

AUTOMATING REPRESSION BEYOND BORDERS:

How AI is Powering Transnational Repression



European Center for
Not-for-Profit Law

Automating Repression Beyond Borders: How AI is Powering Transnational Repression

Authors: Ana Sofia Harrison, Marlena Wisniak

Acknowledgments: Ivana Rosenzweigova, Vanja Skoric, Francesca Fanucci, Esther Meester

Design: Boglárka Szalma



European Center for
Not-for-Profit Law

www.ecnl.org

This paper is available under the Creative Commons license: [CC-BY SA 4.0 Attribution
ShareAlike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/)

March 2026

TABLE OF CONTENTS

I. INTRODUCTION	4
II. AI-ENABLED SURVEILLANCE	8
Biometric surveillance	8
Predictive policing and targeting	10
Algorithmic risk assessments and blacklists	11
Smart cities	12
III. AI PLATFORMS	14
Disinformation and harassment campaigns	14
Social media monitoring and censorship	16
AI agents	18
IV. POLICY OVERVIEW	20
V. CONCLUSION	23

I. INTRODUCTION

Repression no longer stops at national borders. In today's hyperconnected world, states increasingly extend their reach beyond their territories to silence critics, intimidate dissidents and suppress human rights advocacy abroad. This phenomenon, known as transnational repression (TNR), encompasses state-directed actions aimed at deterring, punishing, or criminalising human rights work conducted outside a country's borders. The Office of the United Nations High Commissioner for Human Rights (OHCHR) describes TNR as “acts conducted or directed by a State, or its proxy, to deter, silence or punish dissent, criticism or human rights advocacy towards it, expressed from outside its territory.” In late 2025, the European Parliament adopted a resolution on TNR, proposing a comprehensive definition that reflects evolving scope and severity of these practices, including digital dimensions.¹

The digital transformation of this phenomenon has fundamentally altered its reach and impact. What was once constrained by geography is now amplified by digital technologies, allowing authoritarian power to travel quickly, efficiently, and at unprecedented scale. “Digital transnational repression” (DTR) refers to the use of digital technologies by states or their proxies to monitor, intimidate, silence, or otherwise suppress individuals or groups across borders, frequently targeting diaspora communities, exiled journalists, human rights defenders and political dissidents. Researchers at The Citizen Lab describe it as “the cross-border targeting by states of individuals in exile or in the diaspora using digital technologies to repress any form of dissent.”² Tactics include digital surveillance, online harassment, disinformation campaigns and the deployment of spyware systems.

Often characterised as “threats from a distance,”³ DTR allows states to exert coercive influence without the immediate physical presence of security forces. Yet these practices

¹ “The European Parliament proposes defining transnational repression as attacks and threats by states, including authoritarian regimes and their proxies, that aim to defend and advance their interests by reaching across national borders to coerce, control or silence dissidents, political opponents, journalists, activists, HRDs and diaspora members, through a broad range of physical methods, such as targeted killings, abductions, violence, harassment and enforced returns, disappearances and deportations, and the strategic misuse of legal instruments, including abuse of consular services, extradition procedures or red notices, and arrests, as well as non-physical methods, such as digital surveillance, intimidation, blackmail and threats against HRDs’ families”
https://www.europarl.europa.eu/doceo/document/TA-10-2025-0258_EN.html

² Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online.
<https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

³ In Freedom House’s 2025 Freedom on the Net report, Freedom House identifies four modes of transnational repression: physical attacks, the co-optation of host states where targeted dissidents and activists reside, or of international institutions and processes

rarely exist in isolation. Digital repression frequently operates alongside other forms of transnational coercion, including mobility tracking, exploitation of international legal mechanisms such as INTERPOL, and direct physical violence. Among the most psychologically devastating tactics is “coercion-by-proxy,” whereby states target family members who remain in the country of origin to pressure activists or journalists abroad.⁴ These threats leverage emotional bonds and family responsibilities, generating fear, guilt, and shame that can lead individuals to self-censor, disengage from advocacy, or cease their work altogether.

The integration of artificial intelligence (AI) has dramatically escalated both the sophistication and severity of these practices. As rapid technological development accelerates, AI systems are increasingly weaponised within DTR, exacerbating and accelerating existing repression.⁵ Facial recognition, video surveillance and CCTV networks, predictive policing, deepfakes, social media content moderation, and disinformation campaigns can all be powered by AI to amplify the reach, efficiency and sophistication of TNR. When enhanced by emerging technologies, TNR operates with reduced visibility, making it more difficult to hold authoritarian governments accountable for human rights violations. This technological advancement has made TNR more attractive and useful for governments with authoritarian agendas.

Beyond technological concerns, DTR raises fundamental questions about sovereignty and international law. DTR increasingly blurs the line between intelligence gathering and law enforcement. While traditional, non-digital enforcement abroad without consent is universally prohibited because it breaches the core principle of territorial sovereignty, digital forms of enforcement are not yet treated with the same clarity. Recognising DTR as a form of enforcement, rather than intelligence, makes clear why it conflicts directly with the long-standing prohibition on extraterritorial policing.⁶

The human cost of these evolving practices extends far beyond individual targets. One of the most troubling consequences is DTR's dual impact: the global expansion of authoritarian control and censorship, alongside the normalisation of pervasive self-

(exploiting INTERPOL systems), mobility controls, and 'threats from a distance,' using digital technologies (Schenkkan & Linzer, 2021, p. 9).

⁴ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

⁵ 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. <https://artificialintelligenceact.eu/article/3/>

⁶ Michaelson, M., Thumfart, J. (2023). Drawing a line: Digital transnational repression against political exiles and host state sovereignty. European Journal of International Security. Cambridge University Press. <https://doi.org/10.1017/eis.2022.27>

ensorship.⁷ Citizens, especially journalists, increasingly refrain from reporting on high-risk issues due to the digital surveillance, harassment, and intimidation tactics deployed by repressive regimes, with implications to freedom of opinion and the right to freedom of information.

Understanding and documenting these practices, however, remains profoundly challenging. The boundaries between domestic and transnational repression, and between AI-driven and other forms of digital repression, are often blurred. This ambiguity is exacerbated by limited transparency from technology companies, making it difficult to determine when AI systems are involved and how they are being used, as opposed to digital technologies not based on AI. One example is Pegasus spyware used by the Israeli NSO group; while AI is not currently a fundamental component, integration could enhance such technology.⁸ Because of these ambiguous boundaries, we do not have examples of AI use for every section in this paper. This is an exploratory paper that examines various forms of AI-enabled TNR, presenting concrete examples alongside hypothetical situations grounded in existing research.

This paper proceeds in three parts. It examines: (1) AI-enabled surveillance, including biometric surveillance, predictive policing, algorithmic risk assessments and blacklists and smart cities; (2) AI platforms, covering disinformation and harassment campaigns, social media and censorship and AI agents; and (3) a brief policy analysis addressing regulatory gaps.

Throughout this analysis, we maintain an intersectional feminist and decolonial lens that recognises the uneven distribution of harm, highlighting the gendered dimensions of digital TNR. Women⁹ are disproportionately affected, and an intersectional approach is essential to understand how people with overlapping identities, including race and ethnicity, nationality, religion, class, age, disability, gender expression and sexuality,

⁷ "For example, one interviewee who spent 15 years in public activism work in Yemen before leaving for Belgium spoke about how her device was hacked and her wedding photos were leaked (Interview with research participant from Yemen, Citizen Lab, September 2022). She was not wearing a hijab in these photos, and observed how this was considered problematic in the conservative society in which she lived—a form of social shaming that a male human rights defender would not have experienced. Her appearance in these photos led to a series of online attacks, which culminated in the interviewee engaging in self-censorship and eventually reducing her advocacy work. She continues to fear for her safety despite being abroad in Belgium." While not AI specific, this is a good example of self-censorship from *Gender-based digital transnational repression as a global authoritarian practice* by Siena Anstis and Émilie LaFlèche.

⁸ Sen, R., Farooq, N. (2025, December 24). AI-driven Digital Transnational Repression: Past Lessons, Present Challenges, and Future Directions. The Long Reach of the Strong Arm: Evolving Forms of Transnational Authoritarianism. Springer Nature. https://link.springer.com/chapter/10.1007/978-3-032-04940-7_3

⁹ When we use the word "woman" we use a trans-feminine definition, including all people who identify under the box of "woman," and/or non-binary/gender-non-conforming.

experience technology-facilitated gender-based violence (TFGBV) unevenly.¹⁰ Overall, this report analyses AI tools that profoundly impact civil society, journalists and activists globally, facilitating the transnational expansion of authoritarian power through repression beyond borders.

¹⁰ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

II. AI-ENABLED SURVEILLANCE

1. Biometric surveillance

Biometric surveillance, especially facial recognition technology, has become a sophisticated tool for transnational repression, enabling authoritarian governments to extend surveillance far beyond their borders. This technology allows regimes to monitor citizens who have fled to other countries, undermining the safety that exile once provided.

Governments deploy facial recognition systems at critical transit points including airports, border crossings, train stations, and public spaces where cameras can automatically identify individuals. States can gain access to footage through diplomatic channels or security cooperation agreements, or potentially through covert means such as hacking into foreign surveillance networks. **Perhaps most insidiously, the mere knowledge that facial recognition might identify civic space actors creates a powerful deterrent effect.** People can become reluctant to attend protests, community gatherings, cultural events, or speak out publicly, even when living in countries where such freedoms should be protected. This self-censorship effectively extends governmental repression without requiring any direct action, as the technology itself becomes a mechanism of control through fear and intimidation.

Repressive governments may also scan social media platforms, protest footage, and any publicly available images to identify dissidents living abroad. AI-powered facial recognition systems can process vast amounts of visual data, making it possible to identify targets across multiple platforms and locations simultaneously. A single photograph from a demonstration could be matched against comprehensive government databases, potentially exposing someone who believed they were protected by anonymity in a foreign country.

China's repression of Uyghurs is an example of an authoritarian government, which has used several forms of transnational repression to target the Uyghur Region and Uyghurs abroad. Since 2016, Uyghurs living outside of China and their families have been systematically targeted by an extensive and sophisticated transnational repression campaign conducted by the Chinese Communist Party (CCP). The CCP's goal is to threaten the dissenters' family members with detention in an attempt to control their overseas activities.

Host states such as Turkey and Egypt are increasingly cooperative with China's transnational repressive efforts, leading to heightened surveillance, foreign agent recruitment, smear campaigns and physical attacks on Uyghurs abroad.¹¹ For example, Turkey significantly increased funding for facial recognition technology. In May 2025, authorities granted a 5.7 million lira (148,000 USD) contract to supply facial recognition and such equipment to the Istanbul police headquarters.¹²

In Russia, an example of the use of facial recognition technology in urban surveillance is in Moscow, one of the most heavily surveilled cities in the world, where cameras equipped with AI-facial recognition can track individuals in real time, deterring people from participating in political dissent in public spaces.”¹³ Transnationally, other Russian AI technologies are used to enhance surveillance and influence both within national borders and in international digital spaces. “The GRU, while primarily focused on military intelligence gathering abroad, also leverages AI in certain key areas.”¹⁴ These areas include automating the creation and dissemination of disinformation across social media platforms (see more in section below), and using AI to conduct cyberattacks against foreign governments, institutions and infrastructure.

Modern AI-enabled CCTV networks can also automatically track targeted individuals across entire cities or even across international borders when systems are interconnected. These automated systems require minimal human oversight, allowing continuous monitoring of dissidents' movements, associations and activities. Biometric databases, which contain facial scans, fingerprints and other identifying information, can be shared between allied governments through security agreements or informal channels with cooperative governments.¹⁵ INTERPOL notices, sometimes based on politically motivated charges, can facilitate international tracking, too. Finally, this information can be accessed through data breaches or hacking, further extending the reach of surveillance.

¹¹ Human Rights Foundation Staff. (2024, January 15). Beyond Borders: China's Transnational Repression of Uyghurs. Human Rights Foundation. <https://hrf.org/latest/beyond-borders-chinas-transnational-repression-of-uyghurs/>

¹² Gostoli, Y. (2025, June 5). Turkey's AI-Powered Protest Crackdown. New Lines Magazine. <https://newlinesmag.com/spotlight/turkeys-ai-powered-protest-crackdown/>

¹³ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf)

¹⁴ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.p](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.p)

¹⁵ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.p](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.p)

AI and CCTV networks can enable heightened surveillance, predictive profiling, targeting and real-time monitoring. For example, China's AI-assisted surveillance systems cross-reference CCTV footage with digital behaviour to identify and locate Uyghur individuals engaging in specific behaviour patterns the government deems as unusual.¹⁶ AI-enabled network mapping further allows repressive regimes to identify connections between political activists, dissidents and organising for social movements. Overall, AI-enabled DTR through biometric surveillance technologies such as facial recognition and CCTV networks has obscured the accountability mechanisms that traditionally apply to transnational spying and intimidation beyond a state's domestic jurisdiction.

2. Predictive policing and targeting

AI predictive systems enable authoritarian regimes to identify and pre-emptively target potential dissidents before they engage in activism or dissent. By analysing patterns in behaviour, social connections, speech patterns, online activity and even consumer habits, these systems generate “risk scores” that flag individuals as future threats. This shifts repression from reactive to proactive. People can be targeted not for what they have done, but for what algorithms predict they might do.

AI tools can process and manipulate critical international databases and systems that facilitate cross-border movement and cooperation. This includes the use of immigration databases. It also encompasses automated systems that can flag individuals in visa application processes, leading to denied entry or detention upon arrival in countries that share data with repressive regimes. In addition, INTERPOL AI systems may be used to identify individuals for targeting through Red Notices or other alerts, weaponising international law enforcement cooperation for political purposes. Finally, border control systems relying on predictive algorithms can trigger automatic alerts when targeted individuals attempt to cross borders, including in third countries.

The use of predictive AI can create a self-fulfilling prophecy dynamic. **Someone flagged by an algorithm may face increased surveillance, which generates more data about them, which further confirms the algorithm's “prediction” of threat.** Innocent behaviours such as attending certain places, visiting certain websites, or having friends, family and colleagues who are activists, can be misinterpreted by AI as indicators of future dissent. This creates an effect where entire communities become suspect based on

¹⁶ Sen, R., Farooq, N. (2025, December 24). AI-driven Digital Transnational Repression: Past Lessons, Present Challenges, and Future Directions. *The Long Reach of the Strong Arm: Evolving Forms of Transnational Authoritarianism*. Springer Nature. https://link.springer.com/chapter/10.1007/978-3-032-04940-7_3

demographic or behavioural correlations that may have no causal relationship to actual threats.¹⁷

AI-based profiling and location detection tools enable governments to track and intimidate human rights defenders (HRDs) and their diaspora communities across borders. Mapping diaspora networks is one illustrative example of how AI-enabled profiling and location-detection tools may be deployed. **Diaspora network mapping can be used by governments to analyse social media, communication patterns and financial transactions.** For example, AI systems can be used to also monitor speech patterns through natural language processing which could identify individuals expressing dissent or criticism online, even in encrypted or private forums through metadata analysis. Finally, AI-driven diaspora network maps could be used to predict gathering locations by analysing communication patterns and historical behaviour, including where diaspora communities are likely to assemble for protests or cultural events.

3. Algorithmic risk assessments and blacklists

AI-generated blacklists may be shared with foreign law enforcement or used to flag individuals at borders or embassies.¹⁸ **While the stated purpose of governments using AI systems to analyse financial transactions was to detect potential cases of terrorist financing and money laundering, in practice, these predictive analytics tools can be abused to monitor civil society's activities and financial transactions and to identify who supports activists from abroad.**¹⁹ These algorithmic systems can be used to detect financial support through flagging donations or money transfers to opposition groups or activists, even small amounts sent by family members. Such flagging then leads to authorities suspending transactions or even freezing the accounts of involved activists, under the pretext of opening the terrorist financing or money laundering investigation. This further impedes the ability of people to organise and finance dissent.

Beginning in 2025, ECNL has been researching how to support CSOs and activists in exile that face DTR. More specifically, activists or CSOs who were involuntarily relocated could face issues with access to financial services. While at the time of publishing this report

¹⁷ Amnesty International. (2025, February). Automated Racism. Amnesty International UK.

<https://media.amnesty.org.uk/documents/Automated20Racism20Report20-20Amnesty20International20UK20-202025.pdf>

¹⁸ ECNL. (2023, January 3). CT and Tech: Mapping the impact of biometric surveillance and social media platforms on civic space.

<https://ecnl.org/publications/ct-and-tech-mapping-impact-biometric-surveillance-and-social-media-platforms-civic>

¹⁹ ECNL. (2022, November). Mapping the impact of biometric surveillance and social media platforms on civic space.

https://ecnl.org/sites/default/files/2023-03/TECHNOLOGY%20AND%20COUNTER-TERRORISM_NOV%202022.pdf

we have not yet seen any evidence of this, it is likely civil society organisations have been refused to be serviced by banks of the influence from a third country (financial TNR).

In a 2025 report,²⁰ we exposed how civil society groups and individuals have reported increasing challenges in accessing financial services, including difficulties with opening bank accounts, transferring funds, or facing account closures. The report includes a detailed survey mapping out experiences of CSOs and activists with traditional bank institutions as well as with alternative financial service providers, including online payment systems. We provided the key results from the survey for activists, including the main types of challenges they experienced and provided recommendations. These findings can be applicable to TNR, as we uncover more cases of restriction to financial services linked to TNR.

4. Smart cities

One burgeoning trend is the development of “smart cities,” which is generally presented as “the use of technology-based solutions to enhance the quality of life for citizens, improve interaction with government and promote sustainable development.”²¹ The European Commission’s Digital Strategy encourages the development of smart cities as “a place where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and businesses.”²² However, **“efficiency” without proper human rights safeguards can quickly lead to misuse and abuse, especially with the massive troves of data that is collected, processed and shared. Such integrated data-collection infrastructure facilitates extensive surveillance, harassment and targeting of ethnic minorities, including citizens residing abroad.**

A report by the CSO “Unwanted Witness” presents evidence of smart city tools in Rwanda used for internal and transnational repression. In Rwanda, CCTV and biometric systems were approximately 55% of the country, particularly in the city of Kigali. Since 2019, Kigali has been covered by a CCTV network equipped with facial recognition, managed by the Rwanda National Police and the Rwanda Information Society Authority (RISA) since 2019. These cameras cover urban areas, roads and highways, further amplified by telecom controls. These CCTV networks and telecom controls exemplify how cities, with the use

²⁰ ECNL. (2025, July 2). Access to financial services for CSOs on the move: insights and recommendations.

<https://ecnl.org/publications/access-financial-services-csos-move-insights-and-recommendations>

²¹ Sen, R., Farooq, N. (2025, December 24). AI-driven Digital Transnational Repression: Past Lessons, Present Challenges, and Future Directions. The Long Reach of the Strong Arm: Evolving Forms of Transnational Authoritarianism. Springer Nature.

https://link.springer.com/chapter/10.1007/978-3-032-04940-7_3

²² European Commission. (2025, November 5). Smart cities and communities. <https://digital-strategy.ec.europa.eu/en/policies/smart-cities-and-communities>

of AI, can constantly survey citizens with a lack of transparency in data collection, processing and sharing. The normalisation of “smart cities,” combined with the vague definition of what constitutes a “smart city,” raises concerns about transparency and potential misuse for DTR. City governments' overconfidence in technological practices, specifically the widespread use of AI, misleads citizens into believing these tools enhance public safety. However, as in the case of Rwanda, these tools can be used for mass surveillance and extend beyond borders, as reported by exiled dissidents who continue to face digital threats and harassment abroad.²³

²³Kiira, B., et al. (2025, June 6). Surveillance/Spyware: An Impediment to Civil Society, HRDs and Journalists in East & Southern Africa. Unwanted Witness Uganda. <https://www.unwantedwitness.org/wp-content/uploads/2025/06/Report-06.06.2025-FINAL.pdf>

III. AI PLATFORMS

1. Disinformation and harassment campaigns

According to Freedom House, “activists’ reliance on digital platforms and social media creates multiple points of exposure that authoritarian regimes exploit to prepare, deliver and intensify threats across borders.”²⁴ Furthermore, **AI-generated disinformation and harassment campaigns are a widespread tactic of transnational repression, enabling regimes to quickly disseminate false, personalised attacks against activists both at home and in exile.** AI-generated smear campaigns include doxxing and the dissemination of deepfakes, edited, or non-consensual intimate images or videos that seek to shame victims, often on social media.

Deepfakes (AI-generated content, typically video, images, or audio and increasingly multimodal) and AI bots can enable disinformation campaigns that discredit exiled dissidents and activists flood social media with false narratives and smear campaigns, and undermine the credibility of legitimate opposition voices. Deepfakes can be used to impersonate family, friends or activists, and political dissidents to be used against them or their family members. For example, deepfakes can depict staged audio of activists confessing, saying damaging statements, or committing crimes.

These images and videos disproportionately affect women human rights defenders, activists and journalists, because these campaigns, when targeting women, often shame women through comments about traditionally “female” roles such as being a wife, daughter, or mother.²⁵ This is an example of technology-facilitated gender-based violence (TFGBV). TFGBV can also include various forms of abuse and violence, such as gender-based harassment and abuse, rape threats, digitally facilitated stalking, doxing, and sharing of (often fabricated) intimate images and videos.²⁶ The outcome of this violence can often be self-censorship: 40% of women journalists who reported being the target of digital threats or harassment have since stopped reporting certain stories.²⁷

²⁴ Michaelson, M. (2020). The Digital Transnational Repression Toolkit, and Its Silencing Effects. Freedom House.

<https://freedomhouse.org/report/special-report/2020/digital-transnational-repression-toolkit-and-its-silencing-effects>

²⁵ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

²⁶ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

²⁷ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

Automated threats and intimidation can occur through the misuse of AI tools (e.g., large language models or chatbots) to send phishing emails that employ personalised manipulation. This can include automated harassing messages or threats to dissidents and their families. In a 2019 study of 14,678 deepfakes, researchers found that 96% involved deepfake sex videos and 99% of these deepfake videos depicted faces of women inserted without their consent.²⁸

One example of a deepfake video is a video used by the Chinese government in a malicious smear campaign targeting the Founder and Executive Director, Rushan Abbas, of the Campaign for Uyghurs (CFU), a US-based CSO. Through an attempt to discredit her work on exposing China's ongoing genocide of the Uyghur population, the Chinese government spread fabricated allegations and defamation through multiple digital platforms. According to CFU, these allegations are part of a coordinated effort to silence and invalidate Uyghur advocacy. These kinds of smear campaigns are consistent with Beijing's long-running campaign of transnational repression against Uyghurs.²⁹

Disinformation campaigns can also look like AI-generated smear campaigns on social media. An example of a smear campaign occurred when the Chinese government promoted content aimed at discrediting German anthropologist Adrian Zenz, who is known for his research on Xinjiang and China's alleged genocide of the Uyghur population, as reported by Mandiant.³⁰ This report investigated 72 fake news sites and social media posts that are linked to a Chinese PR firm called Shanghai Haixun Technology Co, which sells content creation packages for English-speaking audiences. According to the report, the content promotes reshaping the image of Xinjiang, criticism of the US, and discredits critics of the PRC government.³¹ This smear campaign exemplifies transnational repression efforts because it involves the Chinese government and their targeting and intimidation of a foreign-based researcher beyond China's borders. In addition, the smear campaign used calculated disinformation to undermine a researcher's credibility and to silence criticism in the international public sphere.

According to NBC News, a Chinese marketing firm operated a network of at least 72 fraudulent news websites across eleven languages, accompanied by fabricated social

²⁸ Antis, S., LaFlèche, E. (2025). Gender based transnational repression as a global authoritarian practice. Taylor & Francis Online. <https://www.tandfonline.com/doi/epdf/10.1080/14747731.2024.2401706?needAccess=true>

²⁹ Campaign for Uyghurs. (2025, September 22). CFU Condemns Malicious AI Deepfake Smear Campaign. <https://campaignforuyghurs.org/cfu-condemns-malicious-ai-deepfake-smear-campaign/>

³⁰ Kong, S. (2022, August 22). China ramps up disinformation campaign on Uyghurs in Xinjiang. Fair Planet. <https://www.fairplanet.org/story/china-ramps-up-disinformation-campaign-on-uyghurs-in-xinjiang/>

³¹ Kong, S. (2022, August 22). China ramps up disinformation campaign on Uyghurs in Xinjiang. Fair Planet. <https://www.fairplanet.org/story/china-ramps-up-disinformation-campaign-on-uyghurs-in-xinjiang/>

media personas, to push pro-Chinese government narratives.³² It is likely that AI was employed in the creation of these social media accounts and website imagery, including AI-generated images and/or bots. This hypothesis is supported by prior evidence demonstrating China's use of AI-driven technologies to generate and amplify disinformation.³³ For instance, while monitoring the activities of Spamoouflage, a pro-Chinese government political influence operation that has been repeatedly exposed since 2019, Graphika identified network assets promoting a distinctive form of AI-generated video content across social media platforms such as Facebook, Twitter and YouTube. Notably, some of this content was developed by tools created by Global North AI companies, including Synthesia, a British AI company.³⁴

2. Social media monitoring and censorship

Social media monitoring and content moderation rely on large-scale data collection and processing and can function as forms of censorship when they suppress political speech or dissent. These practices occur when **AI systems scrape vast amounts of data from social media platforms, emails, or messaging applications. Techniques such as sentiment analysis and natural language processing (NLP) can be used to flag dissenting views or opposition activity, raising serious concerns about overreach and the chilling of legitimate expression,** with direct implications for transnational repression.

This can be done through automated content moderation in general, and expanded with the use of LLMs. When deployed by governments, these tools can serve as instruments of repression. **Governments with authoritarian practices have leveraged content moderation and surveillance systems to monitor, target, and silence dissent both internally and abroad.** These kinds of technologies can be weaponised to track journalists, human rights defenders and marginalised communities. Those linked to flagged content may face severe offline harms, including arbitrary detention, torture, or extrajudicial killings. The combination of predictive surveillance and real-world

³² Collier, K. (2022, August 4). Dozens of fake news websites and social media accounts pushed pro-China talking points. NBC News. <https://www.nbcnews.com/tech/tech-news/dozens-fake-news-websites-social-media-accounts-pushed-china-talking-p-rcna41140>

³³ Lee, C.J. ((2025, July 16). China's AI-Powered Disinformation Tactics: Threats and Implications. University of Virginia National Security Data and Policy Institute. https://nspcbatten.org/content/uploads/sites/2/2025/07/C.Lee_White-Paper.pdf

³⁴ Graphika. (2023, February). Deepfake It Till You Make It Pro-Chinese Actors Promote AI-Generated Video Footage of Fictitious People in Online Influence Operation. Graphika Report. <https://public-assets.graphika.com/reports/graphika-report-deepfake-it-till-you-make-it.pdf>

punishment produces a profound chilling effect, suppressing dissent and driving widespread self-censorship both internally and abroad.³⁵

In Egypt, for example, social media monitoring has increasingly been used for political repression and general forms of surveillance, both within its borders and transnationally to monitor and control dissent. One example is when Egypt deployed an AI system to monitor platforms such as Facebook, Twitter, YouTube, WhatsApp, Viber and Instagram. This system scanned these platforms for 26 topics, “ranging from defamation of religion to calls for illegal demonstrations and terrorism, though the full list remains undisclosed. Since then, successive governments have monitored electronic communications, leading to the arrest and prosecution of activists for their social media posts.”³⁶ This monitoring may also encompass activists living abroad, particularly those with Egyptian roots or family ties.

Indeed, Egypt’s AI-driven surveillance has been used transnationally, by targeting Egyptian expatriates and dissidents living abroad. The government has reportedly employed advanced cyber-espionage tools, including AI-enabled surveillance capabilities.³⁷ In October 2019, a report stated that the Egyptian government had conducted cyber espionage activities to target Egyptian dissidents through installing mobile applications on the targets’ phones to extract files, track locations and identify contacts. This report suggests that the Egyptian government has been repressing dissents both within and possibly outside of Egypt’s borders. The victims of these cyber-attacks were Egyptian journalists, academics, lawyers, opposition politicians and human rights activists.³⁸

Furthermore, so-called shadow banning, which is the practice of limiting the visibility or reach of a user’s content, constitutes another harmful tactic of content moderation. Rather than overtly removing content that governments may consider as problematic, such as protest-related content, platforms may suppress it by not recommending or demoting the content, effectively rendering it invisible. These risks are heightened where

³⁵ Wisniak, M. (2025, April). Algorithmic Gatekeepers: The Human Rights Impacts of LLM Content Moderation. III. Right to Privacy ECNL. https://ecnl.org/sites/default/files/2025-04/ECNL_LLM_CM_Privacy_2025.pdf

³⁶ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.p](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.p)

³⁷ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.p](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.p)

³⁸ Akin. ÜNVER, H. (2024, May). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights. European Parliament. Policy Department, Directorate-General for External Policies. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA\(2024\)754450_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf)

such systems misclassify legitimate speech or assembly-related content as violent or inciting harm.³⁹

Content moderation has the capacity to negatively affect human rights through vague and inconsistent guidelines and enforcement of content policies, and lack of transparency or accountability. Vast content moderation with an overly broad reach can remove entire groups from visibility. This occurs when the algorithm defines certain speech as “violent,” suppressing activists’ freedom of assembly and political organising. LLM-based moderation often lacks cultural and linguistic nuance and therefore reflects biases in training data.⁴⁰ For example, automated content moderation systems disproportionately designate Palestine-related content as “terrorist,” due to biases in training datasets and models, among other things, leading to pro-Palestinian voices being silenced.⁴¹

3. AI agents

AI agents pose unique and heightened risks in the context of DTR due to their autonomy, scale, speed and adaptability. **AI agents operate with very little human oversight and are large models which can run simultaneously, learning from responses and evolving their tactics continuously and independently. The greatest risk of AI agents is the lack of human involvement in automated decision-making and lack of accountability for the actions of these agents.** For example, governments could deny involvement by claiming rogue tools or third-party services or claiming prompt injection, a manipulation technique in which inputs are designed to exploit an AI system’s instruction-following behaviour, leading it to override or improperly reinterpret its predefined operational directives.⁴²

Autonomous surveillance agents are one kind of AI agent which can be used specifically for oversight, monitoring and tracking individuals. **Autonomous surveillance agents can monitor online behaviour in real time, flag keywords, connections or location patterns.** These systems could therefore be used to systematically surveil large populations abroad such as diaspora communities, continuously scanning for dissent.

³⁹ Wisniak, M. (2025, April). Algorithmic Gatekeepers: The Human Rights Impacts of LLM Content Moderation, III. Right to Privacy. ECNL. https://ecnl.org/sites/default/files/2025-04/ECNL_LLM_CM_Privacy_2025.pdf

⁴⁰ Wisniak, M. (2025, April). Algorithmic Gatekeepers: The Human Rights Impacts of LLM Content Moderation, V. Right to Freedom of Peaceful Assembly and Association. ECNL. https://ecnl.org/sites/default/files/2025-04/ECNL_LLM_CM_Privacy_2025.pdf

⁴¹ Wisniak, M. (2025, April). Algorithmic Gatekeepers: The Human Rights Impacts of LLM Content Moderation, V. Right to Freedom of Peaceful Assembly and Association. ECNL. https://ecnl.org/sites/default/files/2025-04/ECNL_LLM_CM_Privacy_2025.pdf

⁴² Chan, A., et al. (2024, June 3). Visibility into AI Agents. ACM Digital Library. <https://dl.acm.org/doi/pdf/10.1145/3630106.3658948>

Perhaps most chilling is how **AI agents could assist governments in systematically prioritising targets for harassment, detention, or other severe human rights violations.** These systems could assist to calculate threat scores by algorithmically assigning numerical risk ratings based on online behaviour, associations, speech patterns and predicted future actions. They could also suggest action priorities by recommending which individuals to target first, optimising repression for maximum impact with limited resources. Finally, AI agents could automate harassment campaigns by coordinating multiple vectors of pressure, such as online harassment, family targeting, or employment sabotage, based on predicted **effectiveness.**

Operating 24/7 at a large scale, these agents could also facilitate cross-platform tracking, specifically identity tracking and profiling. Such cross-platform tracking is likely to enable AI agents to engage in profiling aimed at identifying and de-anonymising activists or whistleblowers living in exile. This could occur by correlating identities through large scale data mining across social media platforms such as X, LinkedIn, Telegram and Instagram, using profile images, usernames and linguistic patterns. **These capabilities enable sophisticated network mapping, allowing authorities to construct detailed relationship graphs of diaspora communities, identify key leaders or influencers, and strategically engage with or disrupt these networks to undermine solidarity and collective action.**

IV. Policy overview

TNR has emerged as a G7 priority, led by Canada's agenda-setting and Germany's focus on digital aspects. Yet this heightened attention coincides with accelerating AI deregulation and pausing regulatory efforts globally. As states with authoritarian practices weaponise AI for cross-border surveillance, intimidation and censorship, democratic nations are dismantling the very regulatory frameworks needed to constrain these abuses.

During the June 2025 G7 Kananaskis Summit,⁴³ TNR was positioned as a cybersecurity and sovereignty crisis. For instance, Dutch intelligence agencies (AIVD/NCTV) reported that multiple countries conduct espionage through diaspora communities in the Netherlands using informant networks and cyber tools, with Iran, Pakistan, Morocco and Turkey confirmed, and others likely involved.⁴⁴ It also stated that countries such as Russia, China and Iran use offensive cyber operations for sabotage, economic espionage, and monitoring dissidents living abroad. Furthermore, Germany's leadership reframed TNR as a digital security issue, with the G7 launching a Digital TNR Detection Academy to build capacity against technology-enabled threats.

Policy challenges related to TNR converge around four connected issues. First, cybersecurity concerns include safeguarding democratic institutions and processes from AI-enabled manipulation, disinformation and cyber intrusion. Second, questions of sovereignty arise in relation to protecting territorial integrity from cross-border digital interference. Third, significant risks emerge for human rights, particularly for marginalised communities, political dissidents, and civil society actors operating in constrained civic spaces. Fourth, these dynamics collectively place growing pressure on democratic governance and the rule of law, with increasingly sophisticated forms of surveillance, repression and information control.

Paradoxically, international security conventions designed to combat terrorism can enable TNR. Examples include UN Security Council counter-terrorism resolutions 2396⁴⁵

⁴³ Leaders of G7 Research Group. (2025, June 17). G7 Leaders' Statement on Transnational Repression. <https://www.g7.utoronto.ca/summit/2025kananaskis/250617-transnational-repression.html>

⁴⁴ Algemene Inlichtingen-en Veiligheidsdienst. (2024, October 17). Across the Border: State Interference in Diaspora Communities in the Netherlands. General Intelligence and Security Service Ministry of the Interior and Kingdom Relations. <https://www.aivd.nl/documenten/2024/10/17/over-de-grens-statelijke-inmenging-in-diasporagemeenschappen-in-nederland>

⁴⁵ United Nations. (2017, December 21). Resolution number 2396. United Nations Security Council. <https://www.un.org/securitycouncil/content/sres23962017>

(2017) and 2671⁴⁶ (2021), the Budapest Convention on Cybercrime⁴⁷ (2001), the recently adopted UN Convention against Cybercrime⁴⁸ (2025), and international policing mechanisms such as INTERPOL cooperation frameworks, all of which facilitate cross-border intelligence and law enforcement collaboration. Indeed, **states that ratified the treaties are bound by intelligence-sharing obligations and may face reciprocal requests from repressive governments, which can exploit these channels to access sensitive information, identify targets, map vulnerabilities, and facilitate criminalisation or persecution.** As a result, multilateral security cooperation mechanisms seemingly intended to protect democratic systems can become conduits for authoritarian overreach. At the same time, recent deregulation efforts in both the European Union and the United States raise significant concerns regarding the governance of AI-enabled TNR, or lack thereof.

First, the European Union’s proposed Digital Omnibus package, announced on 19 November 2025, would delay key compliance deadlines under the EU AI Act until 2027–2028.⁴⁹ Although presented as a regulatory simplification and competitiveness measure, the postponement weakens oversight during a period of rapid AI expansion and growing documented misuse of AI systems. The delay particularly affects safeguards governing high-risk AI applications, which include AI systems described in this paper. By extending implementation timelines, the Omnibus package creates a regulatory gap during which such systems can continue to be used with limited accountability.

Moreover, many AI applications, particularly biometric surveillance and algorithmic risk assessments, remain largely unregulated when it comes to national security. Indeed, broad carve-outs and exemptions for national security in existing AI legislation, including [Article 2 of EU AI Act](#) and [Article 3\(2\) CoE Framework Convention on AI and Human Rights, the Rule of Law and Democracy](#), have created significant regulatory gaps. Compounding this issue, “national security” remains an expansive and poorly defined concept. This ambiguity creates two serious risks. First, it enables governments to sidestep safeguards designed to protect fundamental rights when deploying AI for national security purposes. Second, the wide discretion afforded to states in defining

⁴⁶ United Nations. (2021, December 30). Resolution number 2617. United Nations Security Council. <https://documents.un.org/doc/undoc/gen/n21/424/08/pdf/n2142408.pdf?OpenElement>

⁴⁷ Council of Europe. (2024, August 27). Conventions on cybercrime: The Budapest Convention and the draft UN treaty. Council of Europe Cybercrime Division. <https://rm.coe.int/conventions-on-cybercrime-the-budapest-convention-and-the-draft-un-tre/1680b1631a>

⁴⁸ United Nations Office on Drugs and Crime. United Nations Convention against Cybercrime; Strengthening International Cooperation for Combating Certain Crimes Committed by Means of Information and Communications Technology Systems and for the Sharing of Evidence in Electronic Form of Serious Crimes. <https://www.unodc.org/unodc/en/cybercrime/convention/home.html>

⁴⁹ European Commission. (2025, November 19). Digital Omnibus on AI Regulation Proposal. European Commission, Digital Strategy. <https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal>

national security can facilitate the unchecked expansion of surveillance and executive authority. In practice, we worry that national security risks becoming a convenient shield against transparency and accountability, with direct implications for DTR.

Second, parallel developments in the United States raise related concerns. **The 11 December 2025 Executive Order on a National AI Policy Framework⁵⁰ directs the Department of Justice to establish an AI Litigation Task Force tasked with challenging state AI laws deemed inconsistent with federal policy.** The order further authorises federal agencies to identify “onerous” state regulations and conditions eligibility for certain federal broadband and discretionary funding on compliance with federal AI policy determinations. This approach undermines state regulatory initiatives that have emerged as an important lever for AI governance and accountability, including safeguards addressing surveillance, algorithmic discrimination, and AI-enabled manipulation or censorship. Unsurprisingly, the order does not simultaneously introduce a comprehensive federal regulatory framework addressing AI-related human rights risks or TNR-related harms.

Taken together, these developments suggest a growing policy contradiction. **Democratic governments increasingly recognise transnational repression as a serious and escalating threat to human rights, civic space and democratic governance. Yet current regulatory trajectories are weakening oversight mechanisms that could constrain the deployment and cross-border use of AI-enabled surveillance and censorship.** The delayed implementation of the core EU AI Act jeopardises one of the most comprehensive AI regulatory regimes under development—despite its existing loopholes—while the United States’ federal preemption approach risks dismantling state-level accountability mechanisms without establishing comparable national safeguards. **This regulatory divergence creates permissive environments that facilitate the continued evolution and global growth of AI-enabled TNR practices.**

⁵⁰ The White House. (2025, December 11). Presidential Documents, Executive Order 14365. <https://www.govinfo.gov/content/pkg/FR-2025-12-16/pdf/2025-23092.pdf>

V. Conclusion

AI-enabled transnational repression (TNR) represents an emerging and rapidly evolving challenge at the intersection of technology, human rights and international security. While transnational repression itself is increasingly recognised as a serious threat to democratic governance and civic space, the specific role of AI in enabling and accelerating these practices remains insufficiently understood. **The covert, adaptive and cross-jurisdictional nature of both AI and TNR makes it inherently difficult to document and regulate.** Many incidents involve complex state and non-state partnerships as well as intelligence infrastructures that are shielded from public scrutiny. Targeted communities also frequently face significant risks in reporting incidents, further limiting available evidence. As a result, the current knowledge base likely underestimates both the scale and sophistication of AI-enabled TNR.

This report provides an overview of how AI systems can be used for TNR. **AI does not create transnational repression in isolation; rather, it transforms and intensifies existing digital and traditional repression tactics.** AI-driven surveillance, such as biometric identification, predictive analytics, and large-scale data aggregation allow governments to monitor, profile, and target individuals across borders with unprecedented speed, scale and precision. AI-driven disinformation and harassment campaigns are increasingly used to stigmatise activism and repress dissent. **By lowering operational costs and increasing the efficiency of surveillance and targeting, AI alters the risk calculus of repression and facilitates more systematic and scalable forms of cross-border repression.**

At the same time, **the current policy landscape risks enabling rather than constraining AI-driven TNR.** Existing international security cooperation frameworks, including counter-terrorism and cybercrime instruments, have created extensive cross-border data-sharing and intelligence cooperation mechanisms that are already exploited by governments seeking to identify and target individuals abroad. Furthermore, blanket exemptions for critical sectors such as national security and defence from international regulatory frameworks, delays in implementing AI safeguards, fragmented regulatory approaches, and the erosion of accountability mechanisms risk creating permissive environments in which TNR-enabling technologies can proliferate with limited oversight. These trends reflect a broader contradiction: **democratic governments increasingly**

recognise the cybersecurity and democracy threats posed by transnational repression while adopting (or failing to adopt) governance approaches that may inadvertently facilitate it.

Addressing AI-enabled TNR requires a coordinated, forward-looking response grounded in evidence, accountability and participatory governance. **First, there is an urgent need to strengthen research, documentation and data-sharing among civil society, academia, governments and international organisations to better understand the scale, modalities and drivers of AI-enabled TNR.** Evidence-generation efforts must prioritise rights-based, community-centered methodologies that protect the safety of affected groups. Participatory governance mechanisms are particularly critical in this context. Individuals and communities most vulnerable to TNR, including diaspora groups, human rights defenders, journalists, and marginalised groups, must be meaningfully involved in shaping policy responses, risk assessments and AI governance frameworks.

Second, democratic governments should strengthen international coordination specifically aimed at preventing TNR while embedding robust human rights safeguards within intelligence-sharing, cybersecurity and counter-terrorism cooperation frameworks. This includes integrating participatory oversight mechanisms that allow civil society organisations and affected communities to contribute to monitoring, flagging potential abuse and accountability processes. Such inclusive governance models can help ensure that security measures do not inadvertently facilitate repression, while strengthening regulatory institutions.

Third, regulatory and policy responses must prioritise the protection of democratic actors and their support networks broadly, recognising that TNR frequently operates through indirect forms of intimidation targeting families, colleagues and broader community structures. Rights-based AI governance frameworks should incorporate participatory impact assessments, transparency requirements, and accessible remedies for individuals affected by cross-border surveillance and targeting.

Ultimately, safeguarding democratic systems in the age of AI requires treating transnational repression as both a human rights challenge and a core threat to national security. Without deliberate, rights-based and participatory policies, the rapid growth of AI risks strengthening authoritarian influence, enabling repression across borders, and shrinking the safe space for civic actors, journalists and human rights defenders. The trajectory of AI governance will therefore play a decisive role in determining whether emerging technologies reinforce democratic resilience or accelerate democratic erosion.

Yet this trajectory is not predetermined, and democratic governments can actively shape AI governance to protect those most at risk. Doing so is not only a matter of AI regulation, but of preserving civic space, maintaining public trust and defending core democratic values.

AUTOMATING REPRESSION BEYOND BORDERS:

How AI is Powering Transnational Repression



European Center for
Not-for-Profit Law