YouTube’s recommendation system determines what content should appear on a user’s home page and in the user’s “Up Next” sidebar, which appears next to videos that a user is currently watching. The Up Next feature autoplays recommended content unless a user turns the autoplay off.

YouTube is one of the largest video repositories on the internet, and many users incorrectly equate the site’s popularity with the credibility of its recommendation system. However, despite the fact that YouTube’s recommendation system is responsible for shaping how billions of individuals engage with content on the service, and influencing how they see the world, the company has provided relatively little transparency around how this system works.

This lack of transparency is concerning, as the company’s recommendation system has been found to suggest controversial and harmful videos, including those that promote extremist propaganda, conspiracy theories, and misinformation.

According to a YouTube spokeswoman, in late 2016, the company adopted social responsibility as a core value for the company.**82** During this time, the recommendation system was altered so it considered inputs such as how many times a video was shared, liked, and disliked. **83** These changes were introduced amidst growing pressure on internet platforms to be more proactive in their efforts to combat harmful content, such as extremist propaganda, disinformation and misinformation, and content unsafe for children.**84** Recently, the company has provided more detail around this concept of responsibility, outlining that it consists of the four Rs of responsibility:**85**

* Removing content that violates the platform’s Community Guidelines as quickly as possible
* Raising up authoritative information sources, especially during breaking news moments
* Reducing the spread of content that comes close to violating, but does not violate the platform’s Community Guidelines (known as borderline content)
* Rewarding trusted creators

However, this comes with trade-offs: It is challenging to define authoritative sources across more subjective verticals, as these determinations are based on personal preference and taste.**102**As described above, YouTube provides very little transparency and accountability around how its recommendation system is structured, how it operates, and how it makes decisions.**118** Research has suggested that promoting awareness of the use of algorithmic tools and enabling users to control their own experiences on a platform are fundamental steps in building trust with users. This lack of transparency from YouTube therefore limits the agency users have over their own experiences.**1**

Although YouTube has stated that conspiracy theory videos make up less than 1 percent of all content on the platform, this is still a staggering amount of content, and the problem is compounded whenever the recommendation algorithm promotes this content.

In addition, YouTube has faced particular criticism for creating a “rabbit hole” effect, in which the algorithm delivers personalized recommendations that prompt users to consume harmful or radical content**130** that they did not originally seek out.

former YouTube employee-turned-critic Guillaume Chaslot, have argued that it is in the company's business interest to promote such polarizing and fringe videos and channels, as they drive engagement and greater watch times. Further, critics such as Chaslot have suggested that the recommendation system is biased toward promoting divisive, sensational, and conspiratorial content,**138** perhaps because the system has learned that such content is engaging.**139** Given the vast number of users who consume recommended content, this raises significant concerns about the platform serving as a radicalization pipeline.**140**

This raises concerns that the company is placing profits over ensuring the company’s use of automated tools is responsible, transparent, and accountable.

As highlighted, YouTube does not provide significant transparency around how its recommendation system operates, thus limiting the agency users have over their personal YouTube experience. The company does, however, offer its users a limited set of controls over how this system shapes their platform experience.

Market dominance is, as such, not illegal under EU antitrust rules. However, dominant companies have a special responsibility not to abuse their powerful market position by restricting competition, either in the market where they are dominant or in separate markets.

**Google has abused this market dominance by preventing rivals from competing in the online search advertising intermediation market**.

Based on a broad range of evidence, the Commission found that Google's conduct harmed competition and consumers, and stifled innovation. Google's rivals were unable to grow and offer alternative online search advertising intermediation services to those of Google. As a result, owners of websites had limited options for monetizing space on these websites and were forced to rely almost solely on Google.

Google's practices covered over half the market by turnover throughout most of the period. Google's rivals were not able to compete on the merits, either because there was an outright prohibition for them to appear on publisher websites or because Google reserved for itself by far the most valuable commercial space on those websites, while at the same time controlling how rival search adverts could appear.

currently user-centred approaches do not consider the interests of a variety of other stakeholders—as opposed to just the receivers of a recommendation—in assessing the ethical impacts of a recommender system.

Specifying these parameter choices is highly dependent on the domain of application and the level of abstraction [LoAs, see (Floridi [2016](https://link.springer.com/article/10.1007/s00146-020-00950-y#ref-CR14))][Footnote1](https://link.springer.com/article/10.1007/s00146-020-00950-y#Fn1) from which the problem is considered (Jannach et al. [2012](https://link.springer.com/article/10.1007/s00146-020-00950-y#ref-CR27)). Typically, the literature implements three LoAs: catalogue-based, decision support, and multi-stakeholder environment. Let us consider each of these in turn.

The value of some consequences is often measured in terms of the utility they contain. So, it is reasonable to assume that any aspect of a RS that could impact negatively the utility of any of its stakeholders, or risk imposing such negative impacts, constitutes a feature that is ethically relevant.

 In the case of RS, for example, the risks may involve exposing users to undue privacy violations by external actors, or the exposure to potentially irrelevant or damaging content. Exposure to risks of these sorts can constitute a wrong, even if no adverse consequences actually materialize

Typically, the debate focuses on the fact that these sites are owned by private companies. One side argues that, because these are private companies, they should be allowed to choose who to have as their customers. The other side argues that, because these are vital communication networks, private companies shouldn’t be allowed to ban any individuals.

So creating a regime in which a social media platforms could be designated as a common carrier – whether at the discretion of the platform owner or as a result of criteria laid down in the regime – would not put an end to social media bans. But the imposition of the banning orders would almost certainly be handed to an external agency, as should [the rules](https://simoncarne.substack.com/p/social-media-regulation-publisher-platform) for what can and can’t be posted on such sites.

In the years since, we’ve learned how these kinds of targeted ads can create political [filter bubbles](https://www.techopedia.com/definition/28556/filter-bubble) and [echo chambers](https://www.techopedia.com/definition/23423/echo-chamber), suspected of [dividing people](https://journals.sagepub.com/doi/abs/10.1177/0266382117722446?casa_token=2ayQdG9GskoAAAAA%3ArSCq4yOZ5x33tVexv0EX4jPRqQ7SNCK7z8Pfm42ooHea4Y_VdILuTGVEe7lC3CqJg8Cv1QM9mOx43g&journalCode=bira) and increasing the circulation of [harmful disinformation](https://www.tandfonline.com/doi/abs/10.1080/10584609.2021.1910887?journalCode=upcp20).

Machine learning algorithms are bad at identifying contextual harms. On the contrary, the way targeting works actually amplifies them. Several [audits](https://arxiv.org/ftp/arxiv/papers/2008/2008.09656.pdf), for example, have uncovered how Facebook has allowed [discriminatory targeting](https://venturebeat.com/2020/08/28/ai-weekly-facebooks-discriminatory-ad-targeting-illustrates-the-dangers-of-biased-algorithms/) that worsens socioeconomic inequalities.

The root cause of all these issues can be traced to the fact that consumers have a very isolated experience online. We call this a state of “[epistemic fragmentation](https://www.nature.com/articles/s42256-021-00358-3),” where the information available to each individual is limited to what is targeted at them, without the opportunity to compare with others in a shared space like the London Underground.