Tackling Discrimination in Targeted Advertising

US regulators take very small steps in the right direction – but where is the EU?

Rachel Griffin

23.06.2022 doi: 10.17176/20220623-153440-0

It's a cliché by now to note that social media platforms' real clients are advertisers, not users. On June 21, Meta (formerly Facebook) and the US Department for Housing and Urban Development released a legal settlement that will restrict Meta's ability to offer those clients some of its core ad-targeting products. It resolves (for now) a long-running case over discriminatory targeting of housing adverts – for example, allowing landlords to exclude people of colour, or automatically showing ads to mostly white audiences. Meta is now prohibited from using certain targeting tools in this context, and has promised new tools to ensure more representative targeting. The company has announced it will extend the same measures to employment and credit adverts, also areas where discrimination is illegal and nonetheless pervasive.

What are the implications of this settlement? These announcements raise more questions than they answer, and leave much to be desired when it comes to redressing systemic discrimination. Yet at least the lawsuit and the settlement represent efforts by the US government to proactively tackle algorithmic discrimination, leading to concrete changes in harmful business practices.

It's worth noting, then, that these new measures will only apply within the US. So what are European policymakers doing? European institutions may see themselves as leaders in tech regulation, and defenders of 'European values' like equality, but their actions don't measure up. EU discrimination and data protection law are inadequate for contemporary forms of automated discrimination, and the relevant provisions of the Digital Services Act (DSA) are largely symbolic. This US lawsuit should be a wake-up call for European regulators, reminding them that taking systemic discrimination seriously requires proactive regulatory reform and enforcement.

What has Meta agreed and why?

Discriminatory targeting of information like housing and job adverts can directly exclude minorities from material opportunities. Conversely, it can also facilitate 'predatory inclusion', deliberately targeting poor or vulnerable people with exploitative products like gambling and payday loans. Meta's tools have repeatedly been shown to enable and perpetuate these kinds of discrimination. On multiple occasions, Meta has changed its ad-targeting rules after journalists revealed how advertisers can discriminate – for example, no longer allowing targeting or exclusion based on 'ethnic affinity', the proxy for race that Meta previously assigned based on people's behaviour and interests – only for subsequent investigations to reveal deliberate discrimination is still easy.

Even when advertisers have no intention to discriminate, machine learning-based targeting software can still generate highly unrepresentative audiences, because it's optimised to show ads to the users deemed most likely to click on them. Accordingly, if past data shows that, for example, men were more interested in adverts for senior executive positions, it's logical to show those adverts to male users. As Wendy Chun argues, this is an inherent tendency of today's big data systems, since their usefulness rests on the assumption that the future will be basically similar to the historical data they were trained on. If they are trained on information about racist and sexist social systems – say, a housing market founded on extreme economic inequality and *de facto* segregation, or a job market where men dominate senior roles – they will inevitably make racist and sexist predictions.

Importantly, Meta doesn't even need to collect data on sensitive characteristics like race to enable these forms of discrimination. The same results can be achieved – deliberately or unintentionally – by factoring in proxies that correlate with those characteristics, like neighbourhood, language and interests. For example, if an advert targets or excludes people in Germany who speak Turkish or watch Turkish-language TV shows, it's obvious that most of them won't be white.

The settlement aims to address these issues with two key commitments. First, Meta will stop offering 'Lookalike Audiences' for housing (or credit or employment) ads. This tool allows advertisers to upload a list of existing contacts, and then find other users that Meta's software classifies as similar to those people. It has been directly linked to discrimination, since AI systems designed to find people similar to existing customers are likely to select based on race, gender, sexuality etc., or proxies for those characteristics.

Second, Meta will develop a 'Variance Reduction System' (VRS) to make targeted audiences more demographically representative of 'eligible audiences' who meet the ad's targeting criteria. For example, if an ad targets people in Berlin aged 18-30 who like the music of Nicolas Jaar, and this eligible audience is 60% male, but the automatically-generated audience is 80% male, the VRS will reduce this proportion.

What difference will this make?

As Virginia Eubanks argues, if technologies aren't built with clear intentions to redress historic discrimination, they will intensify it. Making it harder for advertisers to target people demographically similar to existing customers, and trying to make audiences more representative, at least represent intentional steps in that direction. The settlement's terms also suggest the housing department wants to actively oversee and enforce it.

Nonetheless, important questions remain. First, obviously, how will these systems work, and will they actually be effective? Without far more transparency from Meta about their design and outputs, that's very uncertain – as adtech expert Shoshana Wodinsky has pointed out, Meta is essentially asking us to trust that it's replacing a black-box system with another, less discriminatory black-box system. How much Meta will actually invest in designing these new tools also remains uncertain.

More fundamentally, their scope and technical functioning raise major questions. For now, the VRS will only target sex and race – ignoring other legally-protected characteristics that are extremely relevant to housing and employment discrimination, like disability and sexuality. There's also no indication it will consider intersectional discrimination – so an audience for a job ad which is perfectly representative according to sex and race could still completely exclude LGBTQ+ people, or Black women.

Even pursuing representativeness of gender and race is fraught with difficulty. Most Facebook users provide gender identity information, but – unsurprisingly given its history of discrimination controversies – it doesn't explicitly collect data on race. Instead, Meta will rely on a project it launched last year to predict race using surnames, locations and survey data, in order to identify population-level racial disparities. Essentially, therefore, the VRS requires intensified profiling. Moreover, these profiles necessarily use extremely simplistic categories, and arguably reflect a logic of essentialism. Race and gender are understood as fixed, unchanging personal attributes which can be reliably and uniformly predicted, rather than labels imposed on people in particular social contexts.

For example, Facebook allows users to choose from over 50 gender identity descriptors on their public-facing profiles, but its underlying databases reclassify users according to binary gender. Given the extra complexity that representing a diverse range of gender identities would introduce, it seems likely the VRS will (primarily) focus on proportionately representing men and women, overlooking systemic discrimination against non-binary people. Similarly, racial disparities will be assessed using four categories: Hispanic/Latinx, white, Black, and other. This hardly represents the complexities and context-dependence of racial identities. Even if the VRS ensures representativeness of those four categories, many forms of racial discrimination could slip through unnoticed.

Taking the 'eligible audience' as the baseline also leaves a gaping loophole for continuing discrimination, since it can still be selected using proxy criteria which correlate with protected characteristics. If targeted attributes produce an eligible audience mostly comprising white men, the VRS will take these as the 'correct' demographics.

Finally, this settlement only affects Facebook and Instagram ads. A huge part of Meta's business is display advertising on websites, which has just as much potential for bias. Other major adtech companies, notably Google and Amazon, are unaffected (Google's YouTube offers a targeting tool similar to Lookalike Audiences). Overall, then, these measures will at best put a small dent in a massive edifice of discriminatory surveillance advertising.

Meanwhile, in the EU…

From a European perspective, the most obvious limitation of these measures is that they don't apply in Europe at all: they're only happening in the US. This makes sense not only because Meta was forced to introduce them by a US lawsuit, but also because the US is its most valuable market. Developing the VRS will likely be resource-intensive; it would be exponentially more so to adapt it for countries with very different histories and contexts of racial discrimination. Any attempt to redress systemic housing discrimination in Europe would have to consider discrimination against Roma people, migrants from former colonies, intra-European migrants – to name just a few broadly-applicable categories, not to mention the particular groups facing discrimination in individual member states. It would be surprising if Meta invested in doing this voluntarily.

But of course, Europe has its own anti-discrimination laws – and European regulators are generally considered particularly active in legislating and pursuing enforcement against 'big tech'. So why hasn't Meta been asked to take such steps here?

For one thing, it faces few legal risks. As Sandra Wachter's pioneering work on algorithmic discrimination shows, European legal responses are inadequate. Under discrimination law, unless advertisers explicitly target based on protected characteristics, claimants could only allege indirect discrimination, requiring them to show they were unequally impacted relative to a more privileged but otherwise comparable group. This would likely require statistical evidence demonstrating that, for example, women in the 'eligible audience' were less likely to see an ad than men.

The chances of individuals suing Meta because they didn't see an ad, and succeeding, are minimal. Civil society groups suing on behalf of affected user groups might have a better chance, but only if they could access information about how ads are targeted – which, since Meta doesn't make this information available in its Ad Library, and has made aggressive legal threats against third parties unofficially scraping data, is a tall order. The DSA's provisions on data access for researchers could make this easier, but it remains to be seen how much they will empower civil society more broadly. The best option would be for digital services and equality regulators to use existing oversight powers to demand information and take action on systemic discrimination, like the US housing department, but there's little sign of that so far.

The European Parliament debated banning surveillance advertising altogether in the DSA, but rejected it; the final text merely prohibits targeting ads using 'sensitive' data like race, ethnicity and political beliefs. This is clearly intended to acknowledge the discrimination issues discussed here, but its value is – at the very most – wholly symbolic. The GDPR already significantly restricted processing of these data categories, but as Wachter has shown, such bans are totally inadequate to prevent discrimination in practice. They overlook the centrality of predictions and proxy data in surveillance advertising, and the occurrence of discrimination at the population rather than individual level.

For example, Meta doesn't need to process data showing that particular individuals are gay to discriminate based on sexuality. If its algorithms can predict that, say, people who frequently go to gay bars and like watching *Queer Eye* are 25% less likely to be desirable job candidates, and factor that into its ad targeting, this is enough to produce large-scale discrimination.

By just saying advertisers can't explicitly target using sensitive data, the DSA shows a very simplistic understanding of algorithmic discrimination. Gender also isn't qualified as a sensitive characteristic, even though gender discrimination in employment is supposedly an EU policy priority.

Specifically for job ads, the AI Act could offer more effective regulation, since AI systems which affect people's access to employment are classified as high-risk, entailing additional oversight and compliance obligations. Nevertheless, it's not an adequate solution. Commentators have noted many reasons to doubt that the Act will be effectively enforced, given its reliance on self-assessment by companies and privatised technical standards. Even if the high-risk AI requirements are stringently enforced, they represent a very incomplete understanding of the types of algorithmic discrimination discussed here. In particular, bias is only mentioned in relation to data quality. However, as Chun's book illustrates, machine learning can produce highly discriminatory results precisely because it is giving accurate predictions based on an accurate and representative dataset, if that dataset records a reality marked by systemic discrimination.

Meaningfully addressing systemic discrimination in ad targeting – as in many other contexts – therefore requires much more than 'unbiased' AI. Systems must be designed from the start to proactively redress existing social inequalities. Where this isn't possible – as might be the case in targeted advertising, given the technical difficulties discussed above – it might require that they aren't used at all. In this respect, the US settlement is far from being a sufficient response, but at least it represents a serious attempt to understand and proactively address pervasive algorithmic discrimination. European regulators are far behind.

ABOUT THE AUTHOR

Rachel Griffin

Rachel Griffin is a postdoctoral researcher at the Compliant & Accountable Systems Group, RC Trust, Duisburg-Essen University. Her research focuses on EU AI regulation.Rachel Griffin is a postdoctoral researcher at the Compliant & Accountable Systems Group, RC Trust, Duisburg-Essen University. Her research focuses on EU AI regulation.

SUGGESTED CITATION

Griffin, Rachel: Tackling Discrimination in Targeted Advertising: US regulators take very small steps in the right direction – but where is the EU?, Verfassungsblog, 23.06.2022, doi: 10.17176/20220623-153440-0.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20220623-153440-0.