

# Outreach on Google search engine and Data ethics

## **Introduction:**

According to data from August 2022 [1], google has 92 % Market share world wide in search engine and about 5.6 billion estimated searches are done per day in google search engine[2]. YouTube in other hand is the most popular video streaming platform and apparently the second most popular search engine in the world with 3 billion searches a month. Only with these two platforms owned by Alphabet Inc., they own a huge amount of data from their users. Both of these platforms are free for a user to use. And as the popular saying goes, “if you are not paying for the product then you are the product”. In contest of google search and YouTube our data is the price we are paying to use those so to say free platforms. Google Ad revenue in 2021 was 61.85 billion dollar, which accounts to about 58 % of Alphabet’s revenues [3]. This has been made possible with huge amount of data and higher predictive AI models.

When we use internet and use different platforms we get different sorts of advertisements and some of them may grabs our attention in thinking that the microphone of our device might have been listening to what we say. We would think that, I was talking about this product and here now I get and advertisement to this one. This might not be the case that they are listing to our conversation and targeting us with the ads but, it definitely means that the AI models that are behinds those targeted ads are very good at predicting our behaviors with greater accuracy and these targeted ads are working as intended. And this kind of accurate prediction is only possible because of big amount of data that companies like Google are collecting from their users.

Targeted advertisement in itself could be beneficial for both the user and the product seller, as a company will get it’s buyer and a buyer can get the product easily that they might have been looking for. But the effect that these big companies with such huge amount of data and good predictive models can have in a society as a whole, is unprecedented and arises the question on ethics of collecting and using those data from their users.

Data ethics concerns about the proper handling of data, in order to protect and advance a responsible and better engagement. With big tech giants grabbing all the data they could get from a user, it surely uncovers a new ethical issues related to data.

### **Data collected by Google:**

Google offers a service called Google takeout, which let's you download the data they have collected from you. I requested a takeout and got 25 different folder each with each compressed file size of average 5 GB. And it was collected from 50 different products offered from google. And significant amount of data were collected from google search request and the link clicked along with the YouTube search data, and other data on liked commented, viewed, subscribed videos and so on. Google stores location, google knows everything you have ever searched and deleted. Google also creates a profile from your activities online, which they use for advertisement, and this advertisement profile of you can be seen on [google.com/settings/ads/](https://google.com/settings/ads/).

Google collect and link the data to your identity. Some of the data collected are as follows:

Location, search history, browsing history, contact information, device id, user id, financial information, physical address, IP address, email address, audio data, photos, customer support, other user content, crash data, performance data, Name, phone number and so on.

Google collects data from other websites, which are not owned by google.

It has been claimed that about 76% of websites now contain hidden Google trackers and they continuously soak up our personal information every time we visit those sites [9].

### **Data collection, Personalized search results and Targeted marketing:**

Due to profiling and personalized search result, users are only exposed to the information that has been filtered to fit their need, and this effect is well known as “filter bubble” [4.] In this way google and YouTube search results only shows the results that the search engines thinks the user wants to see. This limitation of information with algorithmic filtering causes a user to be less likely to find information that contradicts their profile. This also limits a users to know a different views on that topic [5]. With such huge number of users on google search and YouTube search, it becomes crucial to be aware that how relying on google for information search can affect the society.

Research has shown that by displaying search results that are favorable to one candidate, the search engine manipulation effect can even affect the results of elections [6]. This kind of power that a company has on the political decisions of a country, brings out the question of ever having a fair election in any democratic country. As fair election is the founding base of a democratic country and this kind of effect definitely undermines our democracy.

Recent events with Cambridge analytics Scandal and USA 2016 election also shows us, how powerful can a data driven camping can be. Whistle blower Christopher Wylie claimed that Cambridge Analytics used Facebook data and survey and created a political messaging weapon, that allegedly changed the result of us election 2016 in favor of their client. Although the data was from Facebook and other sources, we can't deny that, google with such huge amount of users' data, could not do something similar.

Similar to Facebook, Google has a business strategy based on monitoring. As much of your information as possible will be collected in order to fulfill the company's declared goal of "organizing the world's information." That data serves as the foundation for a number of undeniably helpful services, which in turn fuel the advertising that accounts for the vast majority of Google's income.

Google does not just collect data and add algorithmic filter to the search result but also sells our data to third party for advertisement purposes. Such collected data are sold to a highest bidder and then comes the targeted advertisement in place. Targeted ad might seems fine at first glance but

from the ethical point of view it is very unsettling and disturbing. The targeted ads can manipulate users and can be used to get people to consume or to behave in a certain way.

Going further deeper, if multiple characteristics of a group's preferences and circumstances are understood, they may be used to develop rewards that either support or dissuade a particular activity. For example, knowing that group A has a preference A1 (e.g. Homosexuality should be punished) and a majority of the same group has a condition B (e.g. being undecided about which party to vote for). Given this situation one can develop a targeted ads that can target the vulnerable group of people and create psychological condition for a group to behave in a specific way (e.g. if one votes for party C one can have rules where homosexuality would be punished).

A study done on search engine manipulation effect (SEME) showed that, biased search results can readily cause swings in uncertain voters' beliefs and voting preferences of 20% or more – and even up to 80% in some demographic groupings[6]. If we look at United States presidential election, fewer than 7.6 percent has determined the winner of half of the presidential elections, and only 4.45 percent determined the winner of the 2020 election[12]. The difference that is required to win the presidential election is well below the shift that can be bought with SEME. As google search results are biased because they produce personalized results to users query, we should be worrying about fair democratic elections.

Price discrimination is one of the side effects of such targeted marketing as well. It is a technique where a seller sells similar items or services at different prices to different buyers. With targeted ads it makes easier to lure those customers who pay higher price for the same product which can be bought in much lower cost. Although it might be super profitable technique in economical point of views but it is highly unethical practice. Consider a scenario where a particular user needs to buy a camera, with their financial information the model can predict how much they can pay for the product and can sell the product on higher price. When a person is in a condition where they can pay more for the product doesn't mean they should pay more.

In addition to this targeted advertisement may contribute to aggravating per-existing discriminatory system. McMillam cotton describes how African American women may be drawn to online higher education through targeted advertisement and persuaded to take out student loans with a higher chance of bankruptcy and negative depreciation. The Harvard Business Review found that people in wealthier regions respond more enthusiastically to online sales. Another way that technology specifically facilitates inequality is this model, where people with more incomes have access to discounted costs.[10]

For an agency or a person with right set of skills and resources, a person's past and present becomes completely transparent, as we are generating and recording big amount of data in Cyber world.

Collecting huge amount of data from a user may be problematic in such case as server can be breached and data from millions of users can be leaked. This leads to many Cyber crimes affecting those users who knowingly or unknowingly shared their personal data. In this sense already collecting data should be a question in concern.

Although it has been and will be claimed that the data they are collecting will always be anonymized, but this doesn't solve the issue here. Re-identification techniques can easily de-anonymize the data [11].

## **Why do people still use Google?**

First of all popularity of a google is huge such that now a day “Googling” has become synonym for searching something on the internet. Secondly, people are not aware that their data has been collected and been used for model training. Even though people know about it they can’t see the prolonged potential harm caused by this process to stop using google. Even though the data that we are generating may not seem valuable in isolation, it is very valuable when it is added to a model for training. Another reason for google being popular is that it is free of cost in a sense of currency that they pay directly. In reality we are paying with our data, to use google services.

Google search engine is a default search engine in every android phones and IOS devices, and Android and IOS mobile operating system has almost 99% of market share in world [7].

People sticks with the default setting unless it is really necessary, and they have not been properly informed on that need.

In many case people willingly give some of their privacy in the name of security or general good to a larger population, this is however not the primary reason they provide and create personal data. It has frequently been claimed that “convenience” is a key element in users’ willingness to grant access their personal data [8].

Another aspect that make people act fine with collection of data is a people thinking “I have no thing to hide”. This delusive expression rose widely after 9/11 attack especially in USA.

People felt that their public welfare is far more important than personal integrity [8].

As google is collecting and modeling it’s machine learning models and Artificial Intelligence, this benefits often goes mutual. With targeted ads and, improved usability with location tracking, it seems fine at first glance. But we should now ask the question, How much data collection is too much, when will be the targeted advertisement becomes intentional manipulation, is this much data collection necessary for an improvements in usability, and finally should the cost of “free services” we are using equivalent to the cost of the data we are generating.

## **What can we do?**

With 100s of services and most of them free of cost, Google is now everywhere. Although it might be quite difficult to avoid google, it is not impossible. There are alternatives to many of the services given by google.

DuckDuckGo(DDG) is one of the better alternatives to google search engine. It was founded by Gabriel Weinberg and was lunched in Feb. 29, 2008. From 2010 DuckDuckGo adapted the idea of Privacy and differentiate itself from other competitors. DuckDuckGo is a default search engine in TOR browser. Google chrome and Moxilla Fire fox also provides extension of the DuckDuckGo search engine. If you are using Google Chrome as browser and DuckDuckGo as search engine, Google can still track you. It is also better to use alternative browser to Google chrome(for e.g Firefox or Tor) and set DuckDuckGo as default search engine. This might not be a silver bullet to the Data vampires like google, that are sucking out all the data from a users, but is one of the steps towards a better internet experience with better privacy.

First thing we can do about this situation is to spread awareness in public. It could start with conversation among family members and friends. This process is slow and may not be very effective. The authorities should be aware of this issues, and should create and change the laws that allows any one to collect and use our data. The top down approach of handling this problem, i.e. rules and restriction on such practices is the better way of handling this problem. For example Profiling is expressly prohibited by the European GDPR (Global Data Privacy Rule), which took effect in April 2018.

### References

- [1] <https://gs.statcounter.com/search-engine-market-share>
- [2] <https://skai.io/monday-morning-metrics-daily-searches-on-google-and-other-google-facts/>
- [3] <https://searchengineland.com/alphabet-q1-microsoft-q3-earnings-search-advertising-383869#:~:text=Zooming%20in%20on%20Google%20search,%2431.9%20billion>
- [4] Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. New York, NY: Penguin
- [5] Burger, V., Hirth, M., Hoûfeld, T., & Tran-Gia, P. (2016). Principles of Information Neutrality and counter measures against biased information. Paper presented at the Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)
- [6]. Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences of the United States of America,
- [7] <https://gs.statcounter.com/os-market-share/mobile/worldwide>
- [8] Richterich, Annika. The big data agenda: Data ethics and critical data studies. University of Westminster Press, 2018.
- [9] <https://www.cnbc.com/2018/01/31/google-facebook-data-privacy-concerns-out-of-control-commentary.html>
- [10] osanagar, "How targeted ads and dynamic pricing can perpetuate bias," Harvard Business Review, 2019.
- [11] Rocher, L., Hendrickx, J.M. & de Montjoye, YA. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10, 3069 (2019).
- [12] <https://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548/>

