

Should Recommender Systems Be Exempt From the Post-Tracking Age?

Originally published at <https://www.unite.ai/should-recommender-systems-be-exempt-from-the-post-tracking-age/>



As first-party data gathering becomes the new lodestar for marketers and data brokers, the increased attention on ‘closed’ data-gathering systems risks to drag one of machine learning’s most fervent research sectors down into controversy and greater regulation.

Actions taken by FAANG players and FOSS producers in the next 12-18 months are [set to close down](#) the culture of cross-domain tracking that engulfed user analytics systems over the past twenty years, and culminated in the Cambridge Analytica scandals and, subsequently, irresistible popular demand for increased online privacy.

Whether or not the implementation lives up to the ideal, and regardless of the extent to which more generalized tracking systems (such as Google’s FLOC and Apple’s SKAdNetwork) can assuage consumer ire and satisfy advertisers, this new wave of concern for user privacy applies only to cross-domain data extraction in a ‘public’ context, and not to closed or proprietary consumer environments, and the bespoke recommender systems that power engagement there.

Rich Data In Walled Gardens

Platforms such as Netflix, Disney+, HBO Max, Roku, and the Amazon ecostructure (including Prime Video and product recommendations), which utilize custom-built machine learning recommendation systems, are among the content services now proliferating and retrenching as the streaming industry [balkanizes](#).

As third-party data-gathering recedes, the advantage these larger streaming players retain in terms of fine-grained access to customer usage data seems likely to inspire envy and imitation, and a renewed emphasis on first-party frameworks as a way of clawing back hyper-personalized targeting from the more generalized new analytics systems.

If this happens, it isn't likely to be as democratic or meritocratic as prior criteria for entry, because the biggest advantage will fall to providers with the most extensive network of first-party platforms; with enough development resources to provide secure local authentication systems; and which are able to manage, analyze and monetize high volume data locally.

This would focus public scrutiny on the privacy aspects of 'closed' recommender systems in a way that they have largely been able to avoid until now, because, prior to this point, they have been exceptional cases, and enjoyed exceptional privileges, operating in a context where the end user has explicitly opted in to aggressive data-gathering practices that are not generally permitted in open networks.

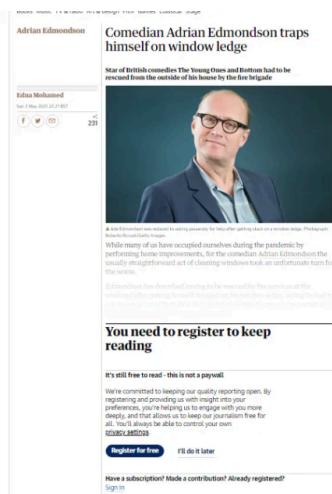
A Wider Return To Hermetic First-Party Environments

An increased emphasis on first-party data seems likely to bring a return to the domain-specific authentication systems that preceded the popularity of third-party methods provided by Google (OAuth 2.0), Facebook and Twitter, as well as other popular bolt-on social platforms such as Disqus.

Ten years ago, the widespread adoption of these third-party authentication platforms solved many security issues for domains with limited development resources, but also made it more difficult to obtain the same granularity of actionable user data that a dedicated and local first-party authentication and monitoring system allows. At the time, it didn't matter that much, because cross-domain tracking could bridge that data gap.

The Login As Solution To An Existential Crisis

Now, the advantage lies in making sure a user is logged in, even if there are no explicit mechanisms to monetize them. One example of this is the growing number of media outlets that are requiring a login to view content, even where there is no paywall in place. For instance, The Guardian is currently experimenting with login requirements for article views that come from Google searches:



Screenshot of a 'login wall' for a Guardian article view that has come from a Google search. This cannot be captured in web archive snapshots, since the restriction is generated either by referrer headers or IP-based systems that reveal Google as the originator of the click.

Restrictions of this type can be difficult to ascertain for an individual viewer, since they may vary across geolocations or other circumstances. For example, the above Guardian article is not restricted in any way when navigated to from within the Guardian website (even if the reader is not logged in), or [when accessed directly](#). Requiring a login from a Google referral is a cheap method of generating a demand-driven increase in membership without alienating 'pre-captured' readers.

Though there have always been data-gathering [advantages](#) in this type of first-party engagement (i.e. a 'local' login), the fall of cross-domain tracking is likely to elevate the practice from 'advantageous' to an existential [necessity](#) in order to avoid the sparser marketing data streams of FLOC and SKAdNetwork.

The Impetus Towards First-party Data Gathering

The evidence of a first-party data 'gold rush' is thick on the ground. According to the view of one industry insider at Forbes, the fall of third-party cookies will lead to new opportunities for companies to [curate and sell second-party data](#), where they have enough first-party infrastructure to effectively become data brokers in their own right.

Analysis elsewhere also forecasts that retailers (who invest heavily in machine learning recommender systems) will [become the new 'media moguls'](#).

In a blog post, monetization platform Setupad exemplifies the intention of the advertising industry to not accede to federated, data-limited systems such as FLOC, [stating](#) that '*behavioral targeting is the answer to future success for advertisers*', and that first-party capture is the absolute prerequisite for this.

Behavioral targeting is [what caused](#) the current tectonic shift in consumer privacy in the first place; and it's what the marketing and professional influencer industries want to win back – by proxy, by stealth or by any other means, no matter that it may eventually drag the recommender system research sector down into the mire with it.

The First-party 'Club'

Besides the requirement for costly infrastructure, as well as security and development resources, another factor indicates why only larger concerns are likely to prosper in the age of first-party data-gathering systems: a company will need compelling market capture in order to coerce consumers back into the local login systems that they were glad to abandon a decade ago.

This is a risky move, even for a major player, and the memory of Digg's [demise](#) in 2010 still haunts the SEO and marketing world. The more compelling a company's market capture, the less damaging this move will be, with more powerful companies able to weather troughs and adapt better to the first-party ecosystem than smaller concerns.

Effects On Recommender System Research

As this situation evolves, it may threaten the relative ‘free pass’ that regulatory oversight has granted to machine learning recommender system research from companies such as Google, Amazon and Netflix.

To an extent, the EU’s [new proposals](#) for AI legislation anticipates greater scrutiny of recommender systems in any case. Though it’s unclear whether the draft’s provision against ‘*subliminal techniques beyond a person’s consciousness in order to materially distort a person’s behaviour*’ will apply to recommender systems, it’s [anticipated](#) that advertisers and recommender system researchers will lobby for exceptional treatment.

But it may be difficult to make a case for ring-fencing recommender system research in the event that the ‘walled garden’ approach becomes the new industry standard, and the leisurely academic pastures that have hosted this sector of machine learning research become a high-volume hotbed for massively commercialized first-party behavioral research development.

Major investment in first-party data workflows may be the only hope for re-creating the same kind of highly effective ‘psychic’ ads and political propaganda that characterized the Cambridge Analytica era; but for the regulators, it may appear that the death of the third party cookie simply moved ‘disreputable’ practices off the streets and into closed premises. If the outward effect of those activities arouse public anger all over again, that may prove scant sanctuary.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG’s Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More