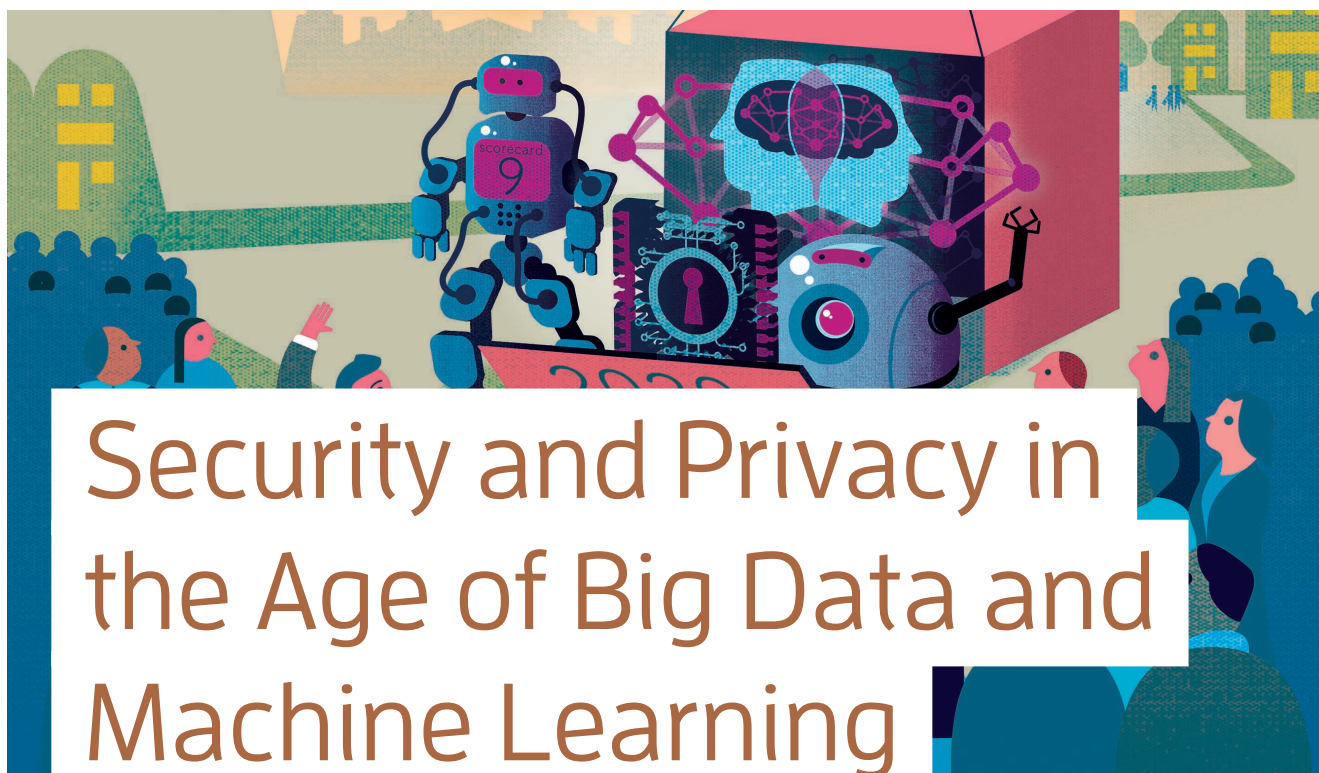




Avi Mendelson, Technion—Israel Institute of Technology

New technologies have the potential to influence society in positive and negative ways. This article explores the possible impact of fast-growing, information-related technologies, such as big data analysis and machine learning, on our privacy and the future of our society and discusses methods to protect against malicious actors. The article ends with a call for action for regulations that need to be made and tools that need to be developed to help to impose them.

The fast development of new technologies always makes a huge impact on society; it can advance civilization, but it may also result in unplanned, destructive results.¹ This article focuses on the potential consequences of the extraordinarily fast development of computer systems, in general, and information technologies, in particular. Special attention is given to the societal implications of using machine learning

and big data. Those technologies have the potential to provide significant contributions to our quality of life; however, they could also be used to cause irreversible damages. Much attention has been devoted to this topic, mainly in the form of privacy protection; new laws were passed and new regulations introduced to protect our rights with respect to the data that are gathered about us. As part of that effort, new measures were defined to prevent the potential harmful usage of that data collection. Unfortunately, it is quite clear that the current efforts are insufficient.

Digital Object Identifier 10.1109/MC.2019.2943137
Date of current version: 22 November 2019

This article focuses on two subjects: the privacy impact and potential misuse of new information technologies and the use of false information to achieve political and commercial goals. It argues that the use of false (fake) information needs special attention and the development of new technologies for its detection and prevention since it has the potential to have a

This part concludes with a discussion on the need for future steps and a call for action.

The second part of the article focuses on the use of false information, such as fake news or modified video/audio, as a means of impacting the outcome of reasoning algorithms, which are based on big data and machine learning. We argue that this type of

- › strive for system security, meaning that only authorized people can access and use it.

THE FAST DEVELOPMENT OF NEW TECHNOLOGIES ALWAYS MAKES A HUGE IMPACT ON SOCIETY; IT CAN ADVANCE CIVILIZATION, BUT IT MAY ALSO RESULT IN UNPLANNED, DESTRUCTIVE RESULTS.

significant negative impact on society. The article does not focus on the use of these technologies for military purposes and “traditional crimes,” such as stealing money, blackmail, and using equipment to perform illegal operations. It mainly focuses on the use of big data and machine learning to change the structure of our society by changing political systems, altering economies, affecting the way people interact with each other, and so on. The first part of the article focuses on privacy-related issues. We look at the new Chinese social-ranking system as the starting point for our discussion and examine how such a system can impact society. Next, we extend the discussion to determine whether such a system can be implemented in the Western world, and if it can, what impact it could have on us. Then, we extend the discussion on some of the recent rules and regulations aiming to prevent the negative impact of big data on privacy.

information misuse is extremely dangerous since it has the potential to change society. We conclude this section with a description of some of the countermeasures that can prevent this potential negative impact and a discussion of what tools must be developed to enable better tracking if a violation of the rules occurs.

PRIVACY-RELATED ISSUES

The Chinese social-ranking system

It has been reported that China is developing a social-ranking system² that will begin to do the following by 2020:

- › track all citizens in the country
- › rank all citizens based on their “social credit”
- › reward and punish citizens according to their social-credit scores

The nature of the ranking algorithm has been kept secret, but it is expected to be based on different types of sensors, including street and surveillance cameras; reports received from banks, schools, and department stores; and so on. It may be impacted by political factors and other unknown causes.

Data gathered as part of this effort could be used for many other purposes. Nicole Kobie³ points out that a “private project” in China already examines the habits and behaviors of more than 400 million customers to profile their family statuses, how much time they spend playing video games, and more. Such a system, even if designed for good reasons (similar to the financial-credit system), violates privacy and has the potential to have a huge impact on society. It will enable a small number of people to gain absolute control over the entire population, manipulate data (if they decide to do so), and make the entire population dependent upon the government. A few cases have already been reported where people were asked to pay to raise their social-credit level.

Can social-credit systems be created in democratic countries?

This article assumes that in Western countries (such as the United States, the United Kingdom, and Germany), there are enough checks and balances to prevent governments from supporting the creation of programs similar to the Chinese social-credit system and any other scheme designed to allow population-behavior control. (Some people might disagree with this point,

but for the sake of the argument the article supposes it.)

Unfortunately, it looks as though programs of a similar nature to the Chinese one already exist or are under development by companies striving to maximize their profit under the umbrella of homeland security and similar efforts. For example, in the article “Germany Edges Toward Chinese-Style Rating of Citizens,”⁴ Cath-erin Schaer points out that in Ger-many and the United States, systems (Schufa and Fair Isaac Corporation, respectively) that are owned by private companies have access to the credit history of millions of customers and that the decisions those corporations are making do not solely depend on that information. The article claims that the systems’ credit rankings also rely on “geo-scoring,” for example, the average socioeconomic rank of a candidate’s neighborhood and friends and information related to her social network. Using that data, banks and other financial institutions estimate the capability of a customer to return a loan. However, the same informa-tion can be used to understand peo-ple’s behavior and the products they may use and for political reasons, that is, guessing the party that a customer will most likely vote for during the next election.

Search engines (such as Google and Bing), social networks (Facebook and LinkedIn), and others gather data related to our privacy that potentially could be used to create a database sim-ilar to the Chinese social-ranking sys-tem. Gerd Gigerenzer, director of the Harding Center for Risk Literacy at the Max Planck Institute for Human Development in Berlin, noted, “If we don’t do anything, then one day a cor-poration or a government institution

will pull all of the information from different data banks together and come up with a social-credit score.”⁴

Countermeasures: privacy

Given that we cannot prevent personal information and private data from being collected (intentionally and unintentionally), lawmakers around the world have begun establishing sets of rules that aim to provide better pro-tection for the way personal and pri-vate data are collected and used. The European Union’s (EU’s) General Data Protection Regulation (GDPR) is a good example of such an attempt. This sec-tion provides a short description of GDPR and compares it to similar sys-tems in the United States.

GDPR. GDPR is a set of rules de-veloped and agreed upon by the EU members. It replaced the Data Protec-tion Act of 1998 and took effect on 25 May 2018. The regulation seeks to pro-protect individuals’ privacy rights and ensure that organizations use personal data appropriately. It sets out the ways

in which privacy rights must be pro-protected and how personal data can and cannot be used. For example, personal data can be defined as names, dates of birth, addresses, phone numbers, email addresses, membership numbers, IP addresses, images and photographs, re-ligions, ethnicities, sexual orientations, and medical information. GDPR also suggests special treatment for criminal and children’s data.

Figure 1 depicts the five stages of the GDPR lifecycle⁵: assess, capture, store, use, and destroy. The stages aim to define the different intermediate states in which data can be held and the oper-ations applicable for each one. From the perspective of this article, special atten-tion must be paid to the destroy stage because it is designed to guarantee that information does not remain in the sys-tem longer than required. It should be noted that, in certain situations, such as data connected with felonies, the destroy rule may be reversed, and infor-mation may need to be accessible for a long time. Hence, the system would

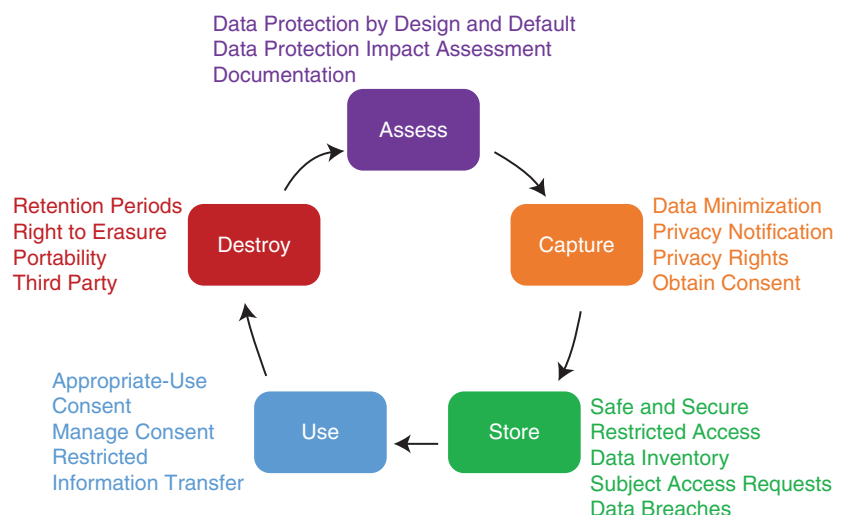


FIGURE 1. The GDPR lifecycle. (Based on “GDPR Overview.”⁵)

need to guarantee that the information was available during the entire required duration.

GDPR also defines a set of rights with respect to stored personal data. That information should include the “right to be forgotten” and must maintain portability, that is, be transparent and easy to access. The user who is the object of the data may ask to apply other restrictions to the kind of processing that is allowed; for example, a person can ask that the data he/she provides will not be used for commercial purposes. GDPR further requires that agencies that keep personal data must report security breaches to a governing body (for example, the Information Commissioner’s Office) and to the individuals whose information was affected, within a reasonably short time period (usually 72 h).

U.S. privacy rules. In the United States, each state enforces different regulations regarding security and privacy, so if personal information is being stored, used, and manipulated, different jurisdictions will treat the situation in varying ways. According to a report by DigitalGuardian,⁶ legislation has been enacted by all 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands to regulate the definition of private data, how it should be stored, who owns it and for how long, and so on. The rules are not identical, but there are many similarities between them.

- › *Data-breach notification laws:* They are similar in every state.
- › *Explicit notification deadlines:* Most states mandate explicit time frames for notifying affected individuals, typically 30–45 days.

- › *Backward compatibility:* This usually expands the current personally identifying information regulations and definitions.
- › *Credit-monitoring requirements:* Regulations in this area are expected to be completed soon.
- › *Private rights of action:* These are expected to be similar to those described in GDPR.

The gap between the U.S. and EU rules for privacy. GDPR requests that sensitive data not be transferred outside the EU (the countries in the European Economic Area). If the transfer of such information is needed (for example, between the United States and Europe), it may be transmitted only after appropriate safeguard measures, which must be agreed upon and signed, are carried out. The current situation is very problematic for international companies, such as Google and Facebook, that are required to separate their U.S.-based data from any that are generated within Europe. So far, companies have paid billions of dollars in fines to the EU for violating the rules.

SECURITY, ADVERSARIAL ATTACKS, AND FAKED DATA

So far, we have focused on the privacy impact from big data and machine learning, but there is growing concern for how “poisoned” data and false information, such as “fake news,” affect the decisions that machines and human beings are making. Such data can be harnessed to achieve commercial and political goals in illegitimate ways. As an example, let’s look at the Chinese ranking system again, but this time, assume that poisoned data are being used.

- › Adversarial attacks can cause the system to make wrong classifications and decisions. That would apply mainly to machine-to-machine interactions via voices and pictures. Under this assumption, some people may be able to negatively impact the scoring of others.
- › When the system gets inputs from many different sources, its security level declines, thus increasing the probability that unauthorized people could affect or even change scores. This is particularly dangerous when private agencies have access to the system.
- › Machine-learning algorithms are statistically based, which makes it difficult to prove their correctness; thus, a malicious change of data can hardly be detected.

This section addresses the impact of three types of false information:

1. changing valid information, such as a student changing grades, through techniques known as fault injections or security attacks
2. adding “noise” to the system so that machines and/or human beings 1) cannot distinguish between the correct and noisy information and 2) make the wrong decision based on the noise instead of the valid information, with examples including adversarial attacks that add noise to pictures to confuse their classification
3. intentionally distributing fake information to confuse people and/or machines.

Although the purpose of each technique is different, they share common ground. As complexity grows and the amount of data reaches a certain point, human intelligence and common sense stop being effective for validating and authenticating the origin and coherency of information. Thus, we depend on “intelligent” machines (that is, machine learning) to enable us to ease the assessment of such data sets. When false information is injected, it compromises the validity of those assessments. The situation requires us to develop tools, methods, regulations, and countermeasures for detecting and preventing the use of faulty data.

Changing valid data

Changes to valid data can be made either by an attacker or the organization that keeps the information. Due to the scope of this article, we are more concerned by opportunities in the data-storage system to modify information and use it against individuals and groups. Such situations can be exploited by criminals and governments. For example, many people oppose the ruling that allows the government to store our biological signatures, such as fingerprints, facial scans, and so on. That information could be used at a later date to frame a person and/or blackmail her. As an individual, it would be very difficult to prove such a fraud, a reality that creates an opportunity for a small number of people to control large populations.

Adversarial attacks

Adversarial attacks proved to be effective in changing the results of classification mechanisms.^{7,8} From our point of view, such attacks can violate our rights since they can hide people's identity and possibly force algorithms

to make wrong decisions. To emphasize the significance of those possible results, consider the potential consequences for a system similar to the

laws lack the mechanisms to handle such situations, Facebook (which holds the rights to the webpage where the footage was posted) has refused to remove

**NOT ENOUGH EFFORT HAS BEEN MADE
TO PROTECT SOCIETY AGAINST THE USE
OF FALSE INFORMATION.**

Chinese social-credit mechanism. An adversary could change the ratings of arbitrary or targeted people and enable individuals to hide their actions.

Fake news and data

Fake news and data are starting to be commonly used by different people and organizations for various reasons. The techniques are made possible by our limited capabilities for verifying the origin and validity of data. The use of false information is extremely dangerous to society since it enables a small group of people, including politicians, to control public opinion and change the way we make decisions. Such actions have already been demonstrated to be very effective for achieving economic goals and controlling people's behavior. Unfortunately, creating fake information is not very difficult; it begins with spreading rumors and extends to creating illegitimate pictures and videos. Interestingly enough, the methods used for adversarial attacks can be employed to create synthetic videos that look similar to the original ones.^{9,10} For example, a doctored video of U.S. Speaker of the House Nancy Pelosi that was distributed online in mid-2019¹¹ was slowed to make her appear drunk. Since current

the doctored video,¹² even after it was proved to be fake.

This article focused on the impact on society of fast-growing, information-related technologies, such as big data analysis and machine learning. It highlighted two significant issues: the need for advanced privacy regulations and enforcement and protecting society against the use of false information, that is, fake news. Currently, a large legislative and technological effort is focused on protecting our privacy. Many countries and U.S. states are busy creating and enforcing rules designed to support the entire lifecycle of private data within systems. The trend is expected to continue, with the belief that a generation of new rules, with better technologies to enforce them, lies in the near future.

Unfortunately, not enough effort has been made to protect society against the use of false information. We believe that such information has the potential to cause major damage by enabling a small group of people to control the way people think, make conclusions, vote, and so on. We need to set up rules and develop sets of algorithms and tools to detect violations and enforce

ABOUT THE AUTHOR

AVI MENDELSON is a professor in the Computer Science and Electrical Engineering Departments at Technion, Israel Institute of Technology. His research interests include computer architectures, distributed systems, hardware security, and accelerators for machine learning. Mendelson received his Ph.D. in computer engineering from the University of Massachusetts, Amherst. He is a Fellow of the IEEE. Contact him at avi.mendelson@technion.ac.il.

regulations. For example, techniques were recently developed to help identify fake news^{13–15} and videos.^{9,16} The author hopes that the growing number of research studies and tools will enable society to better handle this situation, but those developments will not be enough on their own. The next step must involve wide agreements between parties, governments, and elected representatives for the creation of a set of international rules that will define the boundaries of what is allowed and what is not. ■

REFERENCES

1. "Alfred Nobel." Accessed on: Oct. 17, 2019. [Online]. Available: <https://www.britannica.com/biography/Alfred-Nobel>
2. A. Ma, "China has started ranking citizens with a creepy 'social credit' system," *Business Insider*, 2018. [Online]. Available: <https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4>
3. N. Kobie, "The complicated truth about China's social credit system," *Wired*, 2019. [Online]. Available: <https://www.wired.co.uk/article/china-social-credit-system-explained>
4. C. Schaer, "Germany edges toward Chinese-style rating of citizens," *Handelsblatt*, 2019. [Online]. Available: <https://www.handelsblatt.com/today/politics/big-data-vs-big-brother-germany-edges-toward-chinese-style-rating-of-citizens>
5. "GDPR overview." Accessed on: Oct. 17, 2019. [Online]. Available: <http://www.glenariffeoisins.com/wp-content/uploads/2018/05/GDPR-Overview.pptx>
6. "The definitive guide to U.S. state data breach laws," *Digital Guardian*, 2018. Accessed on: Oct. 17, 2019. [Online]. Available: <https://info.digitalguardian.com/rs/768-OQW-145/images/the-definitive-guide-to-us-state-data-breach-laws.pdf>
7. K. Eykholt et al., "Robust physical-world attacks on deep learning visual classification," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1625–1634.
8. N. Srndic and P. Laskov, "Practical evasion of a learning-based classifier: A case study," in *Proc. IEEE Symp. Security and Privacy*, 2014, pp. 197–211.
9. J. Hui, "How deep learning fakes videos (Deepfake) and how to detect it?" 2018. [Online]. Available: https://medium.com/@jonathan_hui/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9
10. L. Jiang, X. Ma, S. Chen, J. Bailey, and Y.-G. Jiang, "Black-box adversarial attacks on video recognition models." 2019. [Online]. Available: [arXiv:1904.05181v2](https://arxiv.org/abs/1904.05181v2), 2019
11. D. Harwell, "Faked Pelosi videos, slowed to make her appear drunk, spread across social media," *Washington Post*, 24 May 2019. [Online]. Available: https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/?no_redirect=on&utm_term=.99465930d3b7
12. J. Waterson, "Facebook refuses to delete fake Pelosi video spread by Trump supporters," *The Guardian*, 24 May 2019. [Online]. Available: <https://www.theguardian.com/technology/2019/may/24/facebook-leaves-fake-nancy-pelosi-video-on-site>
13. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explor. News Lett.*, vol. 19, no. 1, pp. 22–36, 2017.
14. R. Oshikawa, J. Qian, and W. Wang, "A survey on natural language processing for fake news detection." 2018. [Online]. Available: [arXiv:1811.00770v1](https://arxiv.org/abs/1811.00770v1)
15. J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, 2019.
16. S. Li et al., "Stealthy adversarial perturbations against real-time video classification systems," in *Proc. Network and Distributed System Security (NDSS)*, 2019. doi: 10.14722/ndss.2019.23202.