We Need to Think Data Protection Beyond Privacy

Turbo-Digitization after Covid-19 will advance algorithmic social selection and the biopolitical shift of digital capitalism

Rainer Mühlhoff · Technische Universität Berlin · https://RainerMuehlhoff.de

Mar 31, 2020

Photo by Jens Johnsson on Unsplash

The Covid crisis throws some Central European countries into a sudden process of what might be called "turbo-digitization". Many are demanding that the authorities be given access to mobile phone tracking data to slow down the spread of the virus. Schools and universities are hastily making digital learning platforms available after more or less resisting them for a decade. A majority of home office workers are now gaining in-depth experience with video conferencing and cloud services — often through providers with serious privacy issues. All the while Amazon logistics is experiencing a boom like never before.

All these changes involve generating even more data about our professional and private lives and collecting it in central places. Take eLearning platforms as an example: When interacting with these tools, which can also be used to submit homework and take exams, large amounts of data are generated, not only about the grades and performances, but also, for example, about when and how often someone logs in, how active they are in the classroom chat and with whom they communicate in the background via direct messages. This data can — at least theoretically — be used to create detailed profiles of the learning performance and behavior of the vast majority of our students. In countries like the USA, the commercial use of such data has been prevalent for years.

Turbo-digitization and its side effects

Covid-19 is currently leading to a number of political taboos being broken and discourses being shifted. Due to the political reactions to SARS-CoV-2, it is quite possible that the virus will permanently transform our societies. Without scientific evidence of the effectiveness of mobile phone data in containing the virus, the use of such data would be a disproportionate invasion of the right to informational self-determination. Seldom before has there been such an easy majority for legislation that will make this shift possible. And without a comprehensive political debate on data protection and regulation, the widespread collection of behavioral data of particularly vulnerable groups (including data on the learning behavior of students) could produce an irreversible data leak. If this process of rapid innovation and legislation continues at the current pace, the political debate on the risks and side-effects can hardly keep pace. This could then lead to fundamental restrictions of our freedoms with painful long-term consequences.

But why is it so urgent to supplement the digital measures against the virus with a new discussion on data protection and digital fundamental rights? Do we not already have one of the sharpest and most progressive pieces of legislation in the world, the European General Data Protection Regulation (GDPR)? Can we not rely on the GDPR even in this crisis, provided that all measures comply with it?

The point is that the GDPR mainly protects the individual and his or her personal data. However, the main danger posed by the data collection that is now emerging is not the digital exposure of individuals, but the proliferation of algorithmic procedures for population management. Sure, anonymized mass data always carry the risk of re-identification of data subjects; this in itself is already a reason for concern. However, much more serious but less politicized is the use of such data for the training of machine learning models and predictive analyses, which could decide, for example, which population groups will have access to medical services, jobs and educational opportunities, or make predictions about who is in poor health or mentally unstable, who poses a potential security threat (predictive policing), or who has potentially spread the virus according to their location tracking history.

Data protection beyond the EU GDPR

The GDPR does not effectively protect against the use of anonymised data for predictive algorithmic decision making, risk scoring and behaviour-based classification that can be used to treat individuals or groups unequally. Firstly, the GDPR operates with a distinction between anonymous and personal data which is no longer effective today. Secondly, it conceives of data protection and privacy as the right to informational self-determination, i.e. as the right to control the use of one's own personal data. However, algorithmic scoring and decision-making procedures are based on anonymous comparison with the data of many other individuals. By sharing one's own (possibly anonymized) data with a data company, one potentially contributes to an apparatus that harms other individuals and groups. And vice versa, you are potentially harmed yourself by the data many others disclose about themselves (potentially even anonymously) in their daily use of networked services.

The comparatively progressive Article 22 of the EU GDPR does indeed grant certain rights to those affected by a fully automated algorithmic decision. In practice, however, these rights are ineffective because the specific power relations in algorithmic decision-making situations force the subjects to consent to the automated procedure: For example, a precarious person unable to obtain a loan from a traditional bank faces the last option of turning to a predatory payday-lending firm such as ZestFinance to apply for a short-term loan at interest rates of

600–1000% by exposing their social media data. Such a person is, in the end, forced to give their consent. Furthermore, Article 22 does not prohibit the *collection* of the data sets used for algorithmic scoring at all, nor their use in semi-automated decision-making processes. Hence it does not regulate the operator side of predictive analytics by putting reasonable limits to the use of pseudo-anonymized mass data.

Biopolitical turn: From targeted advertising to social selection

The SARS-CoV-2 virus, or the reaction to it in the form of executive orders and hurried legislation, accelerates a "biopolitical turn" of digital capitalism that has already been evident since Cambridge Analytica. This development has so far been discussed primarily in relation to Asian societies, for example in reference to the "social credit scoring system" in China: Data and digital technology are no longer used solely to influence the market behavior of individuals, for example through personally measured advertisements or by calculating individual credit risks on the basis of behavioral data. Rather, data-based algorithms are now used to manage populations. In the biopolitical turn, data-based algorithms increasingly mediate access to the entire social and cultural environment, to jobs and welfare state resources such as social or health services, and to political communication and decision-making — up to the point where they can decide on elections. All this has been evident in Europe and the US during the last few years, but transitioning to this form of data governance particularly in the EU might now be accelerated due to the legislative and infrastructural reactions to the virus.

If (possibly anonymized) behavioral data are collected almost everywhere, the predictive models that are trained with them are able to divide whole populations into risk groups and manage them algorithmically. Data-based algorithms can then organize society into invisible social classes, for example, those who allegedly pose a safety or health risk, those who deserves priority access to scarce medical resources such as respiratory care, those who are allegedly suitable for certain jobs due to their learning behavior at school or university, or those children who are more likely to become victims of domestic violence and should therefore be monitored preemptively by the Child Protective Services.

Collective Data Protection

Networked media are now so widespread that the (possibly anonymous) collection of everyday behavioral data has reached a critical level. So much data is produced that for many business models and use cases the focus is not on the personal data of individuals but on high-resolution yet anonymous mass data. This situation calls for an intensified debate on digital fundamental rights. This debate must take into account that an individual or a group can be treated unequally on the basis of anonymous data that others disclose about themselves, possibly in best faith. This means that data protection is no longer a private matter at everyone's own discretion, as data protection regulation has it. Rather, data protection in the age of predictive analytics is a *collective concern*.

The data that you disclose, for instance by using services such as Gmail, has an impact on others, at least if many people disclose that data. For this effect, it suffices that the data be used in pseudo-anaonymized form by the respective platform company. Algorithms that are suitable for the management of whole populations based on behavioral data are not concerned with names and identities. They get trained on the behavioral data of millions of putatively "normal citizens" who think about themselves as "having nothing to hide". Based on this reference data, which is voluntarily donated by the vast majority, "less normal" or precarious people, allegedly deviant, dangerous or unhealthy individuals can be discriminated against. Hence the societal risk associated with Big Data is not identification or disclosure of personal information, but the *algorithmic selection of societal groups* that are treated differently in terms of access to opportunities, resources and information. Data-driven technology will produce and stabilize social inequalities at local and global scales if its deployment is not regulated by a *collective* understanding of data protection that escapes the liberal paradigm of informational self-responsibility.

We need a debate on digital fundamental rights, now.

The political response to the SARS-CoV2 pandemic provokes a rapid expansion of everyday data collection along with the legitimate purposes for its use. For a political debate on data protection to keep up at this pace, the issue of collective digital fundamental rights should now be on everyone's agenda. The authoritarian rhetoric of a "war-time feeling" and state of emergency all too easily dismisses any responsible discourse on fundamental rights as a lack of solidarity.

But unlike curfews that can be lifted, leaking of data cannot be undone. That's why we should better think twice when it comes to hasty digital transformations. We are at a crucial juncture where we can either take a European path in managing both the innovation potential and the dangers of digital technology in concert with strong, collectively-oriented fundamental rights, or we end up adopting the biopolitical use of digital technology from Asian societies in the course of the coming decades.

After governments in Europe and the US considered mobile phone tracking against Covid in authoritarian states such as Singapore or South Korea as role model, a debate about privacy in this context has erupted in our media. This debate, however, focuses exclusively on individual privacy and neglects anonymous mass data and collective data protection. The profound societal harms that can result from the unregulated collection of anonymized mass data is still astonishingly under-debated. This does not only hold for tracking data, but more generally for our daily use of tool like Gmail, Facebook, Dropbox, digital classrooms, video conferencing, or smart thermometers that is now increasing wold-wide. All these trends show how urgent a discussion about *collective* data protection is. Particularly in times of Covid-induced turbo-digitization, a lack of awareness in the public and among political decision-makers of the possibilities and dangers of anonymous mass data might prove painfully detrimental in the end. We should beware of the easy majority that is available in this moment of crisis to bring forward legislation and infrastructure that will exhibit a profound lack of solidarity in their long-term social effects.

This is an English translation of an article that was first published March 31, 2020, on Netzpolitik.org.