IN-DEPTH ANALYSIS

Requested by the DROI subcommittee

Executive summary of the in-depth analysis

Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights

Authors:
Rasma KASKINA and Angelina CVETKOVSKA (Schuman trainee)
Policy Department for External Relations
Directorate-General for External Policies of the Union

June 2024

EN

Executive summary

Scope and objectives

The aim of this In-depth analysis (IDA) is to address the use of artificial intelligence (AI)-based tools to monitor, influence and suppress opposition or dissent. AI technologies are evolving rapidly, as are their application by an increasing number of countries and the risk of their misuse.

The term 'AI repression' or algorithmic repression does not have a single definition, but can be understood as an intersection of AI applications and repressive actions, including algorithmic surveillance and censorship, automated decision-making and differential technological enforcement. While repression is often associated with authoritarian regime types, it is important to analyse the purposes of techniques, tactics and procedures that could lead to AI-based repression, regardless of the regime type. The purpose of this paper is to provide an overview of AI-based challenges to human rights and to recommend a set of policy options for the European Union (EU) and the European Parliament (EP) to tackle the issue, in the light of the EU's AI Act and in line with standards for the ethical use of AI.

Algorithmic authoritarianism practices that constitute a growing, worrying global trend are illustrated in the IDA through case studies from countries that have frequently appeared in EU statements and reports. These countries impose significant government control over digital channels, underpinned by extensive surveillance systems, and therefore represent 'cases of concern' for human rights and democracy. Another factor underpinning the selection of these case studies was some of these countries' use of unregulated advanced technologies in foreign information manipulation and interference (FIMI) targeting EU Member States, among others.

The IDA also underscores the global nature of the increasing AI-based repression by looking at ways in which Western technology exports have contributed to the deployment of AI-based repression and surveillance tools.

Case studies

China

China's use of AI-based technologies to target the Uighur Muslim minority is one of the most well documented instances of AI-based algorithmic authoritarianism. In Xinjiang, AI surveillance and predictive policing are used to target the Muslim minority under the guise of counter-terrorism. Through the collection of vast data points, authorities can detect potential dissidents before any concrete act is committed. Several data collection initiatives facilitate the surveillance of the Uighur Muslim community, including the Sharp Eyes programme for urban surveillance, and the Integrated Joint Operations Platform, which monitors behaviour considered indicative of potential social instability. The extensive surveillance network limits Uighur citizens' liberty and sense of safety, and leads to self-censorship. Authorities have detained Uighur citizens in re-education camps based on the outcomes of ambiguous AI-based analyses. Investigative journalists, human rights organisations, some Western governments and international organisations have severely criticised the Chinese government for its treatment of Uighur Muslims. Despite the international pressure, including economic sanctions, the surveillance network in Xinjiang persists. The risk of Chinese surveillance technology being adopted by other countries raises concern about a more widespread use of China's methods of social control.

Another example of China's repressive use of AI-based technologies is the country-wide Social Credit System (SCS), which should not be understood as a system to assign a single score for each individual, but rather a whole range of inter-connected policies and technologies. Municipal authorities and even companies implement SCS initiatives, which use social and behavioural data and AI to rank citizens and corporations. The data includes public and private information, such as financial transactions, health records, employment status, compliance with civil duties and any information that citizens contribute voluntarily. This leads to an extensive web of data mining, used to evaluate daily behaviours across different spheres of daily life. The integration process of interlinking data points involves sophisticated algorithms and machine learning models.

The Chinese Communist Party has found a powerful instrument in AI technology to preserve the status quo and suppress dissent in the country. The same technology is also used to target individuals outside China.

Russia

Russia has also increased the use of AI-based tools under the rationale of internal security and stability. The use of such tools has facilitated internet censorship, and repression, including through predictive policing. Urban surveillance has also increased, with AI-driven facial recognition technologies embedded in CCTV cameras in Moscow. The 'Yarovaya Law', a package of anti-terrorism laws, provides a legal basis for widespread repression of dissent and narrative control. The law requires internet service providers to store public and private communication for six months, facilitating AI-driven surveillance. As part of the 'Sovereign Internet Law' (or Runet), which partitions Russia's internet from the rest of the internet, the government adopted Deep Packet Inspection (DPI) technology, allowing authorities to filter internet traffic in real-time.

The full-scale invasion of Ukraine in February 2022 has only intensified Russia's surveillance and content-blocking mechanisms through the use of algorithms. For example, the 'Oculus Project' uses AI-based text detection to limit the proliferation of information about the war in Ukraine and LGBTQ+ movements, and uses bots to flood platforms with pro-Kremlin content. The use of AI-based technologies in Russia is a cause of concern for human rights domestically, and for cyber-espionage, cyber warfare and exertion of influence internationally.

Iran

While Iranian algorithmic repression practices lag behind those of Russia and China, there has been an evolution in the use of AI-based technologies used to monitor internet traffic and apply Islamic moral standards to digital content. With the aim of 'digital sovereignty', Iran developed the National Information Network (NIN) to isolate Iranian users from the global internet by emphasising domestically hosted content. The use of NIN has facilitated state-mandated censorship by blocking counter-narratives that may appear on foreign websites. The Iranian regime portrays the pursuit of digital sovereignty as a safeguard against cyber threats and external interference.

Practically, this approach limits public opposition and contributes to the sense of omnipresent surveillance among journalists, activists and citizens. Examples include the use of facial recognition technology during protests, the use of AI-driven bots and automated accounts to amplify content favourable to the regime, and the use of AI-based tools to produce content in different languages, in order to reach a global audience. To enhance its AI-based capabilities, Iran has partnership agreements with the Chinese regime with major Chinese companies supplying technology to the police and military authorities. Iran imports hardware primarily from the UAE, but also from China, Turkey and India. Iran's use of AI is concerning because of its transnational reach, involving hacking campaigns and cyber espionage, social media manipulation to amplify pro-Iranian narratives abroad, and the use of surveillance technologies to monitor and intimidate Iranian dissidents abroad.

Egypt

Since the Arab Spring in 2011, Egypt has turned to AI as a tool for political repression and surveillance. The crackdown on freedom of expression and assembly in Egypt has increasingly taken digital forms. The Anti-Cyber and Information Technology Crimes Law of 2018 empowers the authorities to block websites and monitor online content to protect national security or the economy. The law, although envisioned to combat extremist and terrorist organisations, has affected dissidents and citizens. Women who create content on social media platforms have been targeted, with some charged with 'violating family principles and values of Egyptian society'. The authorities' increasing use of CCVT cameras and facial recognition technology raises concerns about the identification of participants in political protests. Reports show that the government has engaged in cyber espionage against journalists, academics, lawyers, opposition politicians and human rights activists. Internationally, Egypt's AI-driven surveillance has targeted expatriates and dissidents living abroad.

Sub-Saharan Africa

Ethiopia represents a novel case of algorithmic authoritarianism, particularly relevant for understanding the impacts of AI-driven repression on ethnically and culturally diverse societies. The country has undergone a digital expansion, and while this development has been largely positive, the possibility that AI-based technologies would be used for political repression is concerning. The US has provided Ethiopia with surveillance technology and training. There are examples of the Ethiopia's Information Network Security Agency using its surveillance of private communications to arrest lawful opposition activists.

Beyond Ethiopia, the use of surveillance systems without sufficient checks is a trend in countries in Sub-Saharan Africa. China's engagement with the continent has increased the accessibility of monitoring products. The African Digital Rights Network, a network of activists, academics and analysts researching digital rights in Africa, has signalled concerns about human rights and freedoms in the region in light of the growing technological capacities and legislation eroding digital rights.

Regulatory and governance initiatives

Recent years have seen a huge multiplication of initiatives aimed at establishing basic principles for AI technology governance by international and regional actors or 'blocs' of states, ranging from the UN, the OECD and the Council of Europe to the G20 and BRICS. The wide majority of instruments are non-binding, laying down guidelines for ethical, human-centric and responsible AI development. Notably, the Council of Europe is assessing the feasibility of a legal framework regulating AI based on the Council's human rights standards.

The EU's AI Act is the world's first comprehensive AI law. The AI Act aims to set clear standards that protect fundamental freedoms and safety, while balancing the economic potential of AI technologies. Through the EU's active involvement in different international and regional fora, the EU's framework may influence wider standards of AI governance and ethics.

In parallel, national regulations are emerging in the US, China and India, as well as in other countries, such as Brazil and Kenya.

Recommendations

The key recommendations of this comprehensive paper address the EU as a whole, and the European Parliament (EP) specifically.

For the EU

- To counter AI-based repression, the EU could impose either targeted sanctions against individuals, companies or governments that use AI for repressive purposes, or sectoral sanctions targeting sectors that contribute to the authoritarian use of AI, such as advanced computing, facial recognition

technology or surveillance equipment. To ensure that sanctions are effective, the EU should collaborate with allies, monitor the enforcement of the sanctions, and make the ethical and human rights-compliant use of AI a condition for accessing the EU markets.

- The EU could provide capacity building to civil society organisations and human rights defenders in authoritarian countries to improve their understanding of AI technologies and support them in recognising digital repression.
- The EU could use diplomatic channels to promote global regulatory convergence on AI, including by leveraging the international impact of the EU's General Data Protection Regulation (GDPR) to advocate for similar principles in the AI governance.
- Export controls are another way the EU could curb the misuse of AI technology for repressive purposes. The EU could update existing export regulations with a comprehensive list of sensitive AI technologies. The EU could also increase the transparency of AI exporting licences by requiring companies to report on the end-use of AI exports. Effective implementation would require developing mechanisms to verify end-use, and sanctioning non-compliance. The EU could leverage the Wassenaar Arrangement (a voluntary export control regime covering information exchange on transfers of arms and dual-use goods and technologies) to collaborate with international partners in establishing specific controls for AI technologies.
- The EU could use a multifaceted approach to expand its technical assistance to vulnerable states. Legislative or infrastructure support from EU could be offered, and a dedicated task force within the EEAS specialising in AI governance could be created.
- Finally, the EU could implement an AI auditing and certification mechanism with a clear set of criteria and standards. This mechanism could be leveraged within the EU's international agreements.

For the EP

- The EP could establish a dedicated parliamentary committee aimed at monitoring the implementation of the AI Act on the one hand, and global trends of AI repression and misuse, including new technological developments, on the other hand.
- The EP could play a pivotal role in international AI diplomacy by sharing regulatory best practices with parliamentary bodies beyond the EU, and by leading discussions on AI ethics at global summits.
- The EP could work towards developing a comprehensive list of technologies that could be used for repressive ends and ensure rigorous due diligence in the export process as part of the Parliament's legislative and scrutiny work.
- The EP is in a position to heighten public awareness about the potential risks associated with AI misuse.
- Finally, the EP could play a role in protecting whistle-blowers who expose unethical AI practices and support independent media.

This Executive summary has been prepared on the basis of the In-depth Analysis 'Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights' by H. Akin ÜNVER, Associate Professor, Özyeğin University, Turkey. It will be available in English at: ThinkTank.

Contacts in the European Parliament

Coordination: Rasma KASKINA, Policy Department for External Policies

Editorial assistant: Balázs REISS and Kristina WILHELMSSON

Feedback is welcome. Please write to poldep-expo@europarl.europa.eu

EP/EXPO/DROI/FWC/2019-01/LOT6/1/C/30
PE 754.450