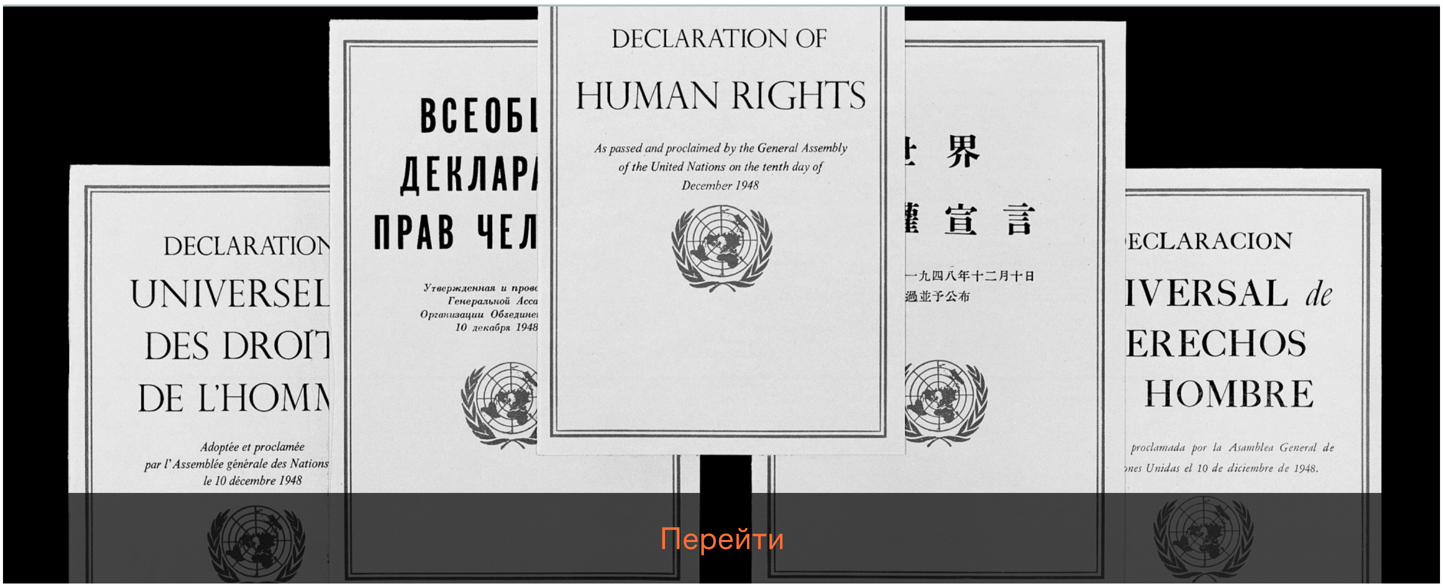

17.05.2023

Human Rights and New Technology in Russia

JOINT SUBMISSION TO THE UN HIGH COMMISSIONER FOR HUMAN RIGHTS

Published: 17.05.2023

Русская версия

OVD-Info is an independent human rights media project that monitors cases of political persecution and violations of fundamental human rights in Russia and provides legal assistance to its victims.

Roskomsvoboda is the first Russian public organization dedicated to the protection of digital rights and digital empowerment.

In this report, we highlight the lack of any standards for the application of new technologies in Russia and the consequent human rights violations. Such violations in Russia are particularly common in the use of facial recognition systems, online censorship technologies, traffic management and filtering equipment, and social media monitoring systems. There are no effective mechanisms for redress and compensation, prevention of further allegations, or for the pre-evaluation of technologies for human rights compliance. Frequently, the public, experts and human rights defenders do not have access to the algorithms of technologies at all. All of the above generates multiple violations of human rights. In the report, we also analyze good practices in human rights standard setting and technology assessment in other countries and suggest how such standards and legal procedures can be established and work.

We recommend to the High Commissioner:

- to draw attention to the Russian situation dedicated to the use of new technologies in the absence of necessary standards that respect human rights;

- to condemn the numerous human rights violations committed by the Russian authorities in the use of such technologies;

- to recommend countries which use mechanisms to assess and analyze the potential human rights impacts of technologies as they are being developed, to facilitate public discussion and debate with representatives of the groups which will be affected or likely to be affected by the technology, in conjunction with with relevant experts;

- to recommend that countries ensure that effective measures are in place to protect and remedy human rights violations caused by the use of new technologies and punish the offenders of such violations, including the developers and users of the technologies.

What examples illustrate in the best manner the relationship between technical standards for new and emerging digital technologies and human rights?

Facial recognition systems

In Russia, the law enforcement authorities use facial recognition systems not only to suppress and investigate crimes, but also to prosecute for political reasons and to search for mobilization purposes.

In particular, Moscow has a surveillance system in courtyards, entrances, and crowded places. Shopping centers and nightclubs in Moscow are also forced to transmit video surveillance recordings to a data processing center to which law enforcement authorities have access. The data from the city's CCTV cameras was actively used to identify participants in the 2021 peaceful assemblies and to prosecute them after they participated in the rallies.

In addition, these technologies are actively used in the Russian transport system, including the metro. In the Moscow public transport, there is a «Sphere» system with facial recognition, processing and storage of the information obtained. A similar system is in operation in St. Petersburg.

From 2022, the facial recognition system began to be used for preventive detentions on public holidays or important public events when, in the authorities' opinion, protest activity is more likely to occur. In most cases in the Moscow metro, those who had been previously prosecuted for participation in protests or for «discrediting» the Armed Forces were detained. People who had not been prosecuted but had participated in protests and had been recognized by video surveillance, or who had registered on opposition platforms' websites, or taken part in democracy forums, etc., were also sometimes detained. Generally, such detentions were associated with being taken to a department of the Russian Ministry of Internal Affairs and a preventive conversation. In total, between 2021 and 2022, OVD-Info recorded at least 595 detentions using the facial recognition system.

Since September 2022, the «Sphere» system has also been used to trace men for mobilization purposes, and in 2023 there was information that a special database with comprehensive data on conscripts will be developed and linked to a facial recognition system, including at transport and border checkpoints.

According to official information for February 2023, 8,998 cameras transmit signals to the «Sphere» system. These cameras are not only located in vehicles, but are also used in mobile units. At the beginning of February 2023, the authorities reported that 7,713 people had been detained using the «Sphere» system. On February 15, 2023, M. Romashin, a deputy head of the Moscow metro security service, said in his speech at a conference on transport security that, if necessary, information from surveillance cameras overlays cellular tracking data on the use of bank cards, fare cards, which allows tracking and detaining a person. However, he also said that «it would be ideal to detain people before they commit a crime or offense, as soon as they ought to break the law.»

Some detainees whose identity was established using the «Sphere»» system challenged their detention and the use of facial recognition systems against them, arguing that their personal data should be removed from the «Sphere». The Moscow courts have currently heard five such cases and rejected the claims of applicants in each of them. Eight more cases with similar claims are pending before the courts. According to a court decision in one of the cases, the detention was initiated due to the fact that the person had previously been held the administrative liability for

«discrediting» the Russian Armed Forces and was wanted for a «rally» on the Russian flag day on 22 August 2022.
In another case, which is still pending before the court, a response was received from the division of the Ministry of Internal Affairs department that the detention and delivery on 23 September 2022 (the day of «referenda» in the Russian-occupied territories of Ukraine) was due to the fact that according to Sphere system, this person was a participant in mass events. However, in most cases, the authorities do not acknowledge that they used the Sphere facial recognition system in connection with protest activity and refer to the law on operative-search activities and the secrecy of information on such activities.

Technologies used for online censorship

In 2012, the first centralized Unified Blocking Register appeared in Russia, which is maintained by the communications and media supervisory body Roskomnadzor (Federal Service for Supervision of Communications, Information Technology and Mass Media). The register is based on Federal Law No. 139-FZ, which amended the Federal Law «On Information, Information Technologies and the Protection of Information» (the «Information Act»), the Federal Law «On Communications», and the Federal Law «On Protecting Children from Information Harmful to Their Health and Development».

The basis for including websites and Internet pages in the Unified Blocking Register and the general procedure for action by Roskomnadzor, hosting providers, and telecommunications operators was originally laid down in Article 15.1 of the Information Act. Over the following 10 years, the list of grounds and procedures for blocking was expanded and specified, other rules appeared in the Information Act aimed at restricting access to specific types of information and online services (e.g., messengers, email services, VPN services), separate blocking registers were created for some of them (nowadays several ones exist, all maintained by Roskomnadzor).

In general, restriction of access to websites or online services under the Information Act is carried out as follows:
an authorized state body or a court decides on restricting access to information or an online service and further sends the decision to Roskomnadzor, then Roskomnadzor enters information in the relevant blocking register (domain name of the website or page index, website network address, short description of the type of banned information, details of the decision that is the basis for blocking, date when the decision was received by Roskomnadzor). Roskomnadzor submits information on websites, web pages, and IP addresses, which need to be blocked, to Russian communications operators automatically (data is uploaded twice a day). Russian telecommunications operators are obliged by law to restrict subscribers' access to websites and online services that receive a data upload from Roskomnadzor (a failure to comply with this obligation will result in a heavy administrative fine).

There are no mandatory technological standards for Internet blocking in Russia. Telecom operators were allowed to block an entire site in order to block one of its pages if the telecom operator did not have the technological capability to perform the blocking on a point-by-point basis. Penalties for non-compliance with the duty to block, the lack of any standards and recommendations of Roskomnadzor have led to the fact that limiting access to information in Russia is often unnecessary and affects content that has not been recognized as illegal in Russia. At the same time, the protection of the right to disseminate legitimate information under extensive blocking in Russia is difficult and almost does not lead to a positive outcome for the distributor of legitimate content, as the established court practice is unwavering in its position of justification of almost any blocking and the courts are not receptive to arguments about the unreasonableness of excessive restrictions on the freedom to disseminate information. Two cases are illustrative in this respect: the case of Vladimir Kharitonov, director of the Internet Publishers Association, and the blocking of the Telegram messenger.

Vladimir Kharitonov, a director of the Association of Internet Publishers, has been trying to challenge in the Russian courts the decision on limiting access to his website as of 2012 — the blocking was aimed at a site with drug propaganda and was done by IP address, so it affected V. Kharitonov's entire website, located on the same IP address. The courts found such a blocking lawful, while the European Court of Human Rights (hereinafter — the ECtHR) found it to be contrary to Article 10 of the European Convention on Human Rights.

In 2018, the situation repeated on a larger scale. Roskomnadzor authorized the blocking of the Telegram messenger by IP addresses (simultaneously by the court order and the request of the Prosecutor General of the Russian Federation). This led to access restrictions and disruptions in the operation of numerous online services and websites, and the owners of some such services and sites tried to challenge the «blocking at the same time» as illegal and recover damages, but their claims were refused by the Russian courts (for example, the owner of the VPN service TgVPN, whose case is currently pending before the ECtHR, did so). Roskomnadzor announced the lifting of restrictions on access to Telegram messenger resources on 18 June 2020.

DPI and CTCS

During the 2-year blocking of Telegram messenger in Russia, the so-called «Sovereign Internet Act» was adopted.

This law introduced mandatory internet traffic routing rules and moved traffic routing under the control of Roskomnadzor. The purpose of the law is to protect the Russian segment of the Internet from external threats. DPI equipment and

CTCS are the technological tools used to implement this law. The law requires telecom operators to install and use on their networks special traffic filtering equipment (DPI) and centralized traffic control systems (CTCS).

DPI stands for Deep Packet Inspection. It is an analysis of the contents of packets (the short blocks by which information is transmitted over the Internet). It is not possible to read the content of encrypted packets, but DPI metadata tells us what sites a user is browsing and, in most cases, what communication protocol he uses. Initially DPI operators studied Internet traffic to set priorities in data transmission. For example, voice messages are more latency-sensitive and need to be prioritized in the traffic stream. Later, solutions appeared to monetize (analyze traffic to sell information and replace ads on sites not using HTTPS with «their own») using DPI equipment. But in the last couple of years we have seen the use of DPI technology also for blocking online content.

In September 2019, DPI equipment from various manufacturers was tested in the Ural Federal District for the purposes of enforcing the law as part of a pilot project by the Roskomnadzor-controlled state enterprise FSUE «Main Radio Frequency Center».

Any documentation with technical standards of this equipment was not disclosed, the testing took place without the participation of civil society institutions, the mechanisms and ways of such participation are not provided.

In any case, not all telecom operators have DPI equipment, and not everyone who possesses it uses it for blocking. In any case, in the end, DPI equipment is not available to all telecom operators, and not everyone who has it uses it.

TSPU *(In Russian «ТСПУ»)* stands for technical means of countering threats. In other words, it is a new hardware and software system that allows restricting access to information whose distribution is banned in Russia. Roskomnadzor has required telecom operators to install TSPA since September 2020. It is doing so within the framework of the law on sovereign internet. The oversight body can now centrally manage the Russian segment of the Internet by filtering traffic using DPI.

TSPU allows not only to block, but to slow down various services and data protocols, as well as to perform «shut down» at the local level. The case of the social network Twitter is an example of «slowing down» with the help of TCSPs.

On 10 March 2021, Roskomnadzor announced Twitter being placed into the list of external threats due to slow removal of material containing banned information, and its decision to slow down Twitter on 100% of mobile devices and 50% of stationary devices. However, there are no explicit grounds for slowdown set by the law. One of the normative acts adopted in pursuance of the law on sovereign Internet states that the «threat to security of Internet functioning on the territory of the Russian Federation» is subject to restriction. Measures to address it include «changing the routes of telecommunications messages» and «changing the configuration of means of communication». No specific technical information on how exactly the supervisory authority implements this very slowdown is publicly available.

Experts made only assumptions and technologies to slow down traffic, so a group of experts, which included a Roscomsvoboda expert, conducted a study and published a report on April 06, 2021. Apart from the technological analysis, the report noted that the Twitter case showed that the mechanism of online censorship in Russia was changing its nature from a decentralized one (telecom operators blocked content based on information from Roskomnadzor) to a centralized one (Roskomnadzor has the ability to unilaterally impose desired restrictions thanks to the mandatory use of TSPU equipment). The researchers also noted that, unlike blocking, in which access to content is blocked, throttling (traffic slowdown) is aimed at reducing service quality, making it almost impossible for users to distinguish between imposed/intentional slowdown and a number of nuances, such as high server load or network congestion.

In light of this, it is important to note that initially content blocking was centralized only, the information about restriction of access to websites and grounds were publicly available from open sources of Roskomnadzor. However, slowdown and blocking under the Sovereign Internet Law through the TSPU are not recorded or displayed in any public registers.

The situation with the slowdown of the social network Twitter in Russia clearly demonstrates that society is completely excluded from the process of shaping technical standards for traffic analysis and filtering technologies such as DPI. Against this backdrop, the use of such dual-use technologies for online censorship makes censorship more invisible to citizens. Meanwhile, citizens who used Twitter attempted to challenge the slowing of traffic to the service through the courts as part of RoskomSvoboda's «Battle for Twitter» public campaign. The court found that Twitter users had no right to challenge the slowdown and refused to accept the claim. The refusal is currently being appealed to the court of cassation.

RoskomSvoboda, in cooperation with the Open Observatory of Network Interference (OONI), has produced a research report documenting cases of internet censorship in Russia over the past year (from January 2022 to February 2023). This report contains a technical analysis of OONI's Internet accessibility and a large legal analysis of all legal provisions enacted since the start of the «special military operation». OONI experts revealed that resources blocked due to military censorship were not all reflected in the state blocking registry.

To conclude, the use of DPI and TCSPA for censorship has not so much exacerbated the issue of respect for human rights in the use of such technologies (including through public participation in standards development) in Russia, as it has drawn attention to the existence of such a problem. Neither at the time of the purely commercial use of these technologies by telecommunications operators, much less their use in implementation of the Sovereign Internet Law by the authorities, was the very issue of human rights (freedom of information and privacy) taken into account, and attempts by public institutions and individual actors to combat human rights abuses due to these technologies are blocked by the courts.

Social media monitoring

In February 2023, Roskomnadzor launched Oculus, a system of automatic search for banned content. According to the technical documentation, the system analyzes images and videos, chats and messenger channels, URLs and other data for banned information in real time and classifies content according to types of banned information.

Also subordinate to Roskomnadzor, the Main Radio Frequency Center is conducting internal tests of the Vepr information bombs search and deactivation system. According to the terms of reference, the system is to counter «dissemination of socially significant information under the guise of reliable messages, which poses a threat to the life and (or) health of citizens, property, a threat of mass disruption of public order and (or) public safety». The system will also aim to detect authors spreading anonymous messages on the Internet, to analyze content on certain topics-triggers, as well as the dissemination of identical messages that, for example, appear on the news websites or Telegram channels and are picked up by some opinion leaders. Experts assume that both Oculus and Vepr systems will be used jointly by Roskomnadzor to replace the «manual» detection of prohibited information on the territory of the Russian Federation. We suggest that these systems could be used both to make decisions to block content and to punish individuals for disseminating information on the Internet. This would reduce the possibility of challenging and appealing restrictions and penalties, as Russian courts do not question the veracity of decisions made in automated mode.

Therefore, both systems are developed at the request of public bodies, the RF public procurement system does not involve analysis of technology in a human rights context and the participation of civil society institutions in the development to set standards in this context.

What standard-setting processes are particularly important to protect and promote human rights in the framework of new and emerging digital technologies?

One of the problems of regulating new technologies is the Collingridge dilemma: «Trying to control technologies is difficult... because in the early stages, when they can be controlled, not enough is known about the social consequences; but by the time those consequences become obvious, control becomes costly and slow». Another serious problem is the so-called pace problem: «technology is changing exponentially, but social, economic and legal systems are changing gradually». The regulatory system works slowly and cannot keep up with new technologies and therefore inhibits development.

New technologies always carry risks to human rights, so it is important to assess technologies at an early stage of development from a human rights perspective and to perceive them as «the interaction between technological capacity and social values».

Support for civil society organizations can dramatically increase awareness of human rights and contribute to their protection. The OHCHR Business and Human Rights in Technology Project is a relevant process that should be developed further as a mechanism for the transparent participation of civil society in protecting human rights in the field of technology.

National authorities should play a central role in assessing legislation and identifying any gaps. Although it is impossible to predict the development of new technologies, there are already established processes that can reduce the risks of human rights violations. One of these is the «human rights by design» approach, based on already well-known and widely used «privacy by design» methodologies. Developers must fulfill the key obligations and requirements of this approach when creating technology.

This way, significant harm can easily be avoided at the early stages. In December 2020, the Ad Hoc Committee on Artificial Intelligence proposed nine principles and priorities to help establish such a framework. Although their focus is on artificial intelligence, these principles can easily be applied to all new technologies in all sectors.

There are also a number of other mechanisms that can help to protect human rights. One is human rights due diligence, which is rapidly becoming a requirement in many jurisdictions due to the introduction of specific regulations at the national level. It is crucial for each organization to conduct a human rights impact assessment to make sure that the design, development and deployment of the technology do not violate human rights.

Countries should consider adopting laws or other binding regulations that reinforce the need to fulfill human rights in the development of new technologies. Such legislation should, inter alia, require that the initial stages of development of any potentially human rights-affecting technologies be subject to assessment by independent experts. Alongside technical experts, human rights experts should be invited to assess and evaluate such technologies and their development processes. If the technology fails such an assessment, the development process should be reconsidered and, if it fails to meet the assessment criteria, it should be halted.

Another solution is the creation of regulatory sandboxes. Regulatory sandboxes are not exactly a new concept. They are already actively used in fintech around the world. Regulatory sandboxes are regulatory tools that allow businesses to test and experiment with the new and innovative products, services or businesses under the supervision of a regulator for a limited period of time.

Human rights risks can also be reduced by more conservative methods, such as certification and audit. If used properly, rather than as a «tick box», these processes can protect human rights.

Prior to the launch of the technology, public discussions and debates should also be organized with representatives of those population groups who will be affected or may be affected by the technology to a larger extent, along with relevant experts. Moreover, if human rights are found to be affected by the technology in the process of use, developers and others responsible should be held accountable and use should be suspended. Victims of such violations should be able to apply to court for protection and restitution of their rights, as well as receive compensation for damages. If the violation of a particular person’s human rights cannot be resolved without abandoning the use of such technology, or if it becomes obvious that the rights of others are being or may be violated, the use of such technology must be terminated, even if its customer and/or developer is the state.

What are the general obstacles to effectively integrating human rights considerations into the processes of establishing technical standards for new and emerging digital technologies?

The proliferation of new and emerging technologies around the world has greatly expanded the state's toolbox of repression and social control, resulting in a gradual deterioration of human rights protection in this area over the past two decades.

If the customer and user of the technology is the state, many of the human rights restrictions arising from the use of such technology are usually justified by national security and public order purposes. In countries where there is virtually no opportunity for expression of independent opinion and participation in public debate, such as in Russia, it seems impossible to have an independent expert assessment of the necessity and proportionality of human rights restrictions caused by the use of technology.

One of the common obstacles to the effective integration of human rights is the adoption of laws and other regulatory acts that restrict human rights through the use of technology. Thus, states are not fighting for human rights but against human rights defenders and independent media. New laws facilitate the legalized collection of citizens' data.The first strategy in Russia is total restrictions, attempts to control the digital environment through the Yarovaya laws, control of the Runet, blocking of independent media websites, and the dissemination of their materials.

If the customer and user of the technology is a non-state company, and there is no legal obligation to conduct ex ante assessments of the technology’s human rights compatibility, they will have little incentive to include civil society representatives in discussions, especially in countries where there is no legal obligation and accountability for human rights abuses in the case of private companies.

Moreover, if there is no legal regulation of a particular technology, as well as a general requirement for prior assessment of the human rights compliance of the technology, there will be no effective means of protection for those affected by the use of the technology. This is the situation in Russia with respect to the use of facial recognition against protesters, as the use of this technology is not regulated in any way. This makes it extremely difficult to challenge and prove such use in the courts.

How accessible are standard-setting processes for new and emerging digital technologies to a wide range of stakeholders, in particular civil society organizations and human rights experts? What indicators measure «access» in this context?

In Russia, standard-setting processes for technology are inaccessible because human rights, including the ability to express independent opinions, to exercise the right to freedom of association, and to participate in public life, are generally severely restricted. The existence of «access» could be measured by the openness of information about the development of such technology, the possibility for civil society to participate in public discussions in relation to the development and use of the technology, to participate in expert assessments of the compliance of the technology with human rights, and the existence of effective remedies for those affected by the use of the technology. All these criteria are not met in Russia.

What are good practices, mechanisms and models for effectively integrating human rights considerations into technical standard-setting processes? Are there particular challenges in their implementation or adoption? What additional measures should be developed and implemented?

For example, the use of remote sensing techniques to document evidence of crimes against humanity in Myanmar in 2017 has shown good results, precisely because of the human rights standards of such technologies and their use without «personalisation». In this case, a comprehensive analysis was applied, including satellite imagery and analysis of the soil layer. The result was the destruction of villages (with ashes in their places) and the emergence of refugee camps.

In litigation, good practice in the use of AI can be observed in Australia, where there is a guide that clearly delineates the limits of such use to prevent human rights abuses. Although it seems impossible and excessive to replace an AI judge in every case, AI can be actively used for technical work as well as for administrative offense cases.

The human right to information, which derives from fundamental conventional rights, is comprehensively violated due to discriminatory and selective practices of social networks and internet services. A good human rights mechanism here is the formation of a system of legal restrictions, similar to antitrust regulations, which would limit the right to unjustified filtering of information by aggregators and social networks.

What are the duties and responsibilities of standard-setting organizations and stakeholders in effectively incorporating human rights considerations into technical standard-setting processes for new and emerging digital technologies?

Consideration of the right to privacy

A prime example of this rights violation is AI applications for facial recognition. This type of technology is already used by police in some countries and risks being used by authoritarian regimes to suppress political dissidents and minorities.

A 2016 study found that half of American adults are already in police facial recognition databases across the country. Because of concerns about privacy and misuse, several major US cities have banned the technology. California, New Hampshire and Oregon have passed laws banning the use of facial recognition technology with police body cameras. Following the Black Lives Matter protests in the US in 2020, IBM, Amazon and Microsoft limited or suspended sales of their facial recognition products.

In Europe, the GDPR prohibits the processing of biometric data to uniquely identify an individual, health data, data about an individual’s sexual life or sexual orientation, and the processing of data revealing racial or ethnic origin.
EU officials initially considered a complete ban on facial recognition in public places, but instead gave member states the option to introduce a ban after strong opposition from some members. In addition, in April 2021, the European Commission proposed new rules and actions to develop robust AI, in which facial recognition is considered a high-risk application and only allowed in certain cases.

In Russia, facial recognition systems are not controlled or regulated in any way, the public and human rights defenders have no access to the development and operation of such technologies, there is no formalized agency responsible for such processes, nor is there controlling legislation. Courts and authorities refuse to provide details of the identification and entry of persons into databases, as well as algorithms for exiting from there.

Consideration of the right to non-discrimination

Cognitive biases often affect non-automated data review systems, and automated AI systems can theoretically help correct or compensate for some of these biases. However, AI systems may also be intentionally or unintentionally biased. Concerns about discrimination arise when particular variables in algorithms implicitly serve as proxies for protected or undisclosed information, such as race, sexual orientation, gender or age. An algorithm may result in users discriminating against a group that correlates with the proxy variable in question.

In the context of crime prevention and predictive policing, discriminatory decision support by AI can lead to serious rights violations. Such examples include the use of artificial intelligence systems to support the identification of potential terrorists based on the content they post online. AI training in such applications is based on current police databases, which often reflect and reinforce existing racial and cultural biases that exist within communities. Existing databases may be biased or incomplete. In Russia, such databases have been known to be used for political, ethnic and ideological profiling, setting dangerous precedents for the further development of such databases and systems and the decisions they make.

It is the duty and responsibility of the organizations developing and applying these technologies to assess in detail each stage of the process of developing and applying these technologies, to report on these processes transparently, and to establish mechanisms to stop the use of algorithms.

More to read

The strangling of Crimea
Denis Shedov and Dan Storyev explain how the repression in occupied Crimea was structured and how it evolved