Big Data: Quasi-public Goods Correlating with National Security and Social Stability

Yunlong Zhang[1(✉)], Jiangbo Wang[2], Yue Liu[3], and Wuji Yan[4]

[1] School of Art and Communication, Beijing Normal University Zhuhai Campus, Zhuhai 519000, Guangdong, China
zhangyunlong0129@126.com
[2] Goldsmiths, University of London, London SE14 6NW, UK
[3] Donald P. Bellisario College of Communications, Penn State University, State College, PA 16803, USA
ykl5395@psu.edu
[4] Beijing No.4 High School International Campus, Beijing 100032, China

Abstract. In the Internet era, data is occupying an increasingly important position, and some important data may even affect the country. More and more people demand that big data correlating with national security and social stability be treated as public goods which hold the characteristics of non-excludable and non-rivalrous. However, due to the economic attribute of big data, we proposes that this type of big data is excludable and non-rivalrous, and should be treated as quasi-public goods. To support this point of view, we quoted two examples, DiDi Chuxing and Cambridge analytics, and give solutions respectively. We believes that the government, as a public service provider, should supervises the use of big data rather than the content of big data. It not only protects the private rights of enterprises to data, but also avoids big data leaks that threaten national security.

Keywords: Quasi-Public Goods · National Security · DiDi Chuxing · Cambridge analytics

1 Introduction

Big data is an aggregation of data and information that specific quantities cannot measure. From a technical perspective, the process from generation to the utilization of big data can be roughly divided into four stages, aggregation, collection, retention, and personalization. Besides the general definition of big data, big data also has the characteristics of velocity, volume, veracity, and variety [1]. Big data and its value can be applied through different means to various fields. The exchange and intercommunication of data will bring benefits to both data collectors and data providers, but often such benefits are contradictory. For example, sharing or leak data between countries may threaten national security. Therefore, big data cannot be exchanged and communicated without distinction and premises. Before relevant legislations and policies adapting to the social changes brought by big data, exploring the relationship between big data and

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 4–12, 2023.
https://doi.org/10.2991/978-94-6463-064-0_2

public good can be a meaningful and beneficial discussion for the good of the entire nation and society.

This paper regards big data as Quasi-public goods as theoretical support. Due to different subjects of data collection and application, big data is still exclusive and competitive. Under certain circumstances, big data will transform between sensitive data and public data. Big data does have the same purpose as public goods, and both are hopeful to provide convenience, generate social benefits, and help social development for human beings. But this does not directly indicate that big data is a public good. Therefore, this paper believes that it is more reasonable to treat big data as Quasi-public goods.

When it comes to national security and social stability, big data could be regulated as a quasi-public good. This paper cites three case studies to discuss the direction in which improper use of big data threatens national security and social stability. They are China's Didi Chuxing leaked data for US, Cambridge Analytica's influence on the US election, and ISIS's use of social media to promote terrorism. These three cases all have one thing in common. When a substantial amount of personal data is collected and utilized for malicious and egotistic purposes, it will cause harm to the interests of the country and the order of society. We believe that big data must be included in the discussion of Quasi-public goods to provide possibilities for the formulation and implementation of legislation and policies.

2 The Definition of Quasi-public Good

Paul A Samuelson was the first to define the concept of public goods, that is, "everyone's consumption and use of such goods will not lead to the reduction of others' consumption and use of such goods" [2], thus deduced that public goods have the characteristics of non-excludable and non-rivalrous. Non-excludable means that the ownership of public goods does not belong to anyone. Under the given supply of public goods, all social members can enjoy the same benefits. Even if it is technically possible, certain people will not be excluded from the benefit group. Non-rivalrous means that the consumption of an item does not decrease as the number of users increases, even if the number of users tends to infinity. This definition was recognized and widely cited by the mainstream academic circles after its birth. However, many items are regarded as public goods with only one characteristic of non-excludable or non-rivalrous, whereas the other characteristic is relatively inapparent. Therefore, we cannot identify any item's belonging to private goods and public goods by the absolute definition of "Public good". In order to solve this ambiguity, this paper employs the concept of "Quasi-public goods."

If the absolute public goods defined by Samuelson are pure public goods, then the quasi-public goods are more similar to a kind of non-pure public goods, which is manifested in its incomplete rivalry and incomplete exclusion. Scholar James M. Buchanan and Elinor Ostrom divided quasi-public goods into club goods and common-pool goods. Buchanan demonstrated that if an item contains only one characteristic of non-excludable and non-rivalrous, it can be defined as "Quasi-public goods" [3].

The definition of public goods in the Internet era is no longer limited to construction resources such as highways and railways or natural resources such as grassland and fishing grounds. Some data resources in the virtual environment should also be considered

and treated as public goods or quasi-public goods. In the process of individuals interacting with the internet, new data and information is generated all the time, which can be divided according to the region where the information producer is located or the relevant field of information. However, the users' information on the internet is mostly owned by companies that occupy a monopoly position in this field. The company integrates the collected user data into big data and determines the company's business model and development direction through comprehensive analysis. They can also sell those datasets to other companies in exchange of money or other user data to maximize benefits. When the data provider knows the purpose of the data and agrees to be collected, the data owner can use the data to push more personalized information or advertising to the user. These data are data whose economic attribute occupies a significant position. This kind of behavior with data exchange as its essential feature is a standard business behavior. Because it can bring economic benefits to the data owner, it is exclusive and competitive to other competitors and can be regarded as an enterprise's private goods.

However, in the data collected by enterprises, although economic attributes are the most crucial component of the data, these data may inevitably involve national security, homeland security, financial security, even bio-security. Once the data involves these realms, they must exist as a kind of public goods and is no longer a private good of the enterprise. Our research regards this type of big data that may threaten national security as "quasi-public goods." Because the main attribute of big data is economical, companies can legally use these data to compete in the market, so it is still excludable. Putting such data under national supervision can ensure that the information is not used illegally. In a democratic society, the government should acts as a representative of citizens to supervise such data on behalf of all citizens.

3 Cases Study

3.1 Big Data and National Defense

Didi Chuxing is China's largest taxi-hailing software. The company once occupied 90% of China's market share of the taxi-hailing industry and had more than 400 million registered users. It uses intelligent algorithms to send message to the taxi driver who are near the passenger through the software to match the appropriate driver and complete the order. Out of consideration for passengers' safety, Didi Chuxing will record the ride route of each customer and the conversation between the driver and the passenger through the App. All the data combined helped Didi Chuxing form an enormous database and analyzed the usage habits of passengers through their algorithm, then provided passengers with personalized services and coupons to attract passengers to use Didi Chuxing more frequently. Didi took this approach to further expand the market in China.

On June 30 2021, Didi Chuxing was officially listed on the New York Stock Exchange. Surprisingly, Didi Chuxing appeared to be extremely low-key. Just when everyone was puzzled by Didi's unusual behavior, the Chinese Cyberspace Affairs Commission officials suddenly announced that it would conduct a cybersecurity review of Didi Chuxing. According to the statement issued by the Commission, Didi Chuxing has serious violations of laws and regulations to collect and use personal information. All app stores are required to remove the Didi Chuxing App following the "State Security

Law of the People's Republic of China" and the "Cyber-security Law of the People's Republic of China." Affected by this incident, Didi Chuxing's stock price has fallen by more than 10%. At present, 25 travel Apps affiliated with Didi have been removed.

According to the statement, Didi Chuxing's database that combines the big data of passenger travel routes and the identity information of passengers after the real-name system is sold to American companies or even the intelligence agencies of other countries. Geographic information is a crucial aspect of national security, and its leakage could cause a disaster. Didi Chuxing has a large amount of national geographic information, if this information is combined with the passenger's personal information, other countries can infer the location of China's confidential departments and institutions. It can even allow spies from other countries to instigate rebellion against people working in these departments and agencies. The data of each order is stored in the company's database in the form of metadata. Regardless of whether the real-name system is implemented, the route of each order will be collected, what the government has to do is not to supervise the specialized personal information provided by the real-name system, but the road data generated in the interaction with the environment. As long as this part of the metadata is supervised, it can effectively prevent these data from threatening national security.

National defense is a recognized public good enjoyed by all citizens of a country, which is entirely non-rivalrous and non-excludable. Data related to national security is an essential part of national defense. This part of data is often legally collected by enterprises, profit is its primary purpose, and it is not entirely non-exclusive. The government should supervise how companies use data and allow companies to use big data for profit while ensuring national security. In fact, Chinese security department has paid more attention to foreign-funded companies, such as Apple, to prevent these companies from collecting customer data in China and submitting it to other countries. But now, it seems that the Chinese government should strengthen the supervision of Chinese companies that go to foreign exchanges to list.

The Chinese government has consciously regarded big data that may threaten national security and stability as quasi-public goods. In June 2017, the "Cybersecurity Law of the People's Republic of China" was officially implemented. In February of the second year, Apple announced that it would transfer iCloud services in Mainland China to Guizhou-Cloud BigData, a wholly-owned state-owned enterprise. Apple's massive amounts of user data are uploaded to iCloud servers in the United States every day. If these data are illegally leaked, the intelligence agencies in other countries can monitor the data of Chinese officials who use iCloud services. They can also find confidential facilities through a massive picture library (iCloud also records where the photos were taken). Therefore, the Chinese government adopted relevant laws to regulate the industry and migrated iCloud big data servers to mainland China. We can compare iCloud to a locked warehouse in which all users' data is stored. Apple has the key to open this warehouse. The Chinese government only requires Apple to move the warehouse to China, and the key still belongs to Apple. As shown in Fig. 1, it can be seen in a statement on Apple's official website that Apple has never provided iCould keys to third-party partners. In other words, the Chinese government cannot open this warehouse, but it can store the

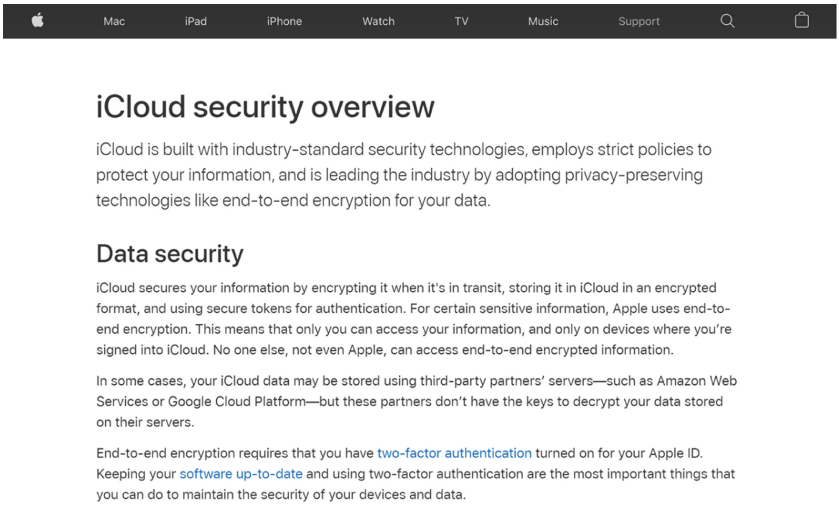

Fig. 1. Apple's iCloud data security statement

data in domestic methods to ensure that the big data will not be illegally provided to others for use.

At present, the government does not have a clear division of which parts of metadata belong to individuals and which parts belong to enterprises. The policies formulated by the government are more inclined to solve the problems that have occurred already and lack foresight. This article puts forward the following three suggestions for solving related problems:

(a) For companies with a market share of more than 50% in the business field, the government should conduct regular supervision of their data technology and data types to ensure the security of the corporate database and prevent companies from collecting data that users do not know through technology.
(b) The government needs to introduce laws and policies to prohibit companies from adding overlord clauses to user agreements, restricting companies from using obscure technical terms in user agreements.
(c) Companies need to establish a data supervision department with professional capabilities to train big data technical talents.

3.2 The Divided Nation: America Under Trump and the Cambridge Analytics

American citizens' political perspectives and openness to an honest and righteous discussion of important issues constitute American democracy and the current political system. As the foundation of American democracy, data and information about citizen's ideology and political opinion must be supervised and protected as "Quasi-Public goods." However, In the United States, Big data's supervision and oversight relied on data aggregator's self-regulation is ineffective and easily corruptible. In this example, the author will interpret an abusive misuse of "Quasi-Public goods." related to politics—the

public's political perspective and participation—will cause political polarization, which may lead to social turmoil and threat to national security.

We can find the presence of Political propaganda and agenda-setting throughout human history. As societies step into the information age, the means of mass communication changed dramatically. The invention of social media exponentially accelerates information dissemination by creating a virtual platform where information transmission can happen between hundreds and thousands of users online. This allows information to transfer and populate on social media like viruses and bacteria among individuals without perception or even awareness [4]. At the same time, users' personal information is tracked and collected into a systematic database, subsequently awaiting analysis to fulfill the data aggregator's objectives. Any individual or organization can forge a tool for massive manipulation by combining social media's efficiency and extension of information dissemination with propaganda's persuasive messaging and audience engagement. Unfortunately, for the past few years, the former president of the United States—Donald J Trump—fall into that description of such an evildoer.

Cambridge Analytics primarily focuses on political communication supported by its psychographic analysis on social media user activities. Essentially, Psychographic analysis can be used to predict a user's personality by five attributes: openness (how open you are to new experiences?), conscientiousness (how much of a perfectionist are you?), extroversion (how sociable are you?), agreeableness (how considerate and cooperative you are?) and neuroticism (are you easily upset?) [5]. Before social media and big data, researchers take a long time collecting and analyzing data from a sophisticated survey to connect the dots between attributes and make a proper prediction. However, the invention of big data systems and worldwide adaption of social media ideally facilitate this analysis process to a point where an individual's personality can be determined in a matter of 200 "likes" from his or her social media account [5].

Although Michal Kosinski, the inventor of this psychographics analysis, relentlessly warned the world about the potential threat of his invention and protected it from falling into the wrong hand. Cambridge analytics somehow managed to obtain his method and utilized it for Trump's election campaign. According to one of Cambridge analytics' former employees, they have used intensive research, data modeling, and performance optimization algorithms to target more than 10,000 different ads to hundreds and thousands of different voters to alter their view on politics and generate conspiracy theories. These propaganda advertisements have been viewed billions of times. We believe that there must be some correlation between Trump's online targeting political advertising and his victory in the 2016 presidential election [6]. According to Pew research center, American political polarization has reached an all-time high under Trump's administration, and he had taken advantage of this phenomenon for his agenda [7]. Trump has continuously utilized social media for targeting political propaganda and posted an enormous amount of prejudicious and inflammatory tweets on racial justice, climate change, law enforcement, and international policy to divide the public opinion into heated discussion and stark disagreement. His irresponsibility and manipulation of citizen's psychographic data subsequently destabilized the society and left America a divided nation [7].

This massive manipulation and subsequent domestic turbulence resulted from Trump and his partner Cambridge Analytica essentially reflected the current legislative void on

Fig. 2. "CCPA Readiness" bar chart comparison of two separate surveys

big data in the United States. Only one of 50 states, California, passed a legal framework—California Consumer Privacy Act (CCPA)—for citizen's data security and digital property rights, and there is no federal bill and regulation regarding the subject matter. Moreover, American firms' altitude and preparation toward CCPA are questionable. According to a survey called "CCPA Readiness" by IAPP (International Association of Privacy Professionals) and company OneTrust, only about 2% of respondents believe that their employing companies are fully prepared to compliance with California Consumer Privacy Act. IAPP and OneTrust conducted two separate questionnaires in April and August 2019. The main question on the survey is: "When do you believe your company can fully comply with the CCPA?" As Fig. 2 shows, the percentage of firms that consider themselves CCPA fully compliant before the bill's effective date was 55%, but strangely, this percentage dropped to 49% in the August 2019 survey [8].

It is reasonable that platforms ought to delete deceitful and propaganda content and ban related accounts quickly. However, Social media platforms generate revenue based on user activities and engagement converted into advertising profit. This business model means that content that can evoke discussion and controversy is more likely to help the company make money. According to section 230 of the US Communications Decency Act, platforms and users rely on self-regulation to identify inappropriate content, and the platforms' legal responsibility for their user's actions is waived [9]. Sometimes people who got banned can open a new account and continue to post hazardous content.

Regarding the legislative flaws in the United States and the commercial behavior of enterprises, we will make feasible suggestions from the perspective of public needs. Cambridge Analytica has obtained citizens-related psychological information through psychological analysis, but it has not been used in the manner expected by citizens. At this level, we believe that the public's participation in politics will be presented to the government in the form of data, so the opinions of the people should also be monitored and used as quasi-public goods. We believe that a more transparent platform should be established for citizens to express their political opinions, and this platform should prohibit the participation of technology companies with prior convictions like Cambridge Analytica. The establishment of this platform needs to be established by technology companies or nonprofit organizations with the highest public trust. The regulator of the platform is the people themselves, not any government or authority, because they may be biased. If it is assumed that the initiator of the establishment of this platform is the government, then the government needs to improve its own credibility as the premise. At

this time, the open government established by Obama has something to learn from [10]. To maintain the impact of transforming personal information into quasi-public goods, in the United States, perhaps technology companies that rely on the trust of the people themselves can better avoid Cambridge Analytica incidents.

4 Conclusion

The development of big data will not stop because of many complex problems. In the future, the problems caused by big data will only increase as the cost of data aggregation will only get lower and lower. How to control big data instead of being controlled by big data is a collective question for every human. We combine the concept of quasi-public goods with a discussion of national security to provide a new direction for the ethical use of big data. As citizens' representatives, the government has the right and obligation to provides citizens visible and invisible protections in the era of big data. The problems brought by the development of big data are the same as its benefits. We hope this article has recognized the challenges and opportunities brought by big data system and constantly reflected on the development of science and technology. We hope citizen's life can peacefully coexist with science and technology. In the future, we will continue to pay attention to this field and conduct research on data-related issues. We hope our elementary discussion and reflection in this paper can help the collective intelligence of the human species develop a comprehensive solution.

References

1. Decuyper, A. (2016). On the research for big data uses for public good purposes. Netcom, (30-3/4), 305–314.
2. Samuelson, P. (1954). The Pure Theory of Public Expenditure. The Review of Economics and Statistics, 36(4), 387-389.
3. Buchanan, J. M. (1965). An economic theory of clubs. Economica, 32(125), 1-14.
4. Quijano, A. M. (2015). Social Media as a Tool in Information Dissemination. Academia. https://www.academia.edu/25078409/Social_Media_as_a_Tool_in_Information_Dissemination
5. Hannes, G. & Krogerus, M. (2017, January 28). The data that turned the world upside down. VICE. https://www.vice.com/en/article/mg9vvn/how-our-likes-helped-trump-win.
6. Paul, L. (2018, March 23). Leaked: Cambridge Analytica's blueprint for Trump victory. The Guardian. https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory.
7. Dimock, M., & Wike, R. (2020, November 13). America is exceptional in the nature of its political divide. Pew Research Center. http://www.pewresearch.org/fact-tank/2020/11/13/america-is-exceptional-in-the-nature-of-its-political-divide.
8. IAPP & OneTrust. (2019, October). CCPA readiness Survey. CCPA Readiness Survey. https://iapp.org/resources/article/ccpa-readiness-survey/.
9. Brannon, V. C., & Library of Congress. (2019). Liability for content hosts: An overview of the Communication Decency Act's section 230. (CRS reports (Library of Congress. Congressional Research Service).)
10. Yan, Y. (2015, July). Research on the National Strategy of Big Data in the United States. China Academic Journal Electronic Publishing House. http://www.cnki.net.