Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 6, June 2017, pg.294 – 296

BIG DATA, AI, MACHINE LEARNING AND DATA PROTECTION

N. Sashi Prabha[1], P. Swathi[2]
tinkusashi@gmail.com [1], peraliswathi23@gmail.com [2]
CSE Dept, Assistant Professor[12]
VBIT[12]

Abstract: This discussion paper looks at the implications of big data, artificial intelligence (AI) and machine learning for data protection, and explains the ICO's views on these. We start by defining big data, AI and machine learning, and identifying the particular characteristics that differentiate them from more traditional forms of data processing. After recognising the benefits that can flow from big data analytics, we analyse the main implications for data protection. We then look at some of the tools and approaches that can help organisations ensure that their big data processing complies with data protection requirements. We also discuss the argument that data protection, as enacted in current legislation, does not work for big data analytics, and we highlight the increasing role of accountability in relation to the more traditional principle of transparency.

I. Data protection implications

Under the first DPA principle, the processing of personal data must be fair and lawful, and must satisfy one of the conditions listed in Schedule 2 of the DPA (and Schedule 3 if it is sensitive personal data as defined in the DPA). The importance of fairness is preserved in the GDPR: Article 5(1)(a) says personal data must be "processed fairly, lawfully and in a transparent manner in relation to the data subject". By contrast, big data analytics is sometimes characterised as sinister or a threat to privacy or simply 'creepy'. This is because it involves repurposing data in unexpected ways, using complex algorithms, and drawing conclusions about individuals with unexpected and sometimes unwelcome effects34. So a key question for organisations using personal data for big data analytics is whether the processing is fair. Fairness involves several elements.

II. Effects of the processing

How big data is used is an important factor in assessing fairness. Big data analytics may use personal data purely for research purposes, eg to detect general trends and correlations, or it may use personal data to make decisions affecting individuals. Some of those decisions will obviously affect individuals more than others. Displaying a particular advert on the internet to an individual based on their social media 'likes', purchases and browsing history may not be perceived as intrusive or unfair, and may be welcomed if it is timely and relevant to their interests. However, in some circumstances even displaying different advertisements can mean that the users of that service are being profiled in a way that perpetuates discrimination, for example on the basis of race36. Research in the USA suggested that internet searches for "black-identifying" names generated advertisements associated with arrest records far more often than those for "white-identifying" names37. There have also been similar reports of discrimination in the UK, for instance a female doctor was locked out of a gym changing room because the automated security system had profiled her as male due to associating the title 'Dr' with men.

III. Expectations

Fairness is also about expectations; would a particular use of personal data be within the reasonable expectations of the people concerned? An organisation collecting personal data will generally have to provide a privacy notice explaining the purposes for which they need the data, but this may not necessarily explain the detail of how the data will be used. It is still important that organisations consider whether people could reasonably expect their data to be used in the ways that big data analytics facilitates. There is also a difference between a situation where the purpose of the processing is naturally connected with the reason for which people use the service and one where the data is being used for a purpose that is unrelated to the delivery of the service. An example of the former is a retailer using loyalty card data for market research; there would be a reasonable expectation that they would use that data to gain a better understanding of their customers and the market in which they operate. An example of the latter is a social-media company making its data available for market research; when people post on social media, is it reasonable to expect this information could be used for unrelated purposes?

IV. Transparency

The complexity of big data analytics can mean that the processing is opaque to citizens and consumers whose data is being used. It may not be apparent to them their data is being collected (e.g., their mobile phone location), or how it is being processed (e.g., when their search results are filtered based on an algorithm – the so-called "filter bubble" effect56). Similarly, it may be unclear how decisions are being made about them, such as the use of social-media data for credit scoring.

V. Conditions for processing personal data

Under the first DPA principle, the processing of personal data must not only be fair and lawful, but must also satisfy one of the conditions listed in Schedule 2 of the DPA (and Schedule 3 if it is sensitive personal data as defined in the DPA). This applies equally to big data analytics that use personal data. The Schedule 2 conditions that are most likely to be relevant to big data

analytics, particularly in a commercial context, are consent, whether processing is necessary for the performance of a contract, and the legitimate interests of the data controller or other parties.

Consent

If an organisation is relying on people's consent as the condition for processing their personal data, then that consent must be a freely given, specific, and informed indication that they agree to the processing.

