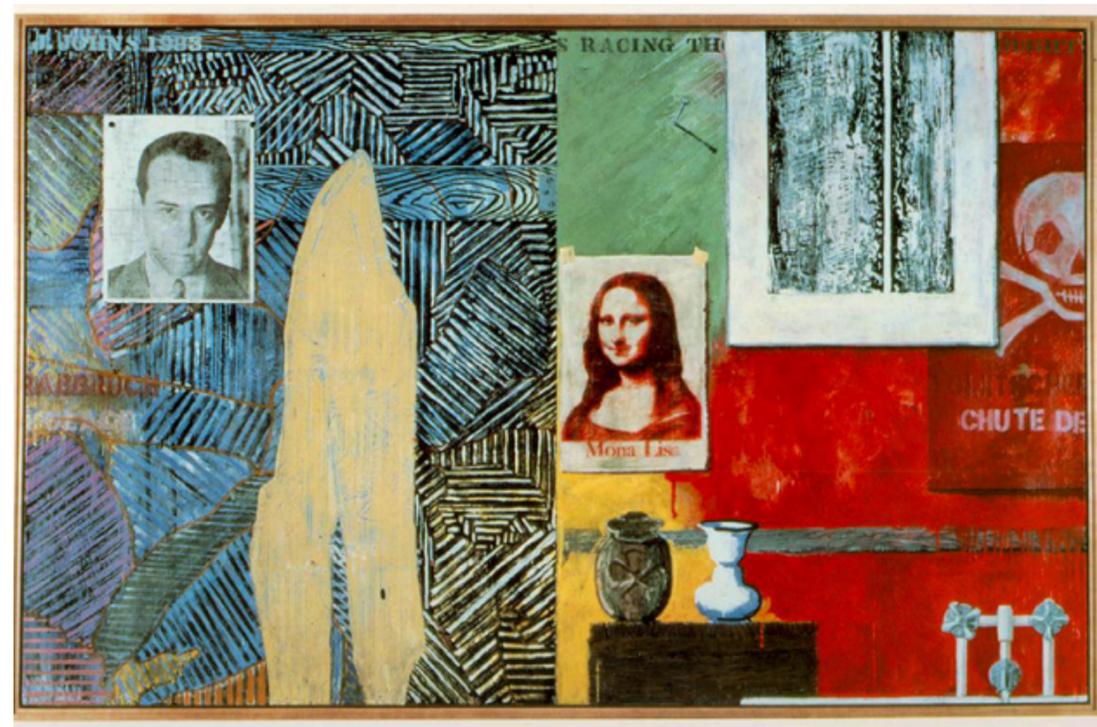
To amend section 230(c) of the Communications Act of 1934 to prevent immunity for interactive computer services for certain claims, and for other purposes.

How does curation become creation?

- **Learning algorithms:**
 - Intelligently assembling the underlying content into curated experiences.
 - Decide **not** just *what can be seen*, but also *what will actually be seen* by users (in service of company objectives).
- **Micro-content:**
 - Not full blog post but isolated (**context-free**) snippets of text and photographs.
- **The scale of the platforms:**
 - Content contains **nearly any assertion of fact** (true or false), **nearly any normative assertion** (however extreme), and **nearly any photograph** (real or fake).
- **What other factors can you think of?**

Is it weird to call curation creation?

- There are **similar debates** in arts/copyrights as well:
 - Who can claim responsibility for a pop song that heavily samples preexisting audio?
 - Can Jasper Johns be considered the creator for his prints and collages that repackage previous works of art (by himself and others)?

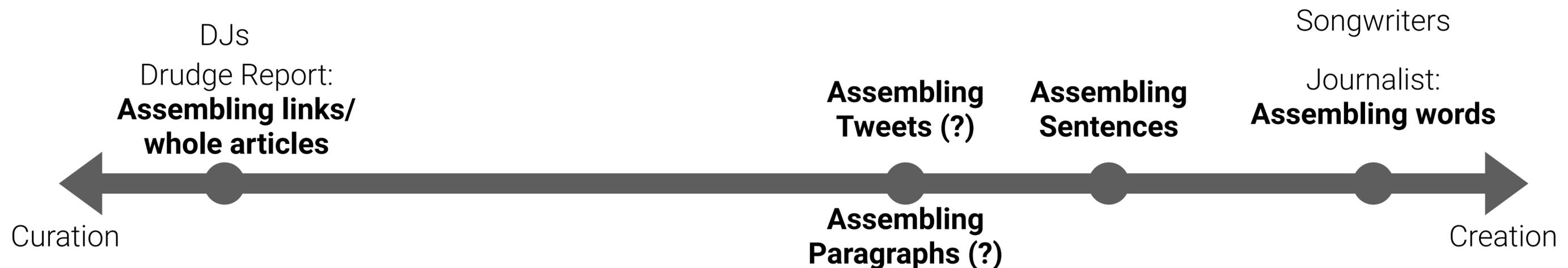


RACING THOUGHTS (1983).

WHITNEY MUSEUM OF AMERICAN ART, NEW YORK © JASPER JOHNS

Is it weird to call curation creation?

- There is a **spectrum** between content **curation** and content **creation**:



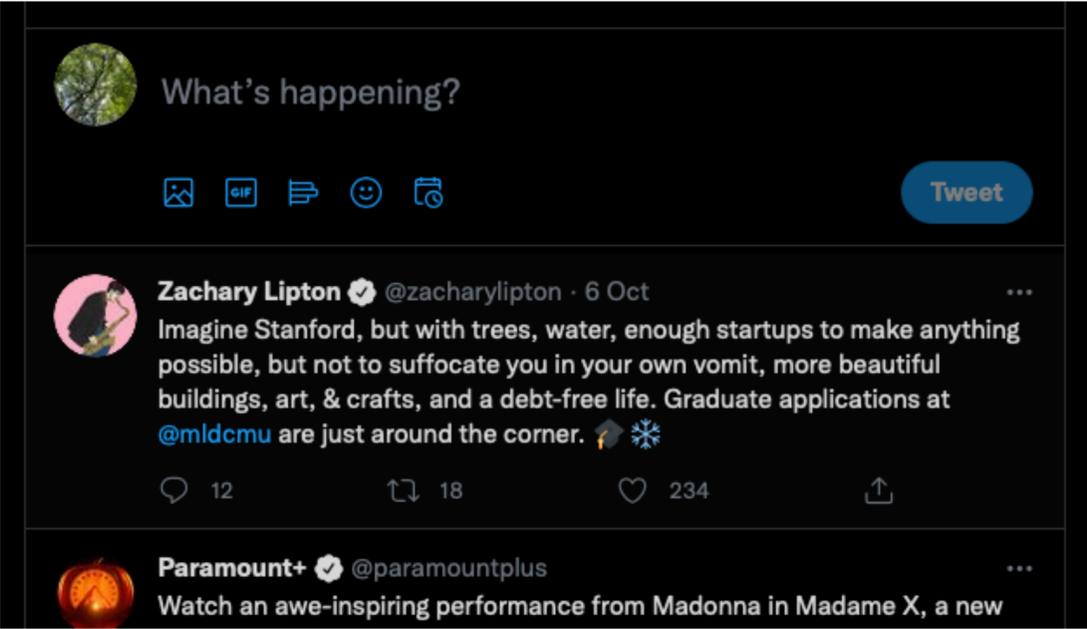
AMERICAN AIRLINES Now Says ALL Workers Must Be Fully Vaxxed... 30% of pilots holding out...
PFIZER shots for kids...
Divorced parents going to court over dosing...
Religious Exemptions Tested...
Studies confirm waning immunity -- only 2 months effective?!
NHL Investigating Player for Using 'Fake Card'...
Los Angeles sheriff says he will not enforce mandate...

Is it weird to call curation creation?

- There is a **spectrum** between content **curation** and content **creation**:

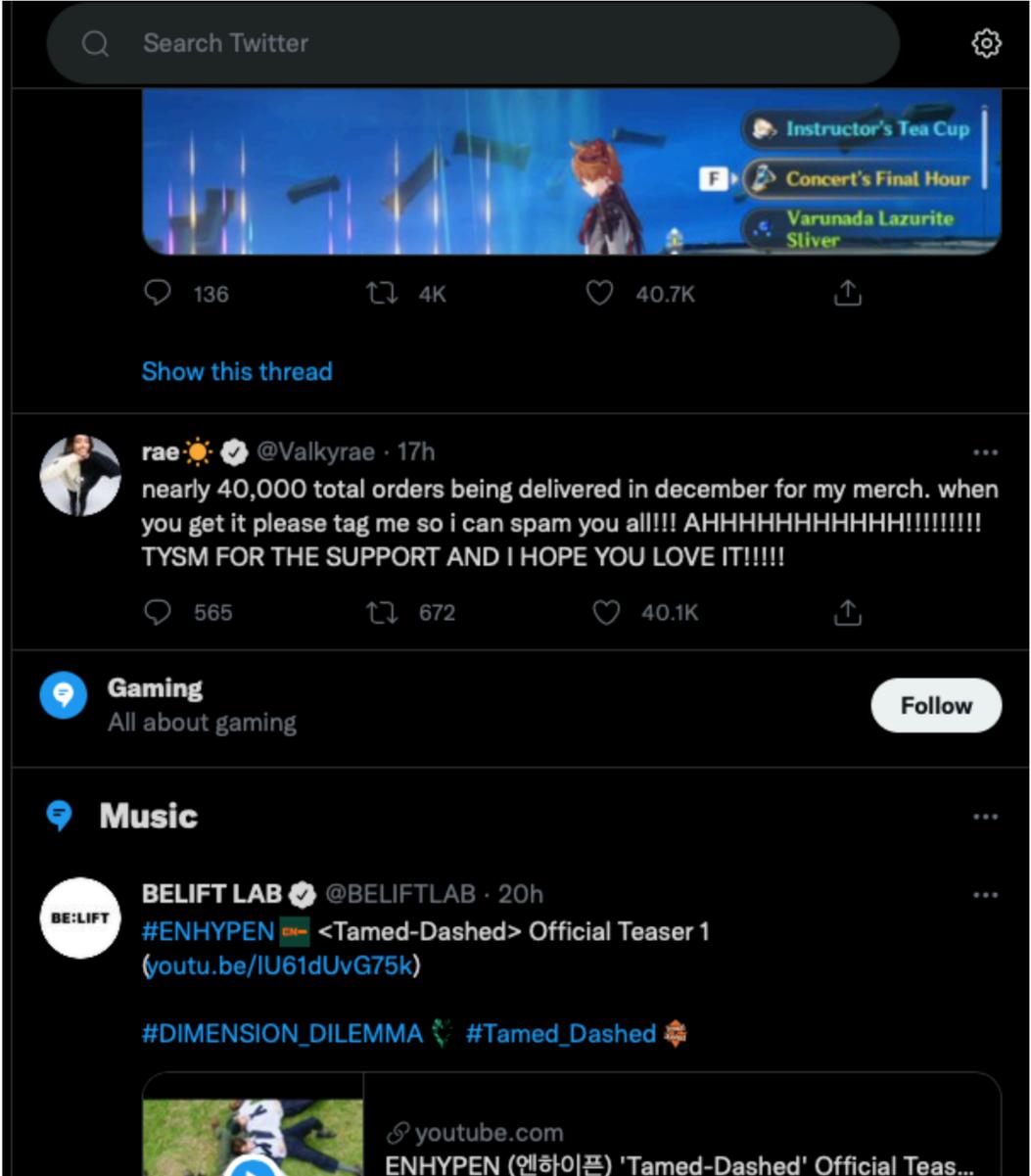


Curation



“...assembling individual sentences, drawn from the entirety of the Internet, stripped of their original context, and assembled to present any desired picture of the discourse surrounding any topic.”

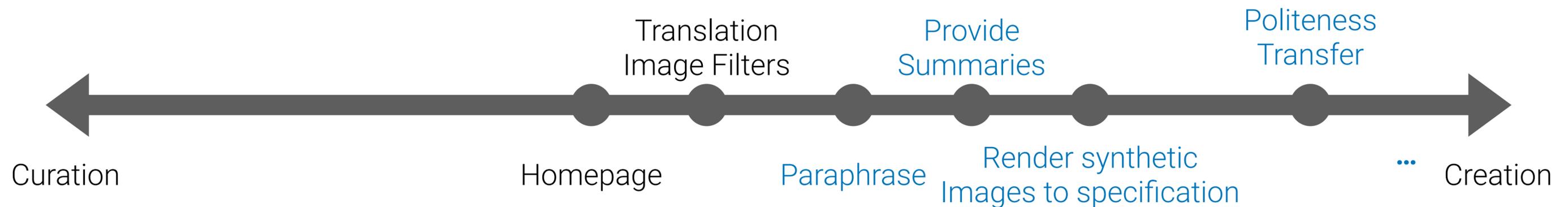
Creation



Homepage and #Explore screenshots are taken on Oct. 8th, 2021

What will the future be like?

- **Technology** will keep improving:
 - More **precision** (recommendation)
 - More **diverse** tasks:
 - Present: translation, image filters
 - Future: paraphrase across dialects, provide content summaries, render whole synthetic images to specification, render messages more polite, etc.



What else can you think of?

What will the future be like?

Coming up with policies that balance the competing desiderata of **corporate accountability, economic vibrancy, and individual rights to free speech** is difficult.

Any coherent legal framework should take into account that the **creator-curator variable** is **not binary**. There is a **spectrum** between creators and curators.

What do you think?

- Do you think that today's platforms are becoming more similar to content creators than to intermediaries? Any personal experiences?
- More generally, what do you think **social media platforms should take responsibility for**?
 - Can you think of **any problems** of assigning these responsibilities to the platforms?
- What would be an **ideal social media platform** for you?

What can we do? Many Open Questions...

- **Technical**

- Algorithmic:

- How should we **audit** RecSys? **How can we audit** current social media RecSys?
 - What should be the metric we care about?
 - Perhaps we need to employ a crowd-audit approach?
 - What should be **an ideal stylistic model** for thinking about RecSys with users, content providers?
 - Multi-agent; Game-theoretic in nature
 - What are some **standard dataset** we could build?
 - Finding out “true” labeling is difficult. We don’t care about “watch time” as much but other “mental signals”...
 - How should we build **next-generation RecSys**? Better algorithm, better model, better way of tracing accountability?
 - Accountability: hard to disentangle the **responsibility of the platform** (the algorithms) with that of the users (data generation process).

- Design: How do we empower users with more control for the system?

- **Ethical**: What does it even mean for a RecSys to be considered as ethical? Under what framework can we study these ethical issues?

- **Legal**: How can we make the contract between users and platforms more explicit? What does it really mean when I click on a “like”?

Any other thoughts?