‘Massive data analysis is more an opportunity than a threat for privacy’

Debates and Ethics of Big Data and Data Science

Muhammad Rizwan Khalid, Jose Antonio Lorencio Abril, You Xu

17th March 2023

Abstract: "The explosion of data and knowledge discovery is changing our world. Big data technologies are being transformative in every sphere of life. Of course, there is some price to pay in terms of privacy, but it is clearly worth it anyway and unavoidable. The benefits we get in terms of new customized and adapted services and preferences is a no brainer: we could not live without massive data analysis tailoring services and products for us."

General Context:

Today, tech giants have access to massive customer and business data. For various reasons, institutions conduct massive data analyses that inevitably contain personal data and sensitive information about individuals. Although the analysis is beneficial in academic research and making a positive social impact, it raises a potential challenge to privacy protection since, if used improperly, it may harm and violate privacy rights.

Proposition Criteria:

As technology advances, data production increases and so does the concern for privacy. This is unavoidable. Nevertheless, thanks to efficient and well-studied data analysis and processing techniques, we have the opportunity to maximize the privacy of people, while leveraging the power of data.

Points of Arguments:

- Data analysis has clear economic and societal benefits. It enables organizations and institutions to make better decisions based on data insights rather than rule-based systems that may not capture the changes in data. Data analysis can reveal patterns and knowledge that may otherwise be hidden. Many industrial sectors use big data technologies to inform their decisions, from IT to agriculture and healthcare. For instance, the European Space Agency (ESA) uses satellite images to monitor the deforestation of the Amazon rainforest [4]. The Copernicus Sentinel missions allow the world's tropical forests to be observed every two weeks. These benefits were also highlighted by the **proposition team**.

- Moreover, public safety and personal health can benefit greatly from massive data analysis. Researchers can use this technique to extract clear patterns and results from complex cases in a short time that human analysis may not match. This is crucial in serious public safety incidents like the Boston Marathon bombing that demand quick action. In the realm of healthcare, massive data analysis has the potential to facilitate the development of individualized care plans for patients. Notably, the American Medical Association (AMA) and various affiliated organizations have introduced

initiatives aimed at tailoring cancer treatments to meet the specific needs of individual patients. [7]. The **proposition team** claimed that these benefits outweigh the potential privacy risks involved. We disagree with this claim, but rather believe that both objectives are not mutually exclusive and can be addressed. For example, there NYT published an article about how data can be used for the public good while protecting individual privacy [5].

- However, these economic and societal gains should not come at the expense of users' privacy. There are several techniques that can protect users' information while preserving the utility of data. One common technique is **anonymization**, which removes identifying information from data so that it cannot be linked back to the original source. This makes it harder to trace a data record to an individual, enhancing their privacy. Anonymization can protect people's privacy in public data and even in data breaches. That is why GDPR requires companies to anonymize all personal data when identification is no longer necessary [1][2].

- Moreover, users and organizations are increasingly concerned about the transparency of how their data is handled. Transparency is not only a legal obligation but also an asset for companies, as it can boost their reputation among customers and peers. By being transparent about how they use data, organizations can foster trust with individuals and ensure their privacy. Transparency also enables individuals to exercise their rights and have more control over their data. It also gives regulators a standard to evaluate and hold organizations accountable. Additionally, it can enhance content moderation and algorithmic performance. Slack is an example of a company that follows the principles of transparency in its privacy policy. The policy is clear and simple, and informs individuals about how their data is used [3][8]. These arguments counter the **opposition team**'s claims that (1) many lives are endangered by the lack of privacy care, and (2) that individuals can be identified from anonymized data, because these techniques are constantly improving to protect privacy and security, and because it is in the companies' interest to adopt better practices and transparency.

- Data minimization is another common technique for data management and privacy that is becoming more important over time. It involves identifying which information is actually relevant for the company's objectives and discarding or avoiding the rest. By minimizing the data they collect and store, organizations can protect privacy and lower the risk of data breaches. For example, Slack does not collect users' location data or browsing history. Slack also deletes messages and files that are older than a certain period of time, depending on the user's plan [8]. This way, Slack minimizes the amount of data it collects and stores, and protects its users' privacy. This counters the **opposition team**'s claim about the danger of data breaches, since we have seen how the use of big data analytics reduces the risk of data breaches.

- The **opposition team** argued that the fines are insufficient or disproportionate to the threat posed by data breaches. In the EU, META was fined twice, as the jury mentioned.

However, big players may regard the fines as a cost of doing business. Moreover, some fines may not be enforceable, such as those imposed by a foreign government on Chinese companies that may not comply with privacy rules or face effective enforcement [6]. Therefore, fines are not the only solution and other strategies for accountability and transparency should be adopted to prevent privacy breaches. Besides, these scandals damage the reputation of the companies, so they have an incentive to avoid them and to make data less attractive for attackers by ensuring high privacy standards.

Final Considerations:

In conclusion, although the threat to privacy is a concern for massive data analysis, it can significantly advance social and economic development that offers limitless opportunities for all human beings' future. Balancing between privacy and the scope of analysis is not easy. However, by carefully considering and thoughtfully evaluating the privacy challenge, we appreciate the fundamental efforts that many institutions, organizations, corporations, governments, and individuals make to minimize the threat while not compromising the opportunity to pursue positive social change for all gaining from the analysis.

References:

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, art 4 (5). url: https://gdpr-info.eu/art-4-gdpr/
2. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, art 5 (5). url: https://gdpr-info.eu/art-5-gdpr/
3. Upland. Great Examples of Transparent Data and Privacy Policies Ahead of GDPR Enforcement. url: https://uplandsoftware.com/adestra/resources/blog/great-examples-transparent-data-privacy-policies-ahead-gdpr-enforcement/
4. "Using a data cube to monitor forest loss in the Amazon." ESA, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Using_a_data_cube_to_monitor_forest_loss_in_the_Amazon.
5. Deming, D. (2021). Balancing Privacy With Data Sharing for the Public Good. The New York Times. url: https://www.nytimes.com/2021/02/19/business/privacy-open-data-public.html
6. "Privacy patchwork: Looking back at the 2021 legislative session", International Association of Privacy Professionals (IAPP), https://iapp.org/news/a/privacy-patchwork-looking-back-at-the-2021-legislative-session/
7. F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine*, vol. 372, no. 9, pp. 793–795, Feb. 2015, doi: 10.1056/NEJMp1500523.
8. Slack. Privacy Policy. url: https://slack.com/trust/privacy/privacy-policy