ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

The Centre for Data Ethics and Innovation calls for evidence on online targeting and bias in algorithmic decision making
14th June 2019

Introduction

This response to the Centre for Data Ethics and Innovation calls for evidence on online targeting and bias in algorithmic decision making is submitted on behalf of the Royal Statistical Society (RSS). It has been prepared by the RSS Section on Data Science, in consultation with its network of practising data scientists. It builds on previous RSS submissions on related topics [1-3] and emphasises the practical implications of what is known about the impacts of online targeting. Although much has been written on the subject of online targeting, public perception is skewed by the disproportionate coverage of unusual cases (such as the Cambridge Analytica scandal), while the less dramatic everyday realities of widespread online targeting remain poorly understood. A key aspect of the ethics debate on algorithmic bias is the focus on explainability as a form of transparency designed to mitigate bias, but harm can be introduced by many kinds of bias at any stage of a prediction process, and the debate - especially when it comes to regulation - needs to be widened.

Online targeting

1. What evidence is there about the harms and benefits of online targeting?

Online targeted advertising has a strong influence on behaviour. For example, Lewis and Reiley [4] reported an effect size of 5% (increase in purchases) from a randomised control trial of online targeting advertising to 1.6 Million *Yahoo* users. Online targeting benefits advertisers and retailers in the form of profit, and - it could be argued - consumers by better connecting them to products and services they may need or want.

One of the most widely recognised ways in which online targeting can cause harm is by affecting the outcome of democratic process through means outside election rules. A *Facebook* study [5] reported that a single day of advertising increased voting participation by 0.39%. This implies that targeted political advertising – so called 'dark adverts' – could have a real impact on political outcomes. The RSS believes that democracy should not be done in the dark, and therefore we would want to see transparency of political advertising on platforms.

To progress to a better understanding of the real and potential benefits and harms of online targeting and work towards effective governance, we need to move beyond thinking about single organisations of one kind or another to a whole-system perspective on online targeting.

2. How do organisations carry out online targeting in practice?

Targeted advertising is used to maximise the value of ad impressions. The aim is to show ads only to the audience that is likely to respond to them. Most organisations use third party software called Demand Side Platforms (DSPs) to bid for advertising impressions on an exchange. There are a number of DSP tool providers, e.g. *Google*'s DoubleClick (see [6] for more). Data is collected using a 'pixel' (often a piece of javascript) integrated into a website that 'cookies' a user, in order that they can be identified in the future on an advertising exchange.

It is also possible to 'onboard' a user (assign them a cookie ID) using personally identifiable information such as their email address. However, even a user who has not been 'cookied' by an organisation can still be targeted using, for example, *Facebook* demographic targeting or targeting based on the search terms entered by users. This relies on *Facebook* already having cookied that user. The most prominent providers of cookie-based targeting tools are *Google* (Google Display Network) and *Facebook* (Audience Insights).

The same advertisers that use DSPs to interact with ad exchanges may also provide real estate on their own site for other advertisers to bid for. This requires Supply Side Platform (SSP) tools to communicate with the ad exchanges that display advertising space available on the site. Companies such as *Google* collect the largest amount of information at an individual level (as many people sign in to *Google* to use their apps), but the accuracy and/or bias in the inferences *Google* makes at an individual level has been questioned [8]. The current use of DSPs, SSPs and ad exchanges means that sites providing display advertising space and those advertising can be blind to each other's content. This can lead to inappropriate or fraudulent adverts being placed on sites, or adverts being placed on inappropriate sites, or next to inappropriate content (see for example the 'YouTube adpocalypse').

Companies may also integrate with an exchange to sell their proprietary data for use in online targeting. *Salesforce* Third Party data marketplace [7] lists over 150 suppliers of third party data. For example, *DataXpand* claim: 'We have premium audiences based on exclusive relationships with leading publishers throughout the rapidly growing regions of Latin America, Europe ... we have more than 200 Million Unique Users in our data segments and growing every month!'.

3. Should online targeting be regulated, and if so, how should this be done in a way that maximises the benefits and minimises the risks targeting presents?

In the EU an organisation must appoint a Data Protection Officer (DPO) if it is engaged in activities that "require regular and systematic monitoring of data subjects on a large scale" (Article 37.1.b of the GDPR). This includes online targeting, but only if it is "systematic" and on a "large scale" - for example, a company buying and selling data for online targeting. An organisation that simply uses a DSP to perform its targeting would not have to appoint a DPO, and would likely use *ad hoc* procedures to monitor online targeting practice instead.

The GDPR has not been enforced on *Facebook*, and yet *Facebook* tracks internet users - even if they are not *Facebook* users - via the widespread use of 'like' buttons. 'Like' buttons are social sharing buttons installed on many internet pages to allow easy sharing of content. Via their use *Facebook* gains data on users, who may never have given consent. This requires investigation in the public interest.

Bias in algorithmic decision making

1. *The use of algorithmic tools*

In the financial services industry, algorithms for decision-making are in use for credit scoring, customer recruitment, anti-money laundering, insider trading, and fraud detection. Algorithms in the financial services industry are mostly developed by banks, insurers, and technology firms (both startups and established companies), and are mostly sold by systems integrators (e.g. PwC, Accenture). They are being taken up slowly and cautiously, although there is a feeling in the financial services industry that as soon as one big bank adopts an AI system or application, others will follow.

The key ethical consideration relevant to algorithmic bias is that the system has been built well, and therefore will act in accordance with an organisation's values and ethics. Most ethical breaches are likely to arise from bugs or lack of expertise rather than through design. We believe that there has been an overemphasis on 'explainability' at the expense of predictive accuracy. An unbiased and highly explainable system that has poor accuracy is just as capable of creating an ethical breach as one with bias and lack of explainability. And as both *Google* [9] and DCMS with GDS [10] pointed out in their written evidence to the UK Parliamentary 2017 Inquiry into Algorithms in Decision Making, bias exists in many non-algorithmic decision making processes, and the application of algorithms to these processes can reveal and reduce bias.

Best practice of development and use of algorithms for decision making includes the involvement of a senior scientist who has the ability to understand and test algorithms robustly. The RSS Data Science Section is developing best practice guidelines for testing software algorithms for both bias and accuracy in consultation with its network of practicing data scientists.

2. Bias identification and mitigation

The only way to prove algorithms are biased is to perform experimentation and analysis. Any argument based on explainability, the capability of an algorithm to produce a justification for its decisions, will fail. This is because of hidden correlations which allow latent and implicit variables to create bias against protected groups. Explainability can lead to a misleading sense of confidence that an algorithm is not biased. Bias can occur anywhere. Every part of the process, from the data used to train the algorithm to the design of the algorithm to the way humans respond to the algorithm output, should be tested thoroughly. These two key papers on bias in predictive policing [9-10] provide examples of how to identify bias.

To mitigate bias, the organisation dealing with the customer should make sure the software they use is not susceptible to unwanted bias, either by direct testing or by working with the software vendor. For direct testing, the standard tools of statistics are sufficient but what is lacking is statistical expertise, particularly in senior management.

Algorithms are designed to discriminate, but the nature of bias is not well defined. There is a potential role for the use of counterfactuals in defining and mitigating algorithmic bias. However, there are unlikely to be blanket solutions. For example, oversight, and therefore governance, of algorithms that use the outputs from other algorithms as their inputs and that could generate damaging feedback loops is a challenge.

Data access is key to detecting and mitigating bias, and we re-emphasise the points below, originally made in the RSS submission to the UK Parliamentary 2017 Inquiry into Algorithms in Decision Making [3]:

Detecting and mitigating issues of bias: the importance of data access

2.14. One of the biggest issues surrounding new machine learning algorithms is the data that they are trained on. The strengths and weaknesses of the input data are, therefore, hugely important, and should be considered as well as the inherent logic or formulation of using an automated or analytical system to address a given problem. Olhede and Rodrigues [http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01012.x/full] write that "Even if we can identify the variables fed into an algorithm, data which reflect poor sampling design, unconscious bias, or which contain irrelevant correlations will have repercussions for the computed output: the algorithm can only work with the data it has."

2.15. Having more open data and data standards, of which the Open Data Institute is one prominent advocate, is important for establishing the quality of inputs. Other mechanisms for audit, for example by sharing data for research in 'safe haven' settings, also need to be supported, so that private data pertaining to important outcomes can be investigated when it is in the public interest to do so [http://blogs.lse.ac.uk/mediapolicyproject/2016/02/10/accountable-algorithms-a-provocation/]. There may also be important scope for more advanced research in these areas, to compare the level of bias in decisions to a counterfactual (the quality of decisions that would have been taken if the algorithmic technology were not in place). Auditing on a post-hoc basis is in either case, not trivial and may at times be technically impossible, therefore those who deploy algorithms in society should also assist by considering fairness issues from the outset – they may be supported on this by accessing training, in addition to 'fairness' principles. Decision-makers can also of course assist transparency for researchers, by publishing the evidence that supports their decisions.

2.16. The importance of exploring and explaining data and algorithms as they are applied is that it should improve the ability to detect failures. For example, self-driving cars, if they are to be successful, will require advanced and complex input from machine learning. The level to which algorithms should be explained should differ depending on the implications of their use, and their possible consequences. We believe that there is a great deal of potential for algorithms to be used for good across society, and that adjustment in the prevailing laws and standards should not shut down innovation which can bring lots of societal gain.

References

[1]Reviewing the data ethics landscape
[2]Algorithms in the justice system
[3]The use of algorithms in decision making: RSS evidence to the House of Commons Science and Technology Select Committee inquiry
[4]http://www.davidreiley.com/papers/OnlineAdsOfflineSales.pdf
[5]https://www.nature.com/articles/nature11421#abstract
[6]https://www.g2.com/categories/demand-side-platform-dsp
[7]https://konsole.zendesk.com/hc/en-us/articles/217592967-Third-Party-Data-Marketplace
[8]https://doi.org/10.1515/popets-2015-0007