GLOBAL CENTER ON COOPERATIVE SECURITY

POLICY BRIEF

Counterterrorism and Violence Prevention

Safeguarding Against the Misuse and Abuse of Artificial Intelligence

By Eleonore Pauwels
June 2022

Contemporary conflicts and efforts to prevent terrorism and violent extremism increasingly involve the use, misuse, and abuse of population data, which have contributed to a new geopolitics of insecurity that cuts across societies and borders. Such threats are of particular concern in conflict-prone and conflict-affected countries due to weak regulatory frameworks and growing cybersecurity and digital divides.

In this context, rapid advancements in artificial intelligence (AI) and machine learning present both challenges and opportunities for terrorism and counterterrorism efforts. Violent extremists and other hostile actors can exploit emerging AI technologies to sow disinformation and exacerbate polarization, target humans and their information systems, manipulate data sets, and attack critical infrastructure. At the same time, the embrace of AI and machine learning by states in the service of counterterrorism has the potential to exacerbate concerns about profiling and human rights. This brief examines these threats and the ways that international organizations, including the United Nations, can and should protect against the misapplication of these technologies by states and nonstate actors alike.

GROWING INTEGRATION OF AI IN GLOBAL SECURITY

The growing integration of AI across myriad aspects of our daily lives from our health and safety to security and economics provides compelling opportunities for optimizing the functioning of cities, businesses, critical services, and infrastructure, but it also presents new risks to civilian populations. Data capture technologies are collecting biometrics and data on human movements, consumption patterns, conversations, and emotions at increasing rates. The vast digital information generated by populations today means that more of these routine behaviors can be understood through AI-led computing. Advances in AI and machine learning can automate actors' capacity for massive data optimization, anomaly detection, behavior prediction, and behavioral systems analysis. Global responses to the COVID-19 pandemic have shown just how quickly algorithms and computing power can be combined with diverse data streams and used in technologies that monitor and control bodies and populations. For instance, in South Korea, algorithmic models use geolocation data, surveillance camera footage, and credit card records to trace coronavirus cases.[1]

1 Justin Fendos, "How Surveillance Technology Powered South Korea's COVID-19 Response," Brookings Institution, 29 April 2020, https://www.brookings.edu/techstream/how-surveillance-technology-powered-south-koreas-covid-19-response.

Although predictive technologies have a profound ability to impact violence prevention and efforts to stem terrorism, AI also has the potential to be destructive in the hands of nonstate actors. Although no evidence to date shows that terrorists have successfully implemented a large-scale physical or cyberattack through the use of AI, terrorism has always relied on access to new technologies. There is no reason to believe that terrorist groups will ignore the potential to weaponize AI. As terrorists have become increasingly sophisticated in their tactics and technology, AI represents a logical step forward in their trajectory.[2]

The convergence of AI, machine learning, and data sets provides a powerful analytical tool for behavioral monitoring and intelligence collection across security, counterterrorism, and violence prevention agendas. These technologies, along with the collection of biometric and other data, have been elevated in the UN Security Council as critical means to combat terrorism, and member states have been called on to share and pool their data.[3] In Resolution 2396, the Security Council required that states deploy and implement systems to collect biometric data, which could include fingerprints, photographs, and digital data on faces to accurately identify terrorists, including foreign terrorist fighters, in compliance with domestic law and international human rights law.[4]

The use of biometrics in counterterrorism is increasingly linked to the use of other emerging data capture technologies for predictive monitoring of recurring patterns in hate speech, violent extremism, and terrorism.[5] The analysis of routine activities can help predict drivers and patterns of violence (e.g., sustained human rights violations and online hate speech targeted at ethnic subgroups) and serve as early-warning signals of impending crises (e.g., social media trends or changes in the movements of refugees, armed groups, or even city traffic). The capture of behavioral biometric data can also help protect critical infrastructure and identify violent nonstate actors by distinguishing the features or activities of a specific group.[6] Algorithms now have access to a high volume and rich variety of data sets collected by smart-sensing technologies in mobile devices and within urban infrastructure. Combinations of technologies such as facial recognition, gait analysis, closed-circuit television cameras, and mobile biometrics can analyze the faces and bodies of individuals in moving crowds.[7] Other sources include communications metadata and Internet connection records, location and activity tracking, financial transactions, and social media activity.

One rapidly expanding domain is the reliance on AI and machine learning solutions within anti–money laundering and countering the financing of terrorism efforts. Within this realm, new technologies are being harnessed to better identify risks and monitor

2 UNCCT and UNICRI, "Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes," 2021, https://unicri.it/sites/default/files/2021-06/Malicious%20Use%20of%20AI%20-%20UNCCT-UNICRI%20Report_Web.pdf.

3 United Nations, "CTED Analytical Brief: Biometrics and Counterterrorism," UNSC CTED, 2020-2021, https://www.un.org/securitycouncil/ctc/sites/www.un.org.securitycouncil.ctc/files/files/documents/2021/Dec/cted_analytical_brief_biometrics_0.pdf. See also UNSC Counter-Terrorism Committee, "CTC Holds Open Briefing on the Work of CTED with Member States of South and South-East Asia Pursuant to Security Council Resolution 2395 (2017)," https://www.un.org/securitycouncil/ctc.

4 UN Security Council, S/RES/2396(2017), 21 December 2017.

5 Eleonore Pauwels, "Artificial Intelligence and Data-Capture Technologies in Violence and Conflict Prevention: Opportunities and Challenges for the International Community," Global Center on Cooperative Security, September 2020, https://www.globalcenter.org/wp-content/uploads/2020/10/GCCS_AIData_PB_H.pdf.

6 Priyanka Chaurasia et al., "Countering Terrorism, Protecting Critical National Infrastructure and Infrastructure Assets through the Use of Novel Behavioral Biometrics," *Behavioral Sciences of Terrorism and Political Aggression*, vol. 8, no. 3 (September 2016): 197–211, https://www.tandfonline.com/doi/full/10.1080/19434472.2016.1146788.

7 Shian-Ru Ke et al., "A Review on Video-Based Human Activity Recognition," *Computers* 2, no. 2 (June 2013): 88–131, https://www.researchgate.net/profile/Jang-Hee-Yoo/publication/285197344_A_Review_on_Video-Based_Human_Activity_Recognition/links/57439d5d08ae9ace841b3c72/A-Review-on-Video-Based-Human-Activity-Recognition.pdf; for "violence detection," see E. Bermejo et al., "Violence Detection in Video Using Computer Vision Techniques," Intel Labs Pittsburgh and Robotics Institute, Carnegie Mellon, 2011, https://www.cs.cmu.edu/~rahuls/pub/caip2011-rahuls.pdf; see also Sadegh Mohammadi et al., "Angry Crowds: Detecting Violent Events in Videos," European Conference on Computer Vision, 2016, https://link.springer.com/chapter/10.1007/978-3-319-46478-7_1.

suspicious activity within large volumes of financial, biometric, and behavioral data.[8] AI programs and machine-learning algorithms can automate the analysis of suspicious transactions and other illicit activity, minimizing the need for continuous human supervision. These tools can reduce inaccuracies and screen against emerging threats in real time, thus streamlining the assessments of customer due diligence (CDD) and risk. Digital identification technology can also strengthen the authentication mechanisms used to prevent money laundering, fraud, and terrorism financing. For instance, India has implemented CDD and customer risk assessments through Aadhaar, its biometrics database. When an individual enrolls in Aadhaar, biometric and personal information such as their name, gender, date of birth, and contact information is captured and incorporated by the Unique Identification Authority of India. Similar digital identification and biometric repositories are being built across the African continent, with Ghana, Kenya, Nigeria, Somaliland, South Africa, Tanzania, and Uganda in the process of logging their populations' biometric data in centralized national databases.[9]

With the convergence of data capture technologies and algorithms, actors across the global research and security industry posit that significant amounts of raw population data can be used to monitor terrorist activity and forecast a range of threats.[10] For instance, the Global Internet Forum to Counter Terrorism (GIFCT) brings together governments, the technology industry, and other actors to promote cooperation on preventing terrorists and violent extremists from exploiting digital platforms. Security experts are increasingly able to monitor how violent nonstate actors operate, collude with transnational criminal networks, finance their operations, and adapt to new domains. This includes a greater understanding of when online hate speech and other harmful social media trends can incite widespread physical violence.[11]

Yet despite these transformative opportunities, the increasing reliance of modern societies on AI and other dual-use technologies makes them vulnerable to exploitation by hostile actors—governments and nonstate actors alike. Three defining sociotechnical trends will affect the near future of human security, counterterrorism, and violence prevention.

TREND 1: BEHAVIORAL SURVEILLANCE, RACIAL PROFILING, AND HUMAN RIGHTS CONCERNS

States and private sector actors in the global security industry may increasingly misuse and abuse AI and population data sets for social surveillance and control, repression, and racial profiling.

In the service of counterterrorism efforts, authorities and private sector actors rely on population data sets, including biometric and financial databases, geolocation, and social media networks, to improve due diligence and better identify, understand, and manage money laundering and terrorism financing risks. Yet, without appropriate governmental and private sector standards and safeguards, the use of AI and mass population data analysis in the financial sector harbors risks of abuse and could potentially undermine privacy, inclusion, and equity. When adopted without proportional and risk-based approaches, digital identification technology can lead to new forms of "de-risking" or "de-banking," ultimately resulting in the financial exclusion of underserved communities. For instance, the rollout of Kenya's biometric identity

8 Financial Action Task Force, "Opportunities and Challenges of New Technologies for AML/CFT," July 2021, https://www.fatf-gafi.org/media/fatf/documents/reports/Opportunities-Challenges-of-New-Technologies-for-AML-CFT.pdf.

9 Eleonore Pauwels, "The Anatomy of Information Disorders in Africa: Geostrategic Positioning & Multipolar Competition over Converging Technologies," Konrad Adenauer Foundation, July 2020, https://www.kas.de/en/web/newyork/single-title/-/content/the-anatomy-of-information-disorders-in-africa.

10 Fred Baradari, "Perspective: AI Is Changing the Game for Situational Awareness," *Homeland Security Today*, 10 April 2019, https://www.hstoday.us/subject-matter-areas/law-enforcement-and-public-safety/perspective-ai-is-changing-the-game-for-situational-awareness.

11 UN Office of Counter-Terrorism (UNOCT), "Countering Terrorism Online with Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia," UNCCT and UNICRI, 2021, https://www.hstoday.us/subject-matter-areas/law-enforcement-and-public-safety/perspective-ai-is-changing-the-game-for-situational-awareness.

program was ruled unconstitutional in October 2021 after significant concerns came to light about the government's lack of a plan to protect Kenyans' biometric data.[12] This was not the first time the program faced controversy. In January 2020, it was reported that at least five million Kenyans had been unable to obtain the documents required to obtain biometric identification, a system that is proving to be particularly burdensome for ethnic, racial, and religious minorities.[13] The additional hurdles these groups face in obtaining an identity card exacerbate entrenched inequalities and ethnic tensions. Human rights defenders in India have reported similar cases of discrimination and privacy breaches related to Aadhaar.[14] These examples contradict the guidelines put forth by the World Bank and the United Nations to address the legal, procedural, and social barriers that populations may face in accessing digital identification systems. Such considerations are particularly critical to prevent the exclusion of traditionally marginalized groups such as women, children, rural populations, migrants, and stateless persons, as well as ethnic, linguistic, and religious minorities.[15]

The use of systemic, automated, and predictive behavioral analysis in counterterrorism and violence prevention is likely to pose further fundamental challenges for protecting human rights in fragile contexts. These implications include limits to self-determination and political agency, privacy and data protection violations, discrimination, exposure to pervasive data security breaches, and new forms of censorship in the virtual civic space.

First, governments could potentially be collecting and managing increasingly large amounts of sensitive data about vulnerable populations and adopting behavioral monitoring technologies created by private sector actors in complex, weakly regulated supply chains.[16] In their report *Use of Biometric Data to Identify Terrorists,* Krisztina Husti-Orbán and Fionnuala Ní Aoláin cautioned that, "in the absence of robust human rights protections, which are institutionally embedded to oversee collection, storage, and use of such evidence, relevant practices are likely to infringe international human rights law standards."[17]AI and machine learning pose profound challenges to privacy because they can automate the detection of "anomalies" or "abnormal behavior" in bulk data routinely collected about individuals and crowds. For example, image and speech recognition algorithms can detect objects in blurry photographs or separate voices in crowded environments. By introducing new opportunities for authoritarian states or violent nonstate actors to control populations, algorithmic surveillance threatens political participation, peaceful assembly, and freedom of movement.

The private surveillance industry has already provided states and nonstate actors with an arsenal of tools for computer interference and mobile device hacking.[18] A 2021 report by cybersecurity and legal experts exposes how, "through 'phishing' operations, social engineering, malware downloads, and gaining access to passwords and networks through security force intimidation, the [Syrian Electronic Army] and the Assad regime have used these practices to monitor

12 Chris Burt, "Kenya's Digital ID Ruled Illegal Until Data Protection Impact Assessment Completed," Biometric Update, October 15, 2021, https://www.biometricupdate.com/202110/kenyas-digital-id-ruled-illegal-until-data-protection-impact-assessment-completed.

13 Abdi Latif Dahir, "Kenya's New Digital IDs May Exclude Millions of Minorities," *The New York Times*, 28 January 2020, https://www.nytimes.com/2020/01/28/world/africa/kenya-biometric-id.html.

14 "India: Identification Project Threatens Rights," Human Rights Watch, 13 January 2018, https://www.hrw.org/news/2018/01/13/india-identification-project-threatens-rights.

15 World Bank, "Principles on Identification for Sustainable Development: Toward the Digital Age," 2021, https://documents1.worldbank.org/curated/en/213581486378184357/pdf/Principles-on-Identification-for-Sustainable-Development-Toward-the-Digital-Age.pdf; United Nations, *Compendium of Recommended Practices for the Responsible Use & Sharing of Biometrics in Counter Terrorism*, 2018, https://www.unodc.org/pdf/terrorism/Compendium-Biometrics/Compendium-biometrics-final-version-LATEST_18_JUNE_2018_optimized.pdf.

16 UN Human Rights Council, "Surveillance and Human Rights, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression," A/HRC/41/35, 2019, https://www.undocs.org/A/HRC/41/35.

17 Krisztina Husti-Orbán and Fionnuala Ní Aoláin, "Use of Biometric Data to Identify Terrorists: Best Practice or Risky Business?" Human Rights Center, University of Minnesota, 2020, https://www.ohchr.org/Documents/Issues/Terrorism/biometricsreport.pdf.

18 UN Human Rights Council, "Surveillance and Human Rights."

and track down activists and human rights defenders in Syria, who are then tortured and killed."[19] In the near future, illiberal governments may have access to a new class of malware that targets users' biometrics. A 2020 study found that AI models can deduce, with an accuracy rate of 75–93 percent, what a human is typing by analyzing their shoulder movements during video calls, showing the ease with which passwords and other confidential information can be gleaned.[20] In a 2018 "red-teaming" exercise intended to emulate the exploitation capabilities of a potential adversary, IBM engineers designed "DeepLocker," AI-powered malware that can hide a cyberthreat such as ransomware in a video conference application and launch only when it identifies the face of the target.[21] Representatives from minority groups could be stigmatized in new and powerful ways with malware designed to target them specifically.

Another major concern is the potential for automated ethnic profiling in conflicts and counterterrorism efforts.[22] A 2021 report by two UN entities notes that drones equipped with facial recognition capabilities can be used to carry out targeted attacks on specific communities.[23] Drones and police body cameras equipped with facial recognition and other biometric-capture capabilities have been used to profile social and racial justice activists, including during peaceful demonstrations.[24] Such forms of behavioral surveillance can help states and nonstate actors anticipate the movement of populations during protests, elections, and religious or social events to better impose control or repressive measures.

The convergence of AI, biometrics, and techniques that use facial, voice, gait, and DNA samples is already used in China to profile individuals based on their ethnicity and phenotype. In recent years, multiple investigations by human rights actors have revealed how the Chinese government has used facial recognition tracking and biometric data collection, including DNA samples and voice samples, against its Uyghur population.[25] Using these data, China maintains a vast system of facial recognition algorithms to identify and monitor Uyghurs based on their appearance, often under the guise of terrorism concerns.[26] Given its dominance in the technology market, China's model of algorithmic surveillance potentially could be exported for use around the world.

Several countries in sub-Saharan Africa are also amassing increasingly sophisticated surveillance technology. For instance, the AI software called Sentry is used to detect "abnormal behavior" on the streets of Johannesburg.[27] Mobile biometric devices deployed by Ugandan police rely on AI to identify an individual

19 Access Now et al., "Digital Dominion: How the Syrian Regime's Mass Digital Surveillance Violates Human Rights," March 2021, https://www.accessnow.org/cms/assets/uploads/2021/03/Digital-dominion-Syria-report.pdf.

20 Conor Cawley, "AI Can Now Guess Your Password by Looking at Your Shoulders," Tech.co, 11 November 2020, https://tech.co/news/ai-guess-password-shoulders.

21 Marc Ph. Stoecklin, Dhilung Kirat, and Jiyong Jang, "DeepLocker: How AI Can Power a Stealthy New Breed of Malware," *Security Intelligence*, 18 August 2018, https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware.

22 In a June 2020 report, the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia, and related intolerance analyzed different forms of racial discrimination in the design and use of emerging digital technologies. Tendayi Achiume, "Racial Discrimination and Emerging Digital Technologies: A Human Rights Analysis," A/HRC/44/57, UN Human Rights Council, June 2020, https://digitallibrary.un.org/record/3879751?ln=en.

23 UNOCT, "Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes," June 2021, https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf.

24 Malkia Devich-Cyril, "Defund Facial Recognition: I'm a Second-Generation Black Activist, and I'm Tired of Being Spied On by the Police," *Atlantic*, 5 July 2020, https://www.theatlantic.com/technology/archive/2020/07/defund-facial-recognition/613771.

25 Maya Wang, "Eradicating Ideological Viruses: China's Campaign of Repression Against Xinjiang's Muslims," Human Rights Watch, 2018, https://www.hrw.org/report/2018/09/09/eradicating-ideological-viruses/chinas-campaign-repression-against-xinjiangs; Danny O'Brien, "Massive Database Leak Gives Us a Window Into China's Digital Surveillance State," Electronic Frontier Foundation, 1 March 2019, https://www.eff.org/deeplinks/2019/03/massive-database-leak-gives-us-window-chinas-digital-surveillance-state.

26 Paul Mozur, "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority," *The New York Times*, 14 April 2019, https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

27 Michael Kwet, "Smart CCTV Networks Are Driving an AI-Powered Apartheid in South Africa," Vice News, 22 November 2019, https://www.vice.com/en_us/article/pa7nek/smart-cctv-networks-are-driving-an-ai-powered-apartheid-in-southafrica.

on the spot, sparking fears that the technology could be used to single out members of the LGBTQIA+ community.[28] The Zimbabwean government hired a Chinese firm to build a national facial recognition and monitoring system through an agreement that involves sharing the biometric records of Zimbabweans without their consent.[29] Such a national identity system, paired with compulsory SIM card registration data, provides the architecture for high-precision population surveillance, from physical movements to online activities.

The convergence of AI and large-scale data collection enables the documentation and profiling of human movements, habits, and even emotions with tremendously powerful applications. The lack of safeguards has enabled state and nonstate actors alike to manipulate individuals' deepest fears, hatreds, and prejudices.

TREND 2: SCALED-UP EMOTIONAL MANIPULATION AND INFLUENCE OPERATIONS

Illiberal regimes, violent extremists, and other violent actors may increasingly rely on emotional manipulation and influence operations to spread disinformation, increase political polarization, and sow social unrest and ethnic conflict.

While the malicious manipulation of information is not a new phenomenon, applying AI to population behavioral data is drastically enhancing methods and techniques in influence operations by changing the strategic communication environments in which intrastate violence and terrorism play out. Both state and nonstate actors can feed their own narratives and mis- and disinformation to their constituents within and across borders. Russian troll factories outsourced business to trolls in Ghana and Nigeria working to foment racial tensions around police brutality in the United States ahead of the 2020 election.[30] In India's West Bengal region, Rohingya refugees have been demonized by the same kinds of extreme threats and online hate-mongering that caused them to flee Myanmar.[31] In Kenya and South Africa, disinformation and hate speech, manufactured in part by political elites, inflamed the racial and socioeconomic divisions that have plagued both countries for decades.[32]

With AI technologies that can mimic and synthetize media from scratch, including text, images, and audio and video samples, the craft of emotional manipulation is becoming ever more powerful and has the potential to inflict harm on specific ethnic groups and other vulnerable communities. In 2019, researchers published a new method for making so-called deepfakes—creating realistic face-swapped videos in real time—with no extensive facial data-training.[33] A sophisticated natural language processing model (OpenAI's GPT-3) can help human operators automate and scale tasks common to modern disinformation campaigns, including promoting false narratives and conspiracy theories.[34] Such emotional manipulation techniques could be used within influence operations

28 Stephen Mayhew, "Ugandan Police Deploy Gemalto Tech for Rapid Capture of Suspects' Biometric Data," Biometric Update, 11 February 2019, https://www.biometricupdate.com/201902/ugandan-police-deploy-gemalto-tech-for-rapid-capture-of-suspects-biometric-data.

29 Samuel Woodhams, "How China Exports Repression to Africa," *Diplomat*, 23 February 2019, https://thediplomat.com/2019/02/how-china-exports-repression-to-africa.

30 Clarissa Ward et al., "Russian Election Meddling Is Back—Via Ghana and Nigeria—and in Your Feeds," CNN, 11 April 2020, https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html.

31 Vindu Goel and Shaikh Azizur Rahman, "When Rohingya Refugees Fled to India, Hate on Facebook Followed," *The New York Times*, 14 June 2019, https://www.nytimes.com/2019/06/14/technology/facebook-hate-speech-rohingya-india.html.

32 Dave Segal, "How Bell Pottinger, P.R. Firm for Despots and Rogues, Met Its End in South Africa," *The New York Times*, 4 February 2018, https://www.nytimes.com/2018/02/04/business/bell-pottinger-guptas-zuma-south-africa.html.

33 Samantha Cole, "This Program Makes It Even Easier to Make Deepfakes," Vice News, 19 August 2019, https://www.vice.com/en/article/kz4amx/fsgan-program-makes-it-even-easier-to-make-deepfakes; Yuval Nirkin, Yosi Keller, and Tal Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, https://openaccess.thecvf.com/content_ICCV_2019/html/Nirkin_FSGAN_Subject_Agnostic_Face_Swapping_and_Reenactment_ICCV_2019_paper.html.

34 Ben Buchanan et al., "Truth, Lies and Automation–How Language Models Could Change Disinformation," CSET Report, May 2021, https://cset.georgetown.edu/publication/truth-lies-and-automation.

led by factions in conflict, from ruling elites and political parties to terrorist groups.[35]

As human behavioral data capture becomes more advanced, malicious actors are increasingly able to rely on AI's capacity for predictive analysis to identify the emotional triggers that push individuals and groups to violence. For instance, criminal groups may spread false claims of violence committed by their enemies to inflame tensions and gain sympathy for their cause. When the Islamic State of Iraq and the Levant increased its power and visibility through social media, its violent propaganda, which used doctored videos and AI bots to magnify messaging, resulted in a wave of online emotional warfare.[36] The violent anti-Islamic backlash that followed was then instrumentalized for the group's recruiting strategies.

A related trend is the increasing tendency of certain actors to exploit digital networks to exacerbate existing racial, social, and economic divisions. In many democracies that are beginning to implement new technologies, where privacy and data protection policies have not been followed by robust regulations, state and private sector actors can extract sensitive personal data from an array of online population databases to target specific ethnic and socioeconomic groups.[37] These platforms can provide precise insights into a person's ethnic background, political affiliation, education level, wealth, location, and contact information. Relying on the aggressive campaigns produced by data analytics companies, political parties can then exploit citizens' personal profiles and information networks to spread rumors, propaganda, hate speech, and mis- and disinformation. For instance, political campaigns in Kenya[38] in 2017 and in Nigeria[39] in 2015 relied on video propaganda that capitalized on ethnic and socioeconomic tensions to target segments of the electorate defined by ethnicity, political affiliation, and age.

At the same time, behavioral surveillance tools can be used to silence civil society and any media outlets that challenge the status quo. In 2020, false news outlets deliberately targeted Spanish speakers in Florida with disinformation designed to sow discord between Black Lives Matter activists and Latino voters, depress the Latino vote, and question presidential candidate Joseph Biden's Catholic faith in attempts to reduce the Democratic share of Latino votes. After voting ended, this propaganda campaign remained active, sharing disinformation that claimed the election was rigged. *The New York Times* found that, in 24 hours, this hypertargeted disinformation effort had generated traffic that eclipsed the 2016 Russian-based election interference campaign.[40]

The protection of civilians and the national interest against cyberterrorism and disinformation also runs the risk of diminishing the privacy of the very civilians that governments are meant to protect. In recent years, Kenya, Nigeria, and other states have proposed legislation and passed cybersecurity laws in the name of defending and protecting the national interest in the fight against terrorism, which in many instances has directly undermined individual rights.[41] Beyond violations of human rights and freedom of expression, such national security measures have resulted in the shrinking of "virtual civic space."

Applying AI to population data sets to monitor and engineer human behavior has direct implications across the counterterrorism, violence prevention, and

35 Pauwels, "Anatomy of Information Disorders in Africa."

36 Antonia Ward, "ISIS's Use of Social Media Still Poses a Threat to Stability in the Middle East and Africa," TheRANDBlog, 11 December 2018, https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html. See Majid Alfifi, Parisa Kaghazgaran, and James Caverlee, "Measuring the Impact of ISIS Social Media Strategy," 2018, http://snap.stanford.edu/mis2/files/MIS2_paper_23.pdf.

37 Robert Muthuri et al., "Biometric Technology, Elections, and Privacy: Investigating Privacy Implications of Biometric Voter Registration in Kenya's 2017 Election Process," Centre for Intellectual Property and Information Technology Law, 2017.

38 George Obulutsa, "Kenya President's Election Campaign Used Firm Hired by Trump: Privacy Group," Reuters, 14 December 2017.

39 Geoffrey York, "Cambridge Analytica Parent Company Manipulated Nigeria's 2007 Election, Documents Show," *Globe and Mail*, 29 March 2018, https://www.theglobeandmail.com/world/article-cambridge-analytica-parent-company-manipulated-nigerias-2007-election.

40 Patricia Mazzei and Nicole Perlroth, "False News Targeting Latinos Trails the Election," *The New York Times*, 5 November 2020, https://www.nytimes.com/2020/11/04/us/spanish-language-misinformation-latinos.html.

41 Humphrey Malalo and Omar Mohammed, "Kenya's President Signs Cybercrimes Law Opposed by Media Rights Groups," Reuters, 16 May 2018.

human rights agendas. AI research labs, private companies, counterterrorism and other security professionals, human rights experts, and other civil society actors will need to diligently monitor the proliferation of training data, malicious codes, and deep-learning techniques that could be harnessed to manipulate the media and influence populations.

TREND 3: THREATS TO CIVILIAN DATA SETS AND CRITICAL INFRASTRUCTURE

Hostile state and nonstate actors can perpetrate adversarial data-manipulation attacks that generate widespread civilian harm, corrupting societies' digital repositories and compromising the functioning of critical infrastructure and industrial control systems.

Malicious cyberoperations, which are increasingly automated, are targeting and weaponizing the growing interdependence among AI, machine learning, and the systems critical to the safety and well-being of civilian populations. Cyberattacks worldwide have targeted civilian sectors from finance and health to industry and energy, affecting vital systems including nuclear power plants, complex supply chains, and biotechnology.[42]

Recent AI and cybersecurity studies have confirmed that a certain type of deep-learning algorithm can be trained to manipulate medical and genome data sets, opening the doors to possible cyberattacks throughout the health, biotech, and biosecurity sectors. In 2019, AI security experts at Ben-Gurion University designed a malicious attack to modify cancer data in hospital CT scans, generating false lung tumors and subsequent misdiagnoses.[43] As part of a red-teaming test, researchers at Sandia National Laboratory also demonstrated how a malware injection can compromise genetic analysis software and manipulate raw genome data.[44] These types of malicious data tampering could result in clinical misdiagnoses with potentially deadly ramifications. The macro implications are also incredibly deleterious, as the corruption of large-scale data sets on pathogens and infectious diseases could undermine the integrity of biomedical information at the global level. Such adversarial techniques pose similar risks to data-driven domains beyond the biotech and medical sectors.

Cyberattacks augmented by adversarial algorithms have the capacity to sabotage industry, governance, and financial systems. Such harmful cyberoperations can increasingly target the automated data protocols that help run critical systems (energy, water, sanitation) and food and industrial manufacturing supply chains. For instance, in 2018 a petrochemical company with a plant in Saudi Arabia was targeted by a new kind of cyberattack designed to compromise its safety protocols and trigger an explosion.[45] These attacks can be carried out remotely and on a large scale, with the potential for significant spillover to other essential humanitarian and civilian services.

The COVID-19 crisis has provided states and terrorist groups with a real-time window into societies' strengths and weaknesses in emergency situations. For instance, in December 2020, cybercriminals and state actors mounted targeted cyberoperations against biotech firms working on the COVID-19 vaccine, demonstrating the potential to access and manipulate

42 United States of America, "Cyberspace Solarium Commission Report," March 2020, https://www.google.com/url?q=https%3A%2F%2Fwww.fdd.org%2Fanalysis%2F2020%2F03%2F11%2Fcyberspace-solarium-commission-report%2F&sa=D&sntz=1&usg=AFQjCNFT-fSIrDsrR-bI6lgXGMoWyvJUyw; "Playing with Lives: Cyberattacks on Healthcare Are Attacks on People," CyberPeace Institute, March 2021, https://cpi.link/sar001; "Advances in Science and Technology to Combat Weapons of Mass Destruction (WMD) Terrorism," UNICRI, June 2021, https://unicri.it/News/Science-Technology-to-Combat-Weapons-of-Mass-Destruction-WMD-Terrorism. "Roadmap for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation," Report of the Secretary-General, June 2020, https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf.

43 Yisroel Mirsky et al., "CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning," Cornell University, arxiv.org, 11 January 2019, https://arxiv.org/abs/1901.03597v3.

44 Corey M. Hudson, "From Buffer Overflowing Genomics Tools to Securing," DEF CON 27 Bio Hacking Village, https://www.youtube.com/watch?v=7du1TltZOJg.

45 David E. Sanger, "Hack of Saudi Petrochemical Plant Was Coordinated From Russian Institute," *The New York Times*, 23 October 2018, https://www.nytimes.com/2018/10/23/us/politics/russian-hackers-saudi-chemical-plant.html.

information about how the vaccine is produced, utilized, and distributed.[46] More broadly, malicious actors—state or nonstate—may capitalize on chaos and public distrust to advance their influence. Rising threats to population data sets and civilian infrastructure could seriously undermine citizens' trust in collective safety measures, from public health guidance to emergency response systems.

Adversarial information operations that target the knowledge, industrial, and governance sectors are a powerful type of hybrid threat. They may target systemic vulnerabilities in civilian and security interfaces, interfere with multiple levels of strategic and emergency decision-making, and involve broad coalitions of malicious actors.[47] The capacity of state and nonstate actors alike to damage public confidence and destabilize critical governance institutions could have significant, long-term implications for peace and security.

RECOMMENDATIONS

Governments across the globe, especially in fragile contexts, face complex challenges in addressing the changing nature of conflicts and violent extremism. Technologically advanced nation-states and proxies are waging increasingly robust influence operations. The global proliferation of AI and other emerging technologies, driven by economic and national security interests, may lead to a diffusion of power among violent nonstate actors, including illicit transnational networks and globally connected violent extremist groups.

There is, thus, an urgent need to anticipate and devise ways for the private and public sectors to harness and regulate the potential of dual-use technologies. Decision-makers in these sectors will have to adapt, revise, and upgrade governance models to mitigate harm to populations affected by conflicts in an era of technological decentralization and automation, especially in cases where responsibility and liability for harm may not be clearly defined.

More sobering, the need to monitor how authoritarian and nonauthoritarian regimes are able to co-opt powerful technological capabilities for population surveillance, disinformation, adversarial data manipulation, and power and resource capture will become ever more urgent. The complex and decentralized emerging technology supply chains could lead to an unprecedented diffusion of power and dual-use potential in conflicts and among violent extremist actors.

In this context, recent developments in the European Union are worth further examination.[48] In April 2021, the European Commission unveiled draft regulations to govern the use of AI applications with the potential to threaten people's safety or fundamental rights in high-risk areas. Some of these, including live facial recognition in public spaces, may be prohibited altogether, with exemptions for national security and other purposes.[49] Under the proposed law, companies deploying AI in high-risk areas would have to provide regulators with risk assessments and proof of human oversight through the AI life cycle, as well as a list of transparency and accountability safeguards. This regulatory move by the EU is significant, as it complements comprehensive data-privacy regulations and related discussions about social media content-moderation laws.

Governments, multilateral bodies such as the United Nations, civil society, and private sector actors must evaluate potential regulatory measures in light of international human rights and humanitarian law, as well

46 David E. Sanger and Sharon LaFraniere, "Cyberattacks Discovered on Vaccine Distribution Operations," *The New York Times*, 3 December 2020, https://www.nytimes.com/2020/12/03/us/politics/vaccine-cyberattacks.html.

47 UN General Assembly, "Use of Mercenaries as a Means of Violating Human Rights and Impeding the Exercise of the Right of Peoples to Self-Determination," A/76/151, 15 July 2021.

48 "Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters," Procedure File: 2020/2016(INI), European Parliament, Legislative Observatory, https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure do?reference=2020/2016(INI) &l=en; European Commission, "A European Approach to Artificial Intelligence," Shaping Europe's Digital Future, https://digital-strategy.ec.europa .eu/en/policies/european-approach-artificial-intelligence.

49 European Parliament, "Use of Artificial Intelligence by the Police: MEPs Oppose Mass Surveillance," News European Parliament, 6 October 2021, https://www.europarl.europa.eu/news/en/press-room/20210930IPR13925/use-of-artificial-intelligence-by-the-police-meps-oppose-mass-surveillance.

as the UN Guiding Principles on Business and Human Rights. Significant normative debates should focus on the need for

- **Bolstering Multistakeholder Engagement—** Sharing due diligence and normative guidance, as well as building policy capacity across the technology, policymaking, civil society, violence prevention, and counterterrorism sectors. In recent years, new cross-sector and interdisciplinary partnerships, such as the GIFCT, CyberPeace Institute, and Biometrics Institute, have allowed UN prevention actors, policymakers, and technology companies to collaborate on normative guidance, early-warning, and accountability mechanisms.
- **Developing Human Rights Impact Assessments—**Conducting ongoing, in-depth human rights impact assessments for cases of potentially sensitive use of AI in counterterrorism and violence prevention. Cross-sector collaborations to anticipate unforeseen misuses and long-term impacts of AI and data capture technologies on vulnerable populations are essential to ensuring accountability. These mechanisms could forecast less predictable outcomes, such as general purpose or civilian technologies being misused by authoritarian states or violent nonstate actors. It will be important to engage with an appropriate and diverse array of external stakeholders, including nongovernmental organizations, researchers, and advocacy groups to ensure that their feedback informs UN, humanitarian, and private sector applications from the outset.
- **Providing Appropriate Safeguards—**Designing effective mechanisms and safeguards to secure population data sets. This includes developing and operationalizing phase-gate processes that allow and require technological firms, policymakers, and actors in the counterterrorism and violence prevention sectors to consider the negative consequences and potential misuses of AI technologies during the full data life cycle, as well as during development, deployment, and postdeployment.

There is an urgent imperative to find the right balance between proportionality in data collection, diligent sharing policy, and effective mechanisms for securing population data sets. State and nonstate actors must also devise adequate methods to weigh the benefits of harnessing AI technologies in the service of counterterrorism, especially with regard to automated and predictive behavioral monitoring, against the costs to civilian security and human rights. Counterterrorism and violence prevention actors should adhere to the purpose limitation principle in the collection, storing, processing, retention, and sharing of population data sets. States and private sector actors will have to manage the tension between the opportunity to apply algorithmic analytics to mass population data and the privacy principles of data minimization and proportionality in intelligence collection.

ABOUT THE AUTHOR

Eleonore Pauwels

Eleonore Pauwels is a Senior Fellow for the Global Center on Cooperative Security. Her research focuses on security, governance, and ethical implications generated by the convergence of artificial intelligence (AI) with other dual-use technologies, including cybersecurity, genomics, and genome editing. She regularly consults for the World Bank, the United Nations, governments, and private sector actors on AI and cybersecurity, the changing nature of conflict, foresight, and global security and counterterrorism.

ACKNOWLEDGMENTS

The Global Center gratefully acknowledges the support for this policy brief provided by the government of Norway. The views expressed are those of the author and do not necessarily reflect those of the Global Center or its advisory council, board, or sponsors or the government of Norway.

ABOUT THE GLOBAL CENTER

The Global Center on Cooperative Security works to achieve lasting security by advancing inclusive, human rights–based policies, partnerships, and practices to address the root causes of violent extremism. We focus on four mutually reinforcing objectives:

- Supporting communities in addressing the drivers of conflict and violent extremism.
- Advancing human rights and the rule of law to prevent and respond to violent extremism.
- Combating illicit finance that enables criminal and violent extremist organizations.
- Promoting multilateral cooperation and rights-based standards in counterterrorism.

Our global team and network of experts, trainers, fellows, and policy professionals work to conduct research and deliver programming in these areas across sub-Saharan Africa, the Middle East and North Africa, and South, Central, and Southeast Asia.

GLOBAL CENTER ON COOPERATIVE SECURITY

globalcenter.org
@GlobalCtr

NORWEGIAN MINISTRY OF FOREIGN AFFAIRS