6

Facebook

Facebook – it's the world's biggest social network by a huge margin, and most of us are used to using it to share details of our everyday lives with our friends and families. It's no secret now that we're also sharing it with their advertisers, but that hasn't put most of us off using it! So here's a brief rundown of how Facebook has been one of the most successful companies in the world at gathering our data and turning it into profit – and why some think its business practices sometimes overstep the mark.

Recently, Facebook has been causing a stir amongst those interested in online privacy and data protection. The latest accusations are that is has been carrying out unethical psychological research – effectively experimenting on its users without their permission. Critics have said that by attempting to alter people's moods by showing them specific posts with either a positive or negative vibe,

and then measuring their response, several ethical guidelines have been broken.

The truth though, is that Facebook (and the internet at large) is making its own rules as it goes along. Putting 1.25 billion people – that's getting on for one fifth of the world's population, if we pretend for a second that none of the accounts are duplicates – within a mouse click of each other was always going to have far reaching consequences. And with hindsight it was a bit silly to have ever expected it to be manageable within established social and legal boundaries.

Of course those of us who love social media believe the potential benefits far outweigh the hazards. Putting aside how much easier it makes keeping in touch with our friends and family, there's clearly a lot to be learned from studying the data generated during that communication. And gathering data from us is the foundation of Facebook's business model.

Don't forget though - although it now seems to be dipping its toes into psychological experiments, Facebook's main motivation for collecting and analysing our data has always been to sell us adverts.

Advertisers benefit from highly detailed profiles users build up over time as they use the site – meaning their messages can be targeted precisely at "women over 40 who love books" or "men under 25 living in the UK who love football".

The huge and speedy success of Facebook was prompted by its simple interface and, somewhat ironically given how things have developed, emphasis on user privacy. This helped it quickly become more popular than other early social networks such as Myspace and

Bebo. But with hindsight, it's clear to see it was always gunning for bigger targets.

A big difference between Google and Facebook is that Google's information on who we are is often a "best guess" based on what sites we are visiting. From the start, Facebook explicitly asks us who we are, where we live and what we are interested in. Yes, Google eventually started to do the same with Google+, but by then, they were simply playing catch-up. Advertisers clearly value this direct approach – ad revenues at Facebook grew by 129% from 2011 to 2013, compared to 49% at Google during the same period.

Like Google and all of the other big tech firms, buying up smaller firms to make use of their IP and, crucially, the data from their user base, is a core business strategy. Notable acquisitions have included Instagram and Whatsapp, both of which came with existing communities of millions of users to add to Facebook's own. Interestingly, their highest profile recent purchase was the makers of the upcoming Oculus Rift virtual reality headset. They are clearly thinking ahead to a time when we may be looking for more convenient methods than existing screens offer to view our data.

Facebook has always said that the privacy worries this causes are addressed by the fact that all information is shared with our permission and anonymized when sold on for marketing purposes. That hasn't stopped a lot of critics taking issue with their practices though. For example, many say that the privacy settings are too complex or not clearly explained, meaning it is too easy for people to share things they didn't mean to. Facebook have tried to fix this several times over the years – often confusing people who had got used to the way they were!

Another feature which caused concern when it was introduced was facial recognition. When you upload a picture, you might see suggestions for people you could tag on it. This is based on analysis of the picture data, which is compared against pictures of people in your Friends list, and prompted an investigation by EU privacy regulators in 2011.

More recently, changes to the way its users' habits are monitored have caused concerns. Its latest monitoring tools record everything from how long a user "hovers" their cursor over certain parts of the page to what websites they visit outside of Facebook. Last month it announced that this information is being used in their algorithms that determine which adverts to show us.

But, at least it is now possible to delete your data permanently, if you don't want Facebook to have access to it at all, any more. Before 2010 you couldn't even really delete your account – although you could remove your profile, everything was kept on their servers for an unspecified amount of time, for unspecified purposes (it would hardly be any use to target adverts at you, if you no longer had an account with which to view the site.) Outcry when this was discovered prompted Facebook to add the ability to erase yourself completely.

Facebook's data strategy is led by its Data Science team – who have their own page, of course, which you can see here. They regularly post updates on insights they have gleaned from analysing the habits of the millions who browse the site.

Overall I think that a lot of the problems caused by Facebook are a symptom of its enormous success. Regulators and lawmakers have shown themselves to be slow to get to grips with the revolution it

(and other social media) have brought to the way we communicate with each other, day to day. And, as with Google, it seems like there are more than enough people who think the problems are worth putting up with, for the convenience it brings.

7

Amazon

Amazon is a big data giant, which is why I want to look at the company in my second post of my series on how specific organizations use big data.

We all know that Amazon pioneered e-commerce in many ways, but possibly one of its greatest innovations was the personalized recommendation system – which, of course, is built on the big data it gathers from its millions of customer transactions.

Psychologists speak about the power of suggestion – put something that someone might like in front of them and they may well be overcome by a burning desire to buy it – regardless of whether or not it will fulfil any real need.

This is of course how impulse advertising has always worked – but instead of a scattergun approach, Amazon leveraged their customer data and honed its system into a high powered, laser-sighted sniper

rifle. Or at least that is the plan – they don't seem to get it completely right yet. I have had some very strange recommendations from Amazon.

Anyway, their systems are getting better and it looks like what we have seen so far is only the beginning – as I've previously mentioned, Amazon has recently obtained a patent on a system designed to ship goods to us before we have even decided to buy it – predictive despatch – you can read more about that here. This is a strong indicator that their confidence in reliable predictive analytics is increasing.

An important factor to consider when looking at Amazon is how commercial its big data is, compared to those of other companies that deal with data on a comparable scale. Unlike, say, Facebook – which might know an awful lot about which movies you like or who your friends are – the vast majority of Amazon's data on us relates to how we spend hard cash.

And having worked out how to use it to get more money out of our pockets, it is now setting out on a mission to help other global corporations do the same – by making that data, as well as its own tools for analysing it, available to buy.

This means that, as with Google, we have started to see adverts driven by Amazon's platform and based on its data appearing on other sites over the past few years. As noted by MIT Technology Review last year, this makes the company now a head-on competitor to Google – with both online giants fighting for a chunk of marketers' budgets.

However, ad sales is not the only arena in which Amazon is taking on Google – its Amazon Web Services offers cloud-based computing

and big data analysis on an enterprise scale. This allows companies which need to run highly processor-intensive procedures to rent the computing time far more cheaply than setting up their own data processing centres – just like Google's BigQuery.

These services include datawarehousing (Redshift), hosted Hadoop solution (Elastic Map Reduce), S3 – the database service it uses to run its own physical warehousing operations and Glacier, an archival service. Recently added to this list is Kinesis, which is a real-time "stream processing" service designed to aid analysis of high volume, real-time data streams.

Amazon has also incorporated big data analysis into its customer service operations. Its purchase of shoe retailer Zappos is often cited as a key element in this. Since its founding, Zappos had earned a fantastic reputation for its customer service and was often held up as a world leader in this respect. Much of this was due to their sophisticated relationship management systems which made extensive use of their own customer data. These procedures were melded together with Amazon's own, following the 2009 acquisition.

Finally, it is worth mentioning the public data sets that Amazon hosts, and allows analysis of, through Amazon Web Services. Fancy digging around in the data unearthed through the Human Genome Project, NASA's Earth science datasets or US census data? Amazon hosts all of this and much more, and makes it available for anyone to browse for free.

Amazon has grown far beyond its original inception as an online bookshop, and much of this is due to its enthusiastic adoption of big data principles. It looks set to continue breaking new ground in this field, for the foreseeable future.