



CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE



New Digital Dilemmas: Resisting Autocrats, Navigating Geopolitics, Confronting Platforms

Steven Feldstein, editor

*Mahsa Alimardani | Afef Abrougui | Arindrajit Basu | Luca Belli | Agustina Del Campo | Iginio Gagliardone
Brian Kot | Irene Poetranto | Jan Rydzak | Janjira Sombatpoonsiri | H. Akin Unver*

New Digital Dilemmas: Resisting Autocrats, Navigating Geopolitics, Confronting Platforms

Steven Feldstein, editor

*Mahsa Alimardani | Afef Abrougui | Arindrajit Basu | Luca Belli | Agustina Del Campo | Iginio Gagliardone
Brian Kot | Irene Poetranto | Jan Rydzak | Janjira Sombatpoonsiri | H. Akin Unver*

The Carnegie Endowment for International Peace thanks the Charles Stewart Mott Foundation for the support that has made the establishment of the Digital Democracy Network possible. Additional valuable support has come from the Ford Foundation and the Open Society Foundations. The authors alone are responsible for the views expressed.

© 2023 Carnegie Endowment for International Peace. All rights reserved.

Carnegie does not take institutional positions on public policy issues; the views represented herein are those of the author(s) and do not necessarily reflect the views of Carnegie, its staff, or its trustees.

No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the Carnegie Endowment for International Peace. Please direct inquiries to:

Carnegie Endowment for International Peace
Publications Department
1779 Massachusetts Avenue NW
Washington, DC 20036
P: + 1 202 483 7600
F: + 1 202 483 1840
CarnegieEndowment.org

This publication can be downloaded at no cost at CarnegieEndowment.org.

Contents

Introduction Steven Feldstein	1
Aggressive New Digital Repression in Iran in the Era of the Woman, Life, Freedom Uprisings Mahsa Alimardani	5
Weaponization of Identity in MENA: How It Manifests Online and the Responsibilities of Tech Platforms Afef Abrougui	11
Strategizing Pushbacks Against Digital Repression: Insights From Southeast Asia Janjira Sombatpoonsiri	15
Defending the 'S Word': The Language of Digital Sovereignty Can be a Tool of Empowerment Arindrajit Basu	19
A Postcolonial Perspective on Digital Sovereignty Iginio Gagliardone	23
To Get Its AI Foothold, Brazil Needs to Apply the Key AI Sovereignty Enablers (KASE) Luca Belli	27
Volume, Speed, and Accessibility as Autonomous Harms: Can Modern Legal Systems Deal With Harmful but Legal Content? Agustina Del Campo	35

The Stalled Machines of Transparency Reporting	39
Jan Rydzak	
The UN Global Digital Compact Must be Multistakeholder and Inclusive	45
Irene Poetranto	
For AI, Does Democracy or Development Come First?	49
H. Akin Unver	
What Matters More for U.S. and EU Tech Export Controls: Human Rights or Geoeconomics?	53
Brian Kot	
About the Authors	59
Carnegie Endowment for International Peace	61



Introduction

Steven Feldstein

Across the globe, the struggle between rights and repression persists. Digital technology remains at the center of these contests. Governments continue to use censorship strategies, mass surveillance measures, disinformation campaigns, and internet shutdowns to counter political protests, rig elections, and consolidate military coups. The 2023 Freedom on the Net report reflected this, indicating thirteen consecutive years of [global internet freedom declines](#).

Despite these challenges, citizens continue to devise creative ways to use digital tools to mobilize demonstrations, circumvent information controls, and deter electoral manipulation. Although in many countries it is hard to escape the impression that the balance of digital technology weighs in favor of the forces of repression, at least several countries experienced positive changes in 2023. Protesters spurred a change in government in Sri Lanka; soon after, authorities lifted widespread social media and communications blocks. In Gambia, online expression is blossoming as the country gets further away from the two-decade rule of deposed strongman Yahya Jammeh. And in Georgia, online mobilization against a bill that would have forced civil society groups to register as “[agents of foreign influence](#)” prompted tens of thousands of demonstrators to descend on the capital and brought about a decisive legislative defeat of the measure.

The digital rights and policy landscape is increasingly entangled by other issues as well. One area of concern is the fraught relationship between large tech companies on the one hand and human rights and transparency concerns on the other. Platforms such as Facebook, Google, Instagram, TikTok, X (formerly Twitter), WhatsApp, and YouTube hold inordinate power to determine what users see, what content will be amplified, and what will be taken down. Platforms make these decisions largely behind closed doors, providing little insight into their reasoning.

A major inconsistency regards how to handle harmful content—for example, hate speech or harassment—that is protected under the law but can still lead to significant damage (because of the volume, scale, and speed of dissemination). Modern legal systems did not anticipate that volume and speed could or should render speech illegal, and it is unclear whether changing free speech protections based on these factors is sensible. Yet there is little doubt that the viral spreading of falsehoods, propaganda, and harassing speech has deeply wounded the world’s democracies and undermined public trust.

Another area of tension relates to the growing impact of geopolitics on technology. Many experts contend that the world is increasingly fragmenting into contesting spheres of influence. In a new book, Anu Bradford [argues](#) that the globe is separating into three “digital empires” led by the United States, China, and Europe, with “each advancing a competing vision for the global digital economy while attempting to expand its sphere of influence in the digital world.” This dynamic is manifesting in competing regulatory approaches and differing market incentives—threatening to permanently split technological innovation. As new technologies, such as generative AI, establish themselves, these challenges will grow. Foundational AI models have tremendous power to [shape discourse and influence democratic deliberation](#), as well as provide new tools for authoritarians to enhance surveillance and disseminate propaganda. So while global consensus frays, there remains a pressing need for governments to develop common understandings around norms, institutions, and approaches.

This latest collection from the Carnegie Endowment for International Peace’s [Digital Democracy Network](#), building on a [volume of essays](#) by the network in 2021, probes four crosscutting themes: digital repression shifts, the interaction between tech platforms and digital rights, digital sovereignty, and the implications of technology on geopolitics and governance.

First, authors explore how digital repression is shifting in different regions. [Mahsa Alimardani’s](#) article describes how Iranian authorities are “aggressively changing their digital repression strategy” in ways that will alter the next decade of information and communications in the country. She argues that since the outbreak of the Jhina (Mahsa) Amini protests, the state has further adapted and mobilized its repression tactics, such as blocking popular platforms, disabling VPNs, flooding content onto sites to distract users, forcing the release of user data from platforms, and relying on biometric surveillance.

[Afef Abrougui](#) examines how political leaders in the Middle East and North Africa have “weaponized identity narratives” for political gain using online propaganda, misinformation, and conspiracy theories. Abrougui uses Tunisia as a case study to illustrate how the state is deploying racial stereotypes and hate speech against migrants to distract and defuse from its political crisis.

On a more hopeful note, [Janjira Sombatpoonsiri](#) describes how civil society groups in Southeast Asia are organizing to counter government repression. She identifies three emerging strategies: using protests and legal action to pressure governments, engaging in cross-regional knowledge-building activities, and forming alliances with domestic policymakers and cross-border civic networks.

The second theme relates to digital sovereignty. This concept has gained attention in relation to [China's cyber sovereignty push](#), in which its leaders argue that the state has the right to “choose its own internet development path, its own internet management model, and its own public policies on the internet.” Many experts criticize China's stance as a thinly veiled attempt to impose oppressive policies on its citizens and promote authoritarian norms abroad. However, both [Arindrajit Basu](#) and [Iginio Gagliardone](#) offer slightly different perspectives.

Basu contends that digital sovereignty can be a positive force for countries. He details how anxieties about digital sovereignty are rooted in power asymmetries. Instead of thinking about digital sovereignty as a means to promote authoritarianism, it is useful to widen the aperture and consider how the concept can help smaller states assert their own security and economic interests, particularly against powerful transnational corporations and large nation-states. Gagliardone, meanwhile, asserts that digital sovereignty proponents have neglected to consider its implications for postcolonial states. A deeper examination of the contradictions behind the concept of internet sovereignty, viewed through an African lens, offers a counterpoint to China's assertions that a bordered internet would advantage countries in the Global South.

[Luca Belli](#) focuses on AI sovereignty, asserting that because of the transformational nature of AI technology, it is critical for countries in the Global South to retain agency and self-determination over these systems. Belli presents a framework he terms “key sovereignty enablers” (KASE) and lays out eight factors essential for achieving “sustainable and strategically autonomous AI ecosystems:” establishing both algorithmic governance and data governance, attaining computational capacity, ensuring meaningful connectivity, securing reliable electrical power, advancing a digitally literate population, maintaining strong cybersecurity, and passing appropriate regulations.

The third set of essays scrutinizes the interaction between platforms, digital rights, and transparency. [Agustina Del Campo](#) presents a novel argument about how “volume, speed, and accessibility” accelerated by information and communication technology warrants a rethinking of legal principles related to free expression. She observes that under current law, justices must first determine whether speech is defamatory or harmful before factoring in speed, volume, and accessibility into the scale of the damages. But for online speech, where harms may stem from secondary factors (such as the effect of hundreds of accounts spamming “harmful but legal” content), courts and legislatures have been unable to properly characterize or deal with these features.

[Jan Rydzak's](#) article warns that tech companies are becoming less transparent about how they handle requests for information and censorship demands from governments. While there are some positive signs, such as new disclosure requirements in the European Union (EU)'s Digital Services Act, he cautions about a global reversal in platform accountability.

The last set of articles address geopolitical and governance questions linked to technology. [Irene Poetranto](#) analyzes the United Nations' efforts to create a Global Digital Compact (GDC), intended to advance "shared principles for an open, free and secure digital future for all." Poetranto writes that if the GDC adopted a collaborative process, it could build trust among governments and convince fence-sitting states about the value of adopting inclusive and multistakeholder governance approaches, rather than following the state-dominated model promoted by China and Russia.

[Akin Unver](#) lays out a provocative issue: while industrialized nations convene dialogues around ethical AI principles and democratic frameworks, "emerging economies are often caught in a turbulent pathway of rapid modernization and authoritarian misuse." For many developing nations and emerging economies, ethical AI is a "luxury," overtaken by the urgent need to leverage AI for technological, economic, and even political gain. Unver asks: does responsible and democratic AI use stimulate development, or is AI development a precursor, enabling nations to subsequently adapt AI use to democratic norms?

Finally, [Brian Kot](#) interrogates the conventional narrative that U.S. export control policy is centered around geoeconomic considerations while EU policy mostly focuses on protecting human rights. He argues that recent developments in the United States' and the EU's export policies "undermine this dichotomized narrative," such as Europe's curtailment of certain advanced technologies to China for geostrategic reasons and U.S. restrictions against commercial spyware due to human rights abuses.

These diverse global perspectives offer insights about new areas of technological competition and emerging trends. The articles underscore the nuance and complexities of how digital technologies are impacting governance, politics, and society. They are intended to help decisionmakers bridge the gap between local perspectives and global conversations.



Aggressive New Digital Repression in Iran in the Era of the Woman, Life, Freedom Uprisings

Mahsa Alimardani

The tragic death of Jhina (Mahsa) Amini due to beatings while in the custody of Iran’s morality police precipitated a massive popular uprising in the country in September 2022. Protests spread across Iran, encompassing more than 140 cities. Despite ebbs and flows in the intensity of the movement, it continues today—in the form of protests in front of prisons, weekly protests after Friday prayers in Sistan and Baluchestan Province, and other actions.

While domestic protests against the Islamic Republic of Iran are not new, the Mahsa Amini protests—sometimes known as the “woman, life, freedom revolution”—have caused the regime to mobilize two decades worth of investment, development, and planning on information and internet controls. Iranian authorities are aggressively changing their digital repression strategy in ways that will reshape the next decade of information and communications in Iran.

First Generation Digital Repression in Iran

New techniques of digital repression came to the fore in [November 2019](#), before the latest movement, when Iranian authorities imposed a near total internet shutdown for almost a week. This allowed authorities to slaughter hundreds, if not [thousands](#), of protesters in the darkness of an internet blackout. At that time, the regime tried to roll out a national internet network, sometimes known as an “intranet” or the National Information Network, while it shuttered access to the global internet. But it encountered many problems and glitches. In the end, the Internet Society estimated that Iran lost [\\$33 million](#) due to the shutdown.

Ever since, the authorities have sought to finesse their tactics of digital repression to become more efficient, in particular by advancing the earlier mentioned nationalization efforts and attempting to pass the [User Protection Bill](#), or [Tarh-e Sianat](#). Multiple goals underpin these efforts: to build Iran's information and communication technology industry, especially in light of international sanctions; centralize control of data and surveillance to serve state interests; and create a resilient system that can withstand temporary interruptions (such as those caused by protests) or long-term disconnections from the global internet.

While the focus of the nationalization effort had always been to move users onto national and controlled internet services and infrastructure, the User Protection Bill has been a more subtle effort to shepherd these ideas into policy. The bill's main focus is to regulate and move users onto national platforms using techniques such as disabling VPNs and criminalizing VPN sales and usage. Research and monitoring [show](#) that this effort to develop technology to disable VPNs has been long in the making by various government authorities.

Digital Repression in Response to the Jhina (Mahsa) Amini Protests

Since September 2022, the government has implemented several mechanisms to curb protesters' use of the internet. One of these methods is to disable VPNs.

In Iran, where most independent internet services are censored, VPNs are synonymous with access to the internet. Recognizing this, Iranian authorities have long worked to advance their ability to disable VPNs, as [tracked by](#) the international human rights organization ARTICLE19. This has included the [use of deep packet inspection technology](#) and other means to disable circumvention tools. At the start of the protests, authorities added the most used applications in Iran—Instagram and WhatsApp—to the long list of censored platforms. Authorities also blocked the Google and Apple app stores. The Apple app store was eventually reopened, but the Google store continues to be blocked. More than [90 percent](#) of Iranians use Google's Android phones, and the majority use their phones to [access the internet](#). With this crucial app store for Android users blocked, finding secure and functioning VPNs has been a further hurdle to access to the internet.

While the intensity of the protests has diminished since late 2022, Iranians' quest for a stable circumvention tool continues. The top concern for Iranians who are still connected to the internet and one of the main points of research and development for the internet freedom community is to find ways to blunt the sophisticated techniques deployed by the regime to disable circumvention technology.

Mobile Curfews and Regional Internet Shutdowns

One of the hallmarks of Iran’s digital repression during the past year has been the implementation of internet shutdowns, particularly severe disruptions to mobile internet connections.

Since November 2019, authorities have increasingly used mobile shutdowns, especially in response to regional protests that started in 2021 in [provinces](#) such as Khuzestan, Kurdistan, and Sistan and Baluchestan. Mobile shutdowns have become favored methods of the regime because most users, particularly those in impoverished communities, rely on smartphones for connectivity. (In outlying provinces, the government has not invested in landline communications, making home broadband connections rare or nonexistent.)

Authorities have disabled mobile internet connections on most days since the protests began. Curfews have meant that mobile internet providers have been unusable, if not completely throttled, from 4:00 p.m. to 1:00 a.m. on major mobile carriers, such as Irancell, Rightel, and Hamrah Aval (MCI), across the country. In addition, home broadband connections have experienced extreme throttling and disruptions when protests have surged.

In many instances, the cities with the harshest internet shutdowns have also suffered extreme forms of brutality from authorities, such as the cities of Zahedan, Sanandaj, and Saqqez, which are all populated by Iran’s persecuted Kurdish and Baluch ethnic minorities. The greatest concentration of casualties during the protests occurred in Zahedan, which experienced what is now known as the Bloody Friday massacre of September 30, 2022, where at least [100 individuals were killed](#). Protests following Friday prayers have continued in Zahedan; these demonstrations have been accompanied by [severe internet disruptions](#).

Flooding the Digital Space to Prevent Dissent

While access to the internet remains a major concern, Iranians are still finding ways to vocalize dissent online. In response, Iranian authorities are exploring and starting to apply newer methods of digital repression, including tactics that rely on fear and/or flooding. As scholar Margaret Roberts explains in the book *Censored: Distraction and Diversion Inside China’s Great Firewall*, authorities implementing information controls online around the world use “[fear and/or flooding](#)” techniques to intimidate demonstrators, as well as curb and distract dissent both online and offline.

Iranian authorities have pursued traditional repression tactics to stop protests, including killing at [least 500 protesters](#), arresting [over 19,000 individuals](#), and blinding, torturing, and raping individuals. This has often been sufficient to stop people from taking to the streets or to prompt parents to forbid their children from joining protests. However, government officials have augmented these efforts with [digital tactics](#). These have included flooding online spaces with disinformation to distract or break unity in the opposition or using propaganda to induce fear. An example is the posting of videos on Telegram of forced confessions and

torture. This raises concerns about platform accountability and whether companies will take down such content, especially considering companies' revised policies for Russia that ban state propaganda following the Ukraine invasion. At times, overtly harmful content has been removed from Meta's platforms and sometimes even on X, formerly known as Twitter. However, the proliferation of dangerous messaging on Telegram, where Islamic Revolutionary Guard Corps (IRGC) content thrives, has continued unabated.

The regime also floods the information space with false or bad information to stymie or prevent citizen mobilization. One particularly egregious case occurred in January 2023. As the government began carrying out executions of demonstrators, protesters in turn started mobilizing outside of prisons. In January 2023, there was widespread mobilization against the potential executions of Mohammad Mehdi Karami and Mohammad Hosseini, both of whom had delivered forced confessions about the murder of a security force member. People protested almost every day outside Karaj prison, where Karami and Hosseini were held, hoping to prevent their deaths. As news spread about the state's final decision to execute them, one particularly insidious rumor captivated the attention of Iranians: the alleged assassination of Judge Abolqasem Salavati. Salavati is known to most Iranians as the "hanging judge" [because](#) of the hundreds of executions and lengthy prison sentences he has given to political prisoners, human rights activists, and media workers. He is among the most notorious perpetrators of human rights crimes in contemporary Iran.

The [source of the news](#) about Salavati's reported death was an anonymous Twitter account known as "[Jupiter rad](#)" with over 100,000 followers at that time and a reputation for being an anonymous opposition activist. (The account was suspended in June 2023.) The rumor quickly overtook Persian language social media. In the meantime, Iranian officials quietly conducted the two executions with few protests outside the prisons. While attribution is not certain without further data, many believe Jupiter rad is a "cyber" account—that is, a troll account run by the regime pretending to be a member of the opposition that is used to distract the public.

Another flooding tactic by the state has been the hacking of diaspora opposition members' private information. The Telegram account of the [Adl Aali](#) hacking group has leaked stolen data and documents from the accounts of prominent opposition members in the diaspora, including Masih Alinejad, Nazanin Boniadi, and Reza Pahlavi. The group claims to have been able to access and hack all three of these individuals. The nature of the content they have released from the alleged hacks from Masih Alinejad indicate that some of the content and private videos of her were taken from her brother's phone, who was held in an IRGC prison in Iran. While some documents from a brief hack of Alinejad's Gmail account in 2019 have surfaced, the content from her brother's phone has allowed ARTICLE19 to pin attribution to the IRGC. (The report will be published in November 2023). Adl Aali used its Telegram account to disseminate defamatory stories about each of the three people at

the height of their efforts to coalesce opposition efforts into one “unity coalition.” While Telegram has removed some of the criminal content after reports and requests from civil society and victims, the company only acted on removals once the information and content had already spread to the information space.

Surveillance Strategies: Digital Apps and AI

The concept of “shonood”—living under authoritarian governance—is a norm in the everyday lives of Iranians that stretches back decades, if not centuries. The regime continues to develop multifaceted surveillance and monitoring tactics as part of its toolkit of digital repression. Certain elements of the regime’s surveillance systems have been hidden, only exposed through leaks and investigations, such as the [mobile surveillance system of SIAM](#), a web program for remotely manipulating cellular connections made available to the Iranian Communications Regulatory Authority. Leaks from mobile internet service providers reveal that operators such as Ariantel are using SIAM to target users to throttle their internet access as well as monitor them. These efforts are part of a broader series of tactics to identify and arrest protesters.

An alarming development has been government exploitation of private tech platforms, such as the Iranian version of Uber and Uber Eats. Authorities have forced companies to share user data and geolocation information with security agents, leading to the [arrests of protesters and activists](#). The integration of private technology with state repression adds further fuel to the situation.

The full extent of the government’s use of AI and facial recognition remains unknown. Various officials, including in the Ministry of Interior, have [made announcements](#) about using biometric technologies based on AI and facial recognition to identify women and girls who are not abiding by compulsory veiling laws. There have been documented instances of women being notified through text messages of infractions detected by traffic surveillance cameras while they were sitting in their cars. These developments have been exacerbated by the passing of the [“Hijab and Chastity Bill”](#) in September 2023, which further criminalizes women for not wearing hijab or promoting protests against mandatory hijab.

Despite these tactics, defiance remains high, and there is a possibility of a [backfire](#) effect from surveillance. As one contact in Iran said to the author in March 2023, “Women are fearless, and these announcements of AI surveillance policing our hijabs is meant to bring back the pre-Mahsa Amini obedience which we refuse to go back to.”

There are indications that authorities have acquired Chinese technology, such as [Tiandy’s traffic surveillance system](#), to monitor women and protests, raising fears about transnational authoritarian flows of digital repression. However, when it comes to policing women and enforcing obedience to mandatory hijab laws, the most effective routes have been to use fear, penalize women in their workplaces or educational institutions for not following the regime’s

rules, or shutter businesses that do not force their women employees to conform to strict codes of behavior and dress. While Iran's AI and facial recognition surveillance capabilities are improving, the technology's biggest advantage is not its actual capabilities, but rather its ability to intimidate women into obedience.

Conclusion

In response to the Mahsa Amini protests, Iranian authorities have implemented extensive internet and communications controls. These involve methods already known in the country, as well as new methods including tactics of censorship that utilize technologies to block circumvention tools and mobile and regional internet shutdowns. The authorities are also experimenting with technologies intended to stop protests by creating fear and intimidating the populace. These approaches of controlling and manipulating the digital sphere are intended to deter any form of dissent against the forty-four-year brutality of the Islamic Republic. Iranian authorities are not only setting the stage for continued repression of domestic activists, but they are also borrowing from and contributing to global tactics for digital repression.



Weaponization of Identity in MENA: How It Manifests Online and the Responsibilities of Tech Platforms

Afef Abrougui

It is common practice for political leaders in the Middle East and North Africa (MENA) region to use online platforms to weaponize identity for political gain, particularly in times of crisis. Those fighting to gain or maintain political power often mobilize support by deploying identity-driven propaganda and conspiracy theories. This serves to distract publics from substantive issues and discredit critics and opposition challengers. Regrettably, social media platforms have taken few meaningful steps to curb harmful speech and protect users. Instead, companies have trotted out superficial partnerships that have led to minimal change.

Tunisia Exemplifies the Weaponization of Identity on Social Media

Events in Tunisia illustrate the polarizing effect of social media, as well as the adept manner in which politicians have weaponized identity narratives for political gain. Tunisia's current political crisis began with President Kais Saied's power grab in 2021. Originally elected in 2019, Saied faced "[three interrelated crises](#)" by 2021: the fallout from the ongoing COVID-19 pandemic, a deteriorating economy, and political paralysis and heightened polarization. In response, he dissolved the democratically elected parliament and orchestrated a process to change the country's constitution,¹ paving the way for an "[ultra-presidential](#)" system that expanded his powers and weakened legislative and judicial checks.

Meanwhile, Saied's critics including [judges](#), journalists, political opponents, and activists have been subjected by his supporters to [waves](#) of harassment and smear campaigns on social media, mainly Facebook. [Dozens](#) of human rights defenders, journalists, unionists, and Saied's critics were [detained and prosecuted](#). Saied intensified his crackdown on free speech and critical voices by enacting a repressive [cyber-crime decree](#) in 2022 that imposed prison sentences on those accused of propagating alleged fake news.

The economic crisis continued to worsen in 2023 as Saied [failed](#) to put in place basic measures to [mitigate its impacts](#) on the most vulnerable people. Rather than take substantive steps to address Tunisia's economic woes, Saied reverted to scapegoating Black migrants from other African countries.

In February 2023, in a meeting with his national security council (later posted on social media by his office), he [denounced](#) undocumented African migrants, stating they were part of a conspiracy to change the demography of the local population. “The undeclared goal of the successive waves of illegal immigration is to consider Tunisia a purely African country that has no affiliation to the Arab and Islamic nations,” he said.

The statement came amid a wave of racist attacks, online and offline, and unprecedented anti-migrant sentiment against African migrants. The campaign was [started and led](#) by a little-known political party, the Tunisian Nationalist Party (TNP), whose leaders espoused racist tropes similar to those adopted by right-wing groups and politicians in Europe and elsewhere. The party's leaders falsely maintained that African migrants were seeking to colonize Tunisia, kick Tunisians out of their country, and benefit from social aid that should go to the local population.

According to [analysis by France24](#), TNP's campaign and its false statements and videos on social media experienced a surge in popularity following Saied's February 2023 statement. While the party launched its initial social media campaign against African migrants in September 2022 (mainly on Facebook and TikTok), its videos garnered significantly more views and reach following the statement.

On social media, [misleading and false videos](#) purportedly showing African migrants causing traffic chaos, committing acts of violence, or arriving in droves, including in armed vehicles, spread on social media, garnering tens of thousands of views. Racism, particularly against Black Tunisians and African migrants, is not new. However, this campaign was unprecedented in its reach. As a result, migrants faced evictions, detentions, racist attacks, and layoffs. Xenophobia and racism became so severe in Tunisia that in March 2023 Côte d'Ivoire and Mali had to urgently [repatriate](#) “frightened nationals who flocked to their embassies for help.”

A Tactic Not Unique to Tunisia

The weaponization of identity for political gains by Saied, his government, and his supporters came at a time when political, financial, socioeconomic crises, as well as a [water emergency](#) exacerbated by climate change, unfolded in the country. But wielding identity politics is not unique to Tunisia.

Across the MENA region, moments of crisis have precipitated rises in hate speech and sectarianism, with those fighting to gain or maintain political power and their supporters deploying identity-driven propaganda and conspiracy theories to distract the public from substantive issues and discredit critics who challenge their rule. Such tactics have peaked

during events perceived to be threats to the status quo, such as protests, popular uprisings, and elections.

In Lebanon, where [sectarianism](#) is institutionalized and enshrined in the constitution, nonstate actors and political groups do not hesitate to [deploy](#) sectarian-based attacks and hateful campaigns against their opponents and critics, including on social media. Lebanese politicians use nationalist and hateful discourses against Syrian and Palestinian refugees, [scapegoating them](#) for a worsening economic situation. For example, calls by the United Nations in 2022 for Lebanon to halt the forced repatriation of Syrian refugees were [met](#) with hateful campaigns and calls for violence. In July of that year, social media users spread xenophobic messages and disinformation about refugees under a Twitter hashtag, “#انضرا_ _يروسلا_ _حزانلل_ شم” (“our land is not for the displaced Syrian”).

Similarly in Algeria, during the Hirak uprising—a series of popular protests that erupted in February 2019 in opposition to former president Abdelaziz Bouteflika’s intent to run for a fifth term in office—racist attacks and racial slurs against those from the Kabylia region [intensified](#). Online trolls used the “[zouave](#)” slur to attack Kabyles²; associate them with France, Algeria’s former colonial power; and accuse them of being separatists. The aim by Arab nationalists was to weaken the protests by creating a narrative that the Hirak movement was a Franco-Amazigh conspiracy.

Social Media Platforms’ Lack of Meaningful Actions

This year, I led [research](#) for the Samir Kassir Foundation in Lebanon on hate speech in the country and the shortcomings of social media platforms’ responses. The findings of the research, [published in July 2023](#), were disappointing but not surprising. Of the four platforms that we examined—Facebook, TikTok, Twitter, and YouTube—none conducted human rights impact assessments in Lebanon (or elsewhere in the MENA region) to identify how their policy enforcement, or lack thereof, affected users’ human rights, including their right to nondiscrimination.

In light of big tech companies’ mass layoffs of [content moderators](#) (along with Twitter’s complete dismantling of its trust and safety apparatus), the future looks bleak. Civil society groups and human rights defenders in the MENA region have frequently called on tech companies to take responsibility for digital rights and ensure user safety and protection from hateful speech. The region represents a [large market](#) for social media, and adoption will continue to grow as internet penetration spreads. Platforms owned by Meta (Facebook, Instagram, and WhatsApp) are already [widely used](#). Egypt is one of Facebook’s biggest markets, with more than 40 million active users. Meanwhile, Instagram registered [172.4 million users](#) in the region, as of April 2023.

Yet, these companies have taken almost no meaningful actions to address harmful speech and safety concerns. When they have taken steps, they have mostly been superficial. For example, many of the platforms have set up partnerships with civil society groups to flag harmful content to be taken down. (YouTube calls them “[priority flaggers](#),” TikTok calls

them “[safety partners](#),” and Meta calls them “[trusted partners](#).”) But the outcomes are underwhelming, and the required time investment by civil society is significant. A [recent report](#) by Internews on Meta’s trusted partner program found that partners “often wait weeks or months for responses to their reports, even on issues relating to imminent harm” and that participation in the program increased their workload. Human rights defenders and digital rights groups in the MENA region are increasingly frustrated; activists describe these partnerships as “customer service support” for the platforms and “free labor.”

Challenging the Status Quo

The current dynamic where platforms set the rules of engagement, consult with civil society, and purport to listen to their concerns but fail to implement any real changes in how they do business needs to be challenged. How can that be achieved?

Civil society and human rights advocates need to explore legal options to force tech companies to respect and protect human rights, such as through strategic litigation in jurisdictions with more robust human rights protections. Perhaps there is something to be learned from litigation occurring in other regions. For instance, in 2021, Rohingya refugees [sued Meta](#) in the United States and the United Kingdom, alleging that the company’s algorithm amplified hate speech against the Rohingya during a genocidal campaign conducted by the Myanmar military starting in 2017. In December 2022, two Ethiopian citizens [filed suit against Meta](#), alleging that the company promoted hate speech on the platform that contributed to the country’s deadly conflict in the Tigray region.

Funding and prioritizing pre-litigation research to explore opportunities for strategic lawsuits against tech companies that display irresponsible conduct is an essential first step. Civil society members, journalists, human rights defenders, and academics should also explore more multidisciplinary approaches and work with developers and the technical community to investigate the effects of platform technologies and algorithms in the MENA region. Gaining a better understanding about how algorithmic systems rank, recommend, and curate content in languages and dialects spoken in the region is not well understood. Hard data that documents harms stemming from tech companies’ business models can later be used through strategic litigation and media coverage to hold tech firms accountable and pressure them to take more meaningful measures.

Notes

- 1 The new constitution was adopted in July 2022, with a landslide majority (94 percent of voters) approving it. The referendum, however, was marked by a low turnout (30 percent of registered voters participated), as opposition groups called for a boycott and political apathy remained widespread among citizens concerned by a deteriorating economy.
- 2 According to Algerian researchers, the word “zouave” was first used to refer to a group of Algerian men from Kabylia who were hired by France in 1830 during its occupation of Algeria to be part of a light infantry unit of the French army. The term is now used as a racial slur against Kabyles.



Strategizing Pushbacks Against Digital Repression: Insights From Southeast Asia

Janjira Sombatpoonsiri

Digital repression facilitates ruling elites' control over online dissent. Such a trend is intense in Southeast Asia where autocratic governments increasingly rely on information-related laws, cyber troops, surveillance and interception technologies, and internet shutdowns to tighten their grip on power and undermine oppositional civil society. Despite these developments, Southeast Asia's civil society has managed to at least partially counter these techniques using three tactics: protests and legal action, knowledge- and capacity-building activities, and alliances with domestic policymakers and cross-border civic networks.¹ To advance this tactical trifecta, activists should better combine institutional and extra-institutional activism, forge cross-border networks, and muster societal support.

Protests and Legal Action

Digital repression aims to raise the costs of participating in online dissent, eventually deterring it altogether. But under certain circumstances—such as when activists' strategic communication generates widespread public support and [alliance shifts](#) among elites—digital repression may [backfire](#) by provoking a wider segment of society to escalate further protests.

Specifically in Southeast Asia, such protests combine online mobilization with legal action. For instance, when the Thai junta announced in 2015 the “Single Internet Gateway”—akin to China's internet firewall—[Thai civic groups](#) created an online petition and mobilized netizens and e-business groups against the proposal. The campaigns effectively pressured the junta to drop the policy. Similarly, in mid-2022, the Indonesian government proposed to

amend several criminal codes, including the Electronic Information and Transactions Law (EIT Law), [criminalizing](#) online and offline defamation of high-ranking officials, dissemination of so-called online hoaxes, and views against the state ideology. This sparked widespread civic protests, stalled the parliament's revision of the hoaxes and defamation articles under the EIT Law, and prompted civic networks to petition to the [Constitutional Court](#) to review whether the amendment violated basic rights.

In parallel, civic groups sometime engage in legal action against actors perceived as responsible for digital repression. In [2020](#) and [2021](#), Thai human rights defenders filed complaints against the army for its alleged involvement in online smear campaigns against them. Further, in December 2022, [eight Thai activists](#) targeted by Pegasus spyware lodged a civil case against the Israel-based spyware manufacturer NSO Group. These activists also accused the army, police, and Ministry of Digital Economy and Society of spying on them, submitting complaints to an [administrative court](#) in June 2023. The outcome of these legal actions remains to be seen, but the court cases appear to have set the stage for future advocacy.

Knowledge-Building Activities

Creating online databases of documented digital abuses allows activists to report cases to platforms more systematically, facilitating measures for risk mitigation. Across Southeast Asia, networks of digital rights groups have built online databases by collecting information from civic actors affected by digital repression. Indonesia-based [SAFENet](#), for instance, created a helpline for complaints regarding judicial harassment of civic actors through cyber laws, internet shutdowns, and digital attacks on activists. SAFENet analyzes and visualizes this information on its website. This, in turn, serves as a basis for flagging potentially harmful content to online platforms, providing legal and health-related assistance for affected individuals, galvanizing legal action and solidarity campaigns, providing digital capacity training for civil society organization, and publishing reports for policy advocacy.²

Similar approaches occur in the Philippines and Thailand, though they are less holistic than SAFENet. For instance, [Vera Files](#), which initially emerged as a national fact-checker in the Philippines, put together a database on cyber libel allegations against journalists, allegations that threatened freedom of the press. Philippine research institutes, such as WR Numero Research, monitored online harassment of human rights organizations amid former president Rodrigo Duterte's war on drugs. In Thailand, civic organizations worked with data scientists to document and detect cyber troop activities intended to influence public opinion during major political events. During the 2020 protests, one civic group shared its analysis of possible military-backed influence operations on Facebook, culminating with the platform [taking down 185 accounts](#) that displayed coordinated inauthentic behavior.³ Big data analytics such as these are crucial; according to a former Facebook staff member in Thailand, platforms prioritize systematic and overarching analyses rather than individual complaints.⁴

Alliances With Domestic Policymakers and Cross-Border Civic Networks

Civic groups increasingly form alliances with sympathetic legislators and policymaking actors to publicly expose state practices in digital space that undermine civil society freedoms. A case in point is the collaboration between Thai opposition parties and digital rights groups to expose [state-backed online smear campaigns](#) against dissidents in 2020 and 2021. Furthermore, after Apple's exposé of Pegasus attacks in Thailand in late 2020, opposition parliament members took up this matter, raising legislative debates about the previous government's involvement in spyware abuses. Similarly, [Vera Files](#), along with human rights lawyers, pressured lawmakers to refrain from passing a so-called fake news bill, filed in September 2022 by Senator Jinggoy Estrada. These groups joined hearing sessions, criticizing the bill for being prone to abuse due to its vague definition of fake news. They proposed rights-based regulation of false online content rather than the criminalization of its purveyors and advocated strengthening media literacy through civic education. So far, the congress has not passed this bill.

This institutional engagement is important as it increases public awareness of measures to tackle digital repression. Lawmakers' debates on issues such as online influence operations or the government's misuse of spyware imply an institutional acknowledgment that digital repression has been carried out under the auspices of the government. Together with civil society campaigns, such institutional activism helps to advance public debates online and promote awareness about government repression. Because of increasing public attention paid to this subject, political parties are then encouraged to incorporate policies mitigating risks of digital repression into their platforms.

Lastly, collaboration between domestic civic groups and regional and global networks of digital rights activists facilitate policy action. The 2022 parliamentary debate about spyware in Thailand built on the investigation by Canada-based Citizen Lab and the Thai nongovernmental organization iLaw that documented Pegasus attacks against anti-government protesters and academics. Their [reports](#), published in both Thai and English, spurred further policy advocacy and ongoing legal action. Regionally, solidarity networks of digital rights groups, such as those led by SAFENet and [Digital Rights Collaborative](#), are core to cross-country learning, movement building, and resource sharing. When dissidents and human rights defenders are digitally harassed in one country, these networks mobilize support from policymakers in other countries to pressure states to cease their abuses.

Ways Forward

Civic activism to counter digital repression in Southeast Asia remains nascent; it still lacks a strategic compass, coordination across organizations and countries, and broader societal support. To remedy these shortcomings, civic groups need to better link their extra-institutional activism—such as advocacy campaigns, legal action, and knowledge-building activities—with institutional channels, such as parliamentary subcommittees and political parties.

In Southeast Asia, state institutions, including the police and the army, [justify digital repression](#) strategies on dubious national security grounds. Exposing these policies will require empowering legislators and progressive segments within security agencies and related ministries. Civic groups can bolster institutional actors by providing data-driven analysis regarding patterns of abuse and offering policy suggestions. Civil society can also muster public support to push back against digital repression. Governments and security agencies often claim that dissidents threaten public order and national security. Their rhetoric aims to undermine civil society's legitimacy by isolating them from the wider population. In response, civil society organizations can connect their concerns to broader societal issues, such as by highlighting the exorbitant amount of state spending on repressive technologies rather than on the population's well-being.

Notes

- 1 This essay focuses on organized actions, thereby excluding so-called small acts of digital resistance, such as linguistic tricks to evade content blocks and draconian cyber laws and spontaneous counter-narratives netizens post in response to cyber troops' smear campaigns.
- 2 Digital Harassment Against Human Rights Defenders Workshop, March 13–14, 2023, Bangkok, Thailand.
- 3 I keep the name of this group confidential for safety reasons.
- 4 Digital Harassment Against Human Rights Defenders Workshop, March 13–14, 2023, Bangkok, Thailand.



Defending the 'S Word': The Language of Digital Sovereignty Can be a Tool of Empowerment

Arindrajit Basu

In 1999, noted U.S. human rights scholar Louis Henkin delivered an emphatic (and now famous) lecture at Fordham School of Law calling for the erasure of “sovereignty,” or “[that S word](#),” from international political and legal discourse. He argued that sovereignty had been used illegitimately to protect state control at the expense of human rights protection and enforcement.

“Digital sovereignty” today has received similar bashing—castigated as an excuse for state control of the digital sphere and an impediment to a genuinely free, open, and interoperable internet. Both Henkin and today’s critics of the “S word” are partially right. States often abuse the language of sovereignty to evade their responsibilities under international law and justify patterns of digital repression.

However, it is important to remember that sovereignty has also been an instrument for voicing the interests of the powerless, especially in postcolonial times. The Janus-faced use of sovereignty stems from the fact that it is not a neatly defined, stable doctrine or legal principle. As much as governments exploit digital sovereignty to justify repressive practices, less powerful states and citizens also use the term to challenge existing political, economic, and social power asymmetries. At its core, sovereignty is an “[essentially contested concept](#).” Its malleability is a key reason for its durability in global political discourse, notwithstanding claims of its demise or calls for its abandonment. Rather than viewing the concept of sovereignty as a coherent doctrine or a desirable state of affairs in its own right, it is instructive to critically evaluate the language of sovereignty within the context of its use. In particular, are states invoking sovereignty to alleviate or entrench existing power asymmetries? Answering this question can clarify thinking about digital sovereignty and how it is deployed in heavily contested geopolitical contexts today.

The Use and Abuse of “That S Word”

Western states [exploited ideas of sovereignty](#) and statehood to create a paradigm where the colonizers had rights to sovereignty but colonized Indigenous communities did not, owing to differences in culture and mechanisms of governance. The colonizers weaponized the doctrine across the world to justify their brazen appropriation of land, property, and traditional knowledge systems which were replaced by colonial systems of government. After World War II, as countries began to decolonize, they started to assert their rights over their territory in the international sphere through the language of sovereignty, equality, and autonomy.

The 1962 General Assembly debate that culminated in [Resolution 1803](#) on the “Permanent Sovereignty over Natural Resources” marked the zenith of solidarity among the developing world around the concept of sovereignty. Yet, the developed world manipulated the tenets of international law to preserve their rights of investment stemming from precolonial times. The plethora of global rules and regulations brokered at international institutions and influenced strongly by global corporations, dubbed the “[Transnational Capitalist Class](#),” resulted in further power asymmetries that compromised the autonomy of postcolonial states.

To use the words of Ghanaian leader Kwame Nkrumah, this dynamic abetted [neocolonialism](#), where a state “is, in theory, independent and has all the outward trappings of international sovereignty. In reality, [the state’s] economic system and thus its political policy is directed from outside.” In essence, global cooperation and capital flows entrenched the power of colonial states and their companies at the expense of decolonized countries that were forced to accept these trade-offs.

Digital Sovereignty and Internet Freedom

The rise of the internet as a global popular phenomenon coincided with the emergence of the [unipolar geopolitical moment](#) of the 1990s, based on liberalism, free trade and information exchange, limited government intervention, and democratic American ideals.

Cyber exceptionalists, such as John Perry Barlow, [defended](#) the concept of an independent cyberspace—free from sovereign control and exercised over analogue spaces and territory. Western governments, particularly the United States, did not disagree with this position; they viewed a decentralized internet as an ideal tool for promoting American ideals and for “[dominating the airwaves as Great Britain once ruled the seas.](#)” Apprehensions about [cyber imperialism](#) did little to temper the liberal euphoria and techno-optimism of the moment. The laissez-faire approach to internet governance birthed corporate technology behemoths that offered disruption, innovation, and profits. Concerns that have taken center stage today, such as the abuse of market dominant positions, the rampant and unregulated spread of disinformation and misinformation, or the unrestrained appropriation and exploitation of data for corporate purposes, failed to initially provoke political or policy responses.

The optimistic ideals of an Americanized internet started unravelling as new geopolitical power centers and normative systems began to emerge. China's Great Firewall, which restricts content from foreign websites in China's information space, marked a key departure from the cyber exceptionalist and internet freedom narratives. China, along with Russia, has been one of the [foremost proponents](#) of cyber sovereignty discourse, which envisages state control over all internet activity and infrastructure within a country's territory. While this idea has generally been dismissed as an authoritarian ploy to further information controls and suppress dissent, a number of democracies have started to use the narrative of digital sovereignty to advance their interests. European states, for example, have introduced their own [digital sovereignty](#) narratives to justify imposing regulatory controls over data, infrastructure, and companies within their jurisdictions. European leaders have stressed that technological or digital sovereignty is needed to protect the rights and security of their citizens. Similarly, emerging economies such as India have used [digital sovereignty narratives](#) to justify a range of domestic restrictions related to cross-border flows and content moderation requirements. India even banned Chinese applications, blaming alleged security risks they pose to Indian citizens. In addition, India has justified assertion of sovereignty over citizens' data and digital infrastructure as the only antidote to of [digital colonialism](#)—described as the expropriation of Indian citizens' data by Western companies.

Sovereignty is also being reimagined, reconfigured, and reclaimed by communities whose autonomy and agency over their own data has been denied. Groups such as the [Māori Indigenous people](#) in New Zealand and the [Homeless Workers Movement](#) in Brazil employ such rhetoric to reassert control over their personal data and enable civil society to circumvent unfavorable political power structures.

Power Asymmetries in the Digital Economy

It is worth emphasizing that an internet unencumbered by the language of the “S word” is not free of power structures and instruments of control. As with any human-designed systems, power asymmetries and hierarchies have emerged in digital spaces. First, these power asymmetries occur between states. The United States still plays a talismanic role in the governance and technical preservation of the internet. The largest tech companies are headquartered in the country and much of the world's [data is stored](#) in 2,701 data centers there. (Germany is a distant second, with 487 data centers.) The United States and China are [battling for control](#) over underwater cables that transmit data, and they are fighting to shape the technical standards that will govern the internet. Both countries conduct online [extra-territorial surveillance](#) against citizens of other countries using data transmitted through internet infrastructure they control; the primary constraint on U.S. surveillance efforts has [come from the](#) European Court of Justice.

Second, there are power imbalances between technology companies and their users. By exploiting weak data protection regimes, companies have been able to manipulate, use, and sell data, including personal data, for profit. [Surveillance capitalism](#), as Shoshana Zuboff

lays out, treats individuals and communities as commodities, undermining human dignity and autonomy.

Third, there are major asymmetries between large and small companies. Competition laws have yet to confront the range of abuses undertaken by large companies in the digital economy. Regulators, especially in the European Union (EU), are waking up to this reality. The EU's Digital Services Act and Digital Markets Act offer promising options for preserving fair and contestable markets.

Finally, there is a significant asymmetry between states and their own people. Internet shutdowns, unjustified online censorship, and unbridled surveillance all exploit this imbalance. A functioning judicial system and constitutional levers of accountability are essential for individuals and communities to withstand and challenge such violations.

Resisting Power Asymmetries in the Digital Economy

If the digital “S word” is removed from public discourse, these power asymmetries will remain in place. An unregulated internet will reinforce the concentration of power in a handful of states and large companies that have the resources to control internet infrastructure. Sovereignty, therefore, can serve as a tool for smaller states and communities to assert their security and economic interests. Using digital sovereignty in this manner turns decades of sovereignty discourse on its head. Communities can reimagine the concept of sovereignty as a means to legitimize and underscore their claims against more powerful states and companies.

The ideal of sovereign equality has been and can be championed by states as the basis for an equal say in the governance and development of the internet. New thinking around digital sovereignty that speaks to the origins and abuses of power in the digital economy and identifies ways to counter it is required. However, it is also important to be wary of the potential weaponization of sovereignty by powerful states seeking to advance digital repression. Censorship, [internet shutdowns](#), or [mass surveillance](#) conducted by any state against its citizens undermines human dignity and autonomy. Justifying such actions through the language of sovereignty recalls similar abuses exerted by Western colonizers in previous eras. This is why the language of human rights must coexist with digital sovereignty discourse.

Much of the focus on countering hegemony in the digital world has focused on democracy and rights. That is an important but not the only piece of the geopolitical puzzle. It is essential to add power asymmetries to the mix. States and other actors will use all the tools at their disposal—military capabilities, economic heft, and convenient interpretations of law and language—to further their interests. Rather than retire the digital “S word,” it is preferable to encourage its coexistence and inclusion with human rights discourse to speak against power, rather than entrench it.



A Postcolonial Perspective on Digital Sovereignty

Iginio Gagliardone

Proponents of a free and open global internet fear its possible fragmentation. To date, experts and policymakers have largely viewed internet sovereignty through the frame of a deepening divide between democracies and autocracies. They have also highlighted the Chinese government's overt moves to theorize and promote the concept of a bordered and sovereign internet.

The promise of internet sovereignty indeed has appeal for authoritarian regimes that have struggled to control journalism, radio, and television in the digital sphere. But this perspective has largely neglected to scrutinize digital sovereignty's implications for postcolonial states. A deeper examination of the contradictions behind the concept of internet sovereignty, viewed through an African lens, offers a counterpoint to China's assertions that a bordered internet would advantage countries in the Global South.

The Current Approach to Digital Sovereignty in Africa

A few countries in Africa have sought to incorporate digital sovereignty premises in formal legislation or concrete policy tools. The Ethiopian government in 2020 passed a law introducing harsh measures against hate speech and disinformation that could threaten "social harmony, political stability, national unity, human dignity, diversity and equality." A year later, after Facebook deactivated some pro-government accounts, Ethiopia's Information Network Security Agency [declared](#) that it would begin developing its own social media platform to rival Facebook, Twitter, and WhatsApp.

In most cases, however, African governments—especially those led by long-standing authoritarian rulers, such as Paul Biya in Cameroon and Yoweri Museveni in Uganda—have [asserted](#) their sovereign rights over digital communication in a cruder fashion, largely through [internet shutdowns](#). Authorities have tended to invoke the need to guarantee security and peace to justify network disruptions.

While these governments have not mentioned China as their inspiration, [Beijing's doctrine](#) that “concerns about Internet security of different countries should be fully respected,” no matter what these concerns are, seems to have gained traction. Internet control measures once considered extreme and disproportionate have progressively become the new normal: in the early days of the internet, countries censoring online communication tended to [deny](#) doing so, or sought to present disruptions as technical failures for fear of becoming the target of international condemnation. Now, the same countries are implementing far more abusive and disruptive measures.

The idea of digital sovereignty is indeed attractive for many governments. As David Fischer [suggests](#), “the fact that digital sovereignty is simultaneously vague and elusive and at the same time charged with idealist imaginaries of political autonomy repurposes the theoretical term into a vessel that governments and other actors attempt to fill with meaning according to their needs.”

One meaning, however, that has so far been absent from the debate on digital sovereignty is its implications for postcolonial states. Or, put differently, what is the place of digital sovereignty in the long history of liberation from colonial rule and other forms of external interference in the politics and economics of postcolonial states?

A Postcolonial View

China has pitched the idea of internet sovereignty as particularly appealing for countries in the Global South, presenting it as a response to U.S. hegemony and interference and as a way to counter imbalances in institutions mandated to shape and regulate the internet. In developing the concept of internet sovereignty, China has paradoxically appropriated and accepted a Westphalian definition of the state—one forged in the history of wars between European powers—even while Chinese authorities present digital sovereignty as a form of resistance to Western influence. For China, the clinging on the concept of sovereignty can be interpreted, in realist terms, as opportunistic and self-interested and a way to exploit widely accepted principles on the international stage, such as territoriality and noninterference, to affirm its own right to shape a distinctive model of the internet within its borders.

The same reliance and attachment to sovereignty takes on different connotations in Africa. Its borders were drawn by [colonial powers](#), which violated demographic, ethnographic, and topographic factors, mostly to prevent conflicts that could have erupted between European countries scrambling for new territories. If colonial borders were kept, rather than being

redrawn, by liberation movements in the 1960s, it was to avoid plunging the continent into war immediately after independence. As then Nigerian prime minister Abubakar Tafawa Balewa [remarked](#) during one of the first conferences convening the leaders of African independent states in 1963:

On the problem of boundaries, our view is that although in the past some of these were created artificially by the European powers, which even went so far as to split some communities into three parts, each administered by a different colonial power, nevertheless those boundaries should be respected and, in the interest of peace, must remain the recognised boundaries . . . Any attempt, on the part of any African country to disregard this fact might bring trouble to this continent.

Since Balewa's speech, two different conceptions of sovereignty have coexisted and competed on the continent. One, Pan-Africanism, sought not only to capitalize on the peaceful coexistence among newly created independent states but also to extend the spirit that informed liberation struggles to chart new avenues for a more united Africa. In this view, replicating the sovereign conception of the nation-state enforced by colonial powers would amount to a form of regression, and only attempts to create a multinational polity could be interpreted as [real progress](#).

The other conception, which prevailed in the end, was about strengthening borders rather than making them more fluid. As Wits University's Achille Mbembe critically described it, what replaced the initial commitment to avoid conflict among liberated states and thus maintain colonial boundaries was a "fetishization" of the concept of the nation-state. To justify controls and restraints on the mobility of people and information, African governments "borrowed concepts from the Western lexicon such as 'national interest,' 'risks,' 'threats' or 'national security' [which] refer to a philosophy of movement and a philosophy of space entirely predicated on the existence of an enemy in a world of hostility," and in so doing disregarded "our own long held traditions of flexible, [networked sovereignty](#)."

Alternatives to China's Model

Some African innovators and intellectuals saw digital media as an opportunity to bring communities split by colonial borders closer together. They saw the potential for online platforms to allow for the re-creation and reinterpretation of modes of being and communication that had characterized precolonial Africa. As Mbembe wrote:

Precolonial Africa might not have been a borderless world. But where they existed borders were always porous and permeable. . . . Networks, flows and crossroads were more important than borders. What mattered the most was the extent to which flows intersected with other flows.

At a concrete level, in September 2006, Celtel International, the telecommunications company founded by Sudanese engineer Mo Ibrahim, launched an initiative called One Network, which sought to harness the power of some of these flows. Establishing the world's first borderless mobile network, Celtel's scheme allowed users in Kenya, Tanzania, and Uganda to make transboundary calls at [local rates](#), free of roaming charges. Defying stereotypes of Africa as condemned to catch up with or become a late adopter of technical and normative innovations, African citizens were the first to experience the power of mobile technology to redraw boundaries and encompass preexisting or imagined communities. (In comparison, it took another ten years for the European Union to [end roaming charges](#) among member states.) And yet, despite its symbolic potential, One Network failed to inspire a critical mass of African leaders and entrepreneurs to experiment with similar forms of borderless communication.

Conclusion

Imageries of networked sovereignty, to date, appear more aspirational than practical when it comes to concretely uniting African communities across borders. However, critical reflections and contestations of the idea of sovereignty, as rooted in the history of colonialism and liberation in Africa, should not be excluded from debates on the future of the internet. A deeper appreciation of the contradictions behind the concept of internet sovereignty, examined through an African lens, offers a counterpoint to Chinese claims that a bordered internet would benefit the Global South.



To Get Its AI Foothold, Brazil Needs to Apply the Key AI Sovereignty Enablers (KASE)

Luca Belli

As a [transformational technology](#), AI will have a global impact and considerable ramifications for national economies, democracies, and societies. Although many countries are [developing](#) AI governance frameworks, the regulation of AI is only one of the essential elements that countries need to consider to achieve AI sovereignty.

AI sovereignty is not a universally defined concept. Building upon previous work on digital sovereignty and, particularly, what I have [described](#) as “good digital sovereignty,” I here define AI sovereignty as the capacity of a given country to understand, develop, and regulate AI systems. As such, AI sovereignty is essential to retaining control, agency, and self-determination over AI systems.¹ Importantly, AI sovereignty will likely become an increasingly relevant and strategic topic as AI technologies continue to develop, are adopted, and acquire a significant role in various aspects of society and democratic governance, beyond the digital economy. The impact of AI advancement [includes](#) a wide range of critical sectors such as defense, infrastructure management, healthcare, and law and justice.

A layered framework is needed to analyze which elements are essential to establish a country’s AI sovereignty. These elements make up what I call key AI sovereignty enablers (KASE) and include sound governance for data and personal data as well as algorithmic governance, strong computational capacity, meaningful connectivity, reliable electrical power, a digitally literate population, solid cybersecurity, and an appropriate regulatory framework.

I argue that sound governance,² regulation, research, and development in all the elements of the AI value chain are essential not only to achieving economic growth, social justice, and industrial leadership, but also to asserting AI sovereignty and avoiding overreliance on

exclusively foreign AI systems in a country, which would likely transform the recipient country into a digital colony. Importantly, the purpose of this article is not to advocate for AI autarchy, that is, fostering fully self-sufficient national AI industries, nor to deny the ample benefits that digital trade and cooperation can produce, but rather to discuss how countries could achieve a sufficient level of strategic autonomy, which entails grasping the functioning of AI systems, developing such systems rather than being mere consumers, and regulating them effectively.

Through careful consideration of each of the KASE and their interconnection, countries can build what I call an “AI sovereignty stack.” In this governance framework, the authorities in charge of overseeing each KASE should be enabled to cooperate with other authorities from different sectors (including in consumer protection, data privacy, financial services, energy, and telecom infrastructure) in order to facilitate smooth organization and information sharing. This layered structure may reduce the country’s exposure to the technological choices of foreign private or public actors and simultaneously increase its agency and self-determination over and through AI systems.

For countries in the Global South, AI sovereignty should be a policy priority. The KASE require considerable planning, resources, and implementation capacity, but they are highly strategic objectives that reinforce national sovereignty and empower countries to resist possible adverse conditions spanning from the extraterritorial effects of foreign regulation to the imposition of foreign sanctions and the increasingly frequent disruption of supply chains.

It is important to acknowledge that not every country may be interested or need a fully self-sufficient national AI industry. In the case of Brazil, the proposed KASE framework will illuminate whether Brazilian policy choices and governance arrangements can allow the country to assert AI sovereignty or will lead to AI dependency.

Exploring the KASE of Brazil

Data Governance

Data make up the lifeblood of AI systems. Access to diverse, high-quality data is essential for training and improving AI models. Importantly, depending on the type of AI at stake, the data fed into AI systems can be personal, governmental, confidential, and/or copyrighted, among other data types. The multitude of data types introduces a fair amount of complexity and the need for regulatory compliance in the processing of this information. Hence, developing AI capabilities and sovereignty requires both the availability of large volumes of heterogeneous data and control over such data, including rules governing how they are collected, stored, processed, and transferred to third countries.

Countries with large and diverse populations, as well as consolidated data collection practices and well-structured data policies, will indubitably have a competitive advantage in securing their AI sovereignty. However, few countries enjoy this privilege. So, countries should consider establishing shared data policy frameworks at the regional level or within existing international governance mechanisms, so that national data assets can be shared under agreed norms. ([Latin America could craft its own regional framework](#), learning useful lessons from existing mechanisms, such as the Council of Europe’s Convention 108 and the Malabo Convention.) This strategy would allow the usage of much larger and diversified data pools, providing juridical certainty for AI researchers and developers while protecting the rights of personal data subjects, defending intellectual property rights, and preserving the public interest at the same time.

Particularly, sound data governance allows a country to protect its citizens’ data privacy, ensure national and informational security, and harness the value of data for national development. Brazil made considerable progress in data governance by structuring one of the most progressive and refined [open data policies](#) and by adopting a last-generation data protection framework, the [Lei Geral de Proteção de Dados](#) (LGPD). The [enforcement](#) of the LGPD, however, remains very embryonic, [especially regarding new generative AI](#) systems.

Furthermore, personal data collection is concentrated in the hands of a few foreign tech giants, primarily as a result of so-called [zero-rating](#) mobile internet plans, in which data usage for a few applications selected by the mobile internet operator—typically, dominant social media companies—does not count toward the users’ total data consumption. Thus, the government is unable to harness personal data as a national asset. Lastly, data security also remains [very patchy](#), given the lack of a cybersecurity law and regulation on personal data security.

Algorithmic Governance

Software algorithms are the foundation of AI systems, enabling machines to perform tasks and make decisions. Importantly, algorithms can be both the subject and facilitator of regulation. On the one hand, the development and deployment of algorithms can at least partly give rise to risks and social problems, triggering the need for regulatory intervention. On the other hand, algorithms can support the regulatory intervention itself, as they are increasingly useful in the elaboration and implementation of regulation.

Algorithm development, deployment, and regulation are all equally important dimensions of algorithmic governance. Developing and owning proprietary software provides a considerable competitive advantage and allows a country to embed its normative values within the software. Investing in research and development of AI algorithms, while also addressing the potential risks that they pose, can enormously enhance a country’s technological capabilities and reinforce AI sovereignty.

Hence, the promotion of multistakeholder cooperation to develop software algorithms can enhance AI sovereignty either when domestic players are stimulated to develop proprietary software or when software is developed open-source through a collaborative process embraced—or even led—by national stakeholders. President Luiz Inácio Lula da Silva’s first administration was a true pioneer in [employing](#) a collective approach to digital sovereignty, having promoted free and open software as a strategic objective for national development as early as 2003. Such a policy not only enhanced Brazil’s strategic autonomy from foreign software producers but also increased national understanding and development of software. Unfortunately, this policy was reversed by president Michel Temer’s administration in 2016, de facto bringing about the so-called platformization of the country’s public administration, relying primarily on foreign software providers.

Despite the political turbulence, over the past two decades, Brazil has developed several industrial policy instruments aimed at fostering the national software industry. However, the software development sector has not thrived as much as it could, primarily due to inconsistent policies and an absence of regulations focused on stimulating organic software development and implementation, including a lack of capital to jump-start the industry. Particularly, Brazilian software policies have lacked complementary instruments to stimulate supply and demand, especially compared to countries like China, where public procurements of nationally developed software are common; India, where digital public infrastructure has been established through the [India Stack](#); or South Korea, where in the late 1990s capacity building efforts were organized to foster demand.

Computational Capacity

Training complex AI models and processing large datasets require substantial computational resources. Particularly, the most advanced AI systems, such as generative AI, can be remarkably computer-intensive due to their increased complexity. Ensuring continuous access to sufficient computational capacity should be seen as a key strategic priority.

The availability of high-performance computing infrastructure depends on access to multiple factors, spanning from semiconductors—including chips specifically designed for AI applications as well as latest-generation graphics processing units—to specialized servers tailored to AI specificities in data centers. In this respect, it is interesting to note that some of the first policies adopted by the current Lula administration have been the reintroduction of the [national support program for the development of semiconductors](#) (known as “PADIS”) as well as [the suspension of the decision](#) from former president Jair Bolsonaro’s administration to sell the National Center for Advanced Electronic Technology (Ceitec), which is the only semiconductor producer in Latin America.

The availability of cloud computing resources by itself is not enough to assert AI sovereignty; cloud providers have to also be fully compliant with national legislations. A telling example

is the online education platforms that [operate](#) in Brazil. Two major U.S. tech companies supply these platforms nationally, [but neither company even mentions whether they have been complying with the Brazilian LGPD](#), despite the law being fully in force since 2021.

Meaningful Connectivity

Meaningful connectivity—allowing users to enjoy reliable, well-performing, universally accessible internet infrastructure for an affordable price—plays an instrumental role for AI systems to function optimally and be accessible to a wide population. Seamless connectivity facilitates data exchange, collaboration, and access to cloud-based AI services. It enables real-time applications and supports the development and deployment of AI technologies across various sectors, contributing to the construction of a country’s AI sovereignty.

Over the past ten years, Brazil has made [enormous progress](#) in promoting internet penetration. The cost of connectivity has considerably declined while the connected population has doubled in a decade. Yet, such a rosy picture belies less visible digital divides in the quality of internet access. Most of the internet-connected Brazilian population is de facto only partially connected due to low-quality access.

In fact, more than 70 percent of the Brazilian connected population, and around 85 percent of the lower income population, has access only to a [reduced set of apps](#) included in so-called zero-rating plans. As such, user attention and data collection are concentrated in a remarkably limited number of services—typically dominant social media platforms—making it particularly challenging for any other business to develop complete sets of personal data that can be used to train AI models.

Reliable Electrical Power

As AI systems grow in relevance and size, they require a stable and robust [supply](#) of electrical power to operate effectively. Ensuring reliable power infrastructure and access to affordable electricity is necessary for maintaining uninterrupted AI operations. In this regard, Brazil is probably one of the best-placed countries to support the expansion of AI infrastructure: it is not only energy independent but also in recent years has reached [approximately 85 percent of its annual energy production](#) via renewables, [especially hydropower](#).

However, the national power grid is not without criticism. In the short term, Brazil’s energy supply is relatively secure thanks to the complementarity of various energy sources to hydropower, but the lack of structural planning and the possibility of adverse effects from hydrology—which has been observed in recent years—can considerably increase the cost of energy. Hence, despite having developed a strong power infrastructure, Brazil’s capability to support the deployment of power-hungry technologies requires stronger contingency planning to prevent potential dependencies on external sources.

Digitally Literate Population

Enhancing the digital literacy of the population through capacity building, training, and multigenerational education is [essential](#) not only to achieving a skilled AI workforce but also to fostering cybersecurity and, ultimately, national sovereignty. Investing in AI education, research, and development helps nurture a pool of talented AI professionals, while spreading an understanding of how to make the best use of technology. A sound education strategy is therefore vital to upskill the national population from passive consumers of digital technology to prosumers that can develop technology and innovate.

A robust talent pipeline of AI researchers, engineers, and data scientists enables a country to develop and maintain its AI capabilities, increasing its ability to export technology and reducing its likelihood of becoming a digital colony. It is highly promising that the recently elected federal government in Brazil has already adopted a new [National Policy for Digital Education](#).

However, digital literacy is only a priority for new generations of students, ignoring the fact that virtually no one in Brazil—as in most other countries—has received this type of education, making the majority of the population digitally illiterate. Such a situation is particularly risky in the context of accelerated digital transformation and automatization, where it is necessary for all individuals to understand the functioning of technology, especially those whose labor, social, and economic conditions are likely to be affected by the deployment of AI systems.

Strong Cybersecurity

AI systems are susceptible to cybersecurity threats and can be used to perpetrate cyber attacks, and AI critical infrastructure can come under attack. Brazil has recently enacted personal data protection laws as well as a [considerable number](#) of sectoral cybersecurity regulations, spanning the telecom sector, the banking sector, and the electricity sector. While such progress has allowed the country to [climb](#) the International Telecommunication Union's Global Cybersecurity Index, this positive advancement must be considered again with a grain of salt.

Indeed, Brazil still lacks a cybersecurity law and a national cybersecurity agency, although both have been recently proposed by a study produced by the [Center for Technology and Society](#) at Fundação Getulio Vargas and by a [draft bill](#) formulated by the Brazilian presidency. The existence of a highly fragmented approach to cybersecurity—driven by the initiatives of sectoral agencies with no general competence in cybersecurity and frustrated by the lack of a coherent national strategy—represents a big vulnerability. Brazil has not yet managed to create a solid governance framework to connect, coordinate, and leverage the incredible amount of talent that it produces for the cybersecurity sector.

Appropriate Regulatory Framework

A comprehensive governance framework that encompasses ethical considerations, data protection laws, and AI regulations is crucial for AI sovereignty. The Brazilian National Congress is discussing a [new bill](#) for an AI regulatory framework to help protect citizens' rights, promote fairness, and prevent discrimination and other potential risks, establishing sustainable, clear guidelines and standards for the development, deployment, and use of AI technologies.

While this ongoing initiative is laudable, it is not yet clear to what extent it can effectively regulate AI. The latest version of the proposed bill provides a necessary level of flexibility on key issues such as AI systems transparency, data security, data governance, and risk management. However, such flexibility, which is critical for the law to adapt to technological evolution, must be matched with a mechanism that allows specification through regulation or standardization.

In the absence of such specifications, the law risks being ineffective. The recent Brazilian experience regulating data protection illustrates that the adoption of modern law and the establishment of a new regulatory authority are only the beginning of the regulatory journey. The usefulness of underspecified legislation could be [jeopardized](#) if the pressing task of specifying the law is delegated to a regulator that seems “ineffective by design.”

Conclusion

Importantly, these AI sovereignty enablers are interconnected and mutually reinforcing. This consideration is particularly relevant, as legislators and governments around the world devise measures to regulate AI technology. Unfortunately, policymakers often overlook the importance of other elements of KASE. Understanding the interconnectedness of the KASE and leveraging their interdependence through an integrated approach are essential factors to achieving AI sovereignty and avoiding digital colonialism.

However, such an approach seems to be absent from the current Brazilian strategic vision for AI. Indeed, the 2021 Brazilian Artificial Intelligence Strategy has been [widely criticized](#) for including only general considerations about how AI could be implemented in several sectors, without defining how policymakers could better coordinate, assess, and allocate responsibility for the strategy's implementation.


The Brazilian administration should consider implementing these principles in the next revision of its strategic approach to AI. An integrated approach considering the KASE is instrumental to achieving AI sovereignty, developing indigenous AI capabilities, diversifying supply chains, increasing the digital literacy of the population, fostering strategic investments and partnerships, and safeguarding the security of critical AI infrastructure.

Not all countries will be able to elaborate and implement the necessary strategic, policy, and institutional changes allowing them to build an AI sovereignty stack. Such an effort might be especially herculean for countries in the Global South, which typically depend on foreign technologies. However, a careful mix of creative thinking and much-needed political vision regarding technological development may allow low-income countries to overcome some of the most burdensome obstacles, for instance by using open software to reduce the financial costs of procuring foreign software. The elaboration of an AI sovereignty stack, therefore, should be seen as a goal that all governments should strive to achieve even if not easily accomplished for every country.

Ultimately, countries that possess strong capabilities in the KASE areas are not only better positioned to maintain control over their AI technologies, policies, and data, but they also will likely increase their technological relevance, reducing dependence on external sources and preserving their national interests and autonomy in the AI landscape. Countries lacking such capabilities need to reconsider thoroughly their strategic approaches to AI in order to minimize the considerable risks of AI dependency that will likely exacerbate the already ongoing phenomenon of digital colonization.

Notes

- 1 The [right to self-determination](#) is a so-called primary principle or principle of principles, as it plays an instrumental role to allow individuals to enjoy their human rights, thus being an enabler of other fundamental rights. For this reason, it is enshrined as the first article of both the Charter of the United Nations, the International Covenant on Civil and Political Rights, and the Universal Declaration of Human Rights. According to these three international legal instruments, states have agreed that “all peoples have a right to self-determination” and that “by virtue of that right they are free to determine their political status and to pursue their economic, social and cultural development.” It is essential to emphasize the relevance of the [internal dimension](#) of self-determination, that is, the individual right to freely determine and pursue one’s economic, social, and cultural development, including by independently choosing, developing, and adopting digital technologies.
- 2 For the purposes of this article, governance is [defined](#) as the set of processes and institutional mechanisms that stimulate, facilitate, organize, and coordinate the stakeholder interactions of different stakeholders in a political space, to confront different opinions and interests regarding a specific issue and, ideally, achieve the proposal of the best possible regulatory solution. Regulation is intended as the product of governance, consisting of an ample range of instruments that can foster the stability and proper functioning of complex systems, where the presence of multiple actors with varying or divergent interests can naturally lead to instability and dysfunction.



Volume, Speed, and Accessibility as Autonomous Harms: Can Modern Legal Systems Deal With Harmful but Legal Content?

Agustina Del Campo

Volume, speed, and accessibility are among the many features that the internet brought about that significantly changed and positively impacted the information ecosystem. The amount of information shared online and the reach of such information is unprecedented and has been deemed a [democratizing force](#). The ability to keep and find information online has dramatically enhanced people's access to information, political participation, research and learning capacities, hiring practices, due diligence, and so forth. While international standards over freedom of expression, content, and dissemination have remained mostly unchanged, volume, speed, and accessibility are novel. But are they sources of new redressable harms?

Has the Internet Changed How Legal Systems Determine Harm?

Historically, volume, speed, and accessibility have been looked at by the law to assess damages. These features do not speak to the nature of the speech—whether it is harmful or not, legal or illegal—but rather to the nature and scope of the damages potentially generated by illegal or legally reprehensible content. As a matter of law, the content is what causes the harm. Therefore, if the content is ruled to be harmful—and the harm is legally redressable and attributable—then most courts around the world would look to the volume, speed, and accessibility, or the permanence, to determine the damages. For a defamation lawsuit, for example, advocates and judges would first need to establish the defamatory nature of the statements. Only then would they look at the accessibility and reach of such statements to assess damages. The volume or replication of that speech would also serve as a basis to

calculate the compensation owed to the injured party. However, speed and accessibility would not, in of themselves, render the statement defamatory.

With the advent of the internet, however, there seems to be increasing consensus that speed, volume, and accessibility should be considered specific sources of harm. Yet, advocates and judges have faced severe difficulties in fitting these categories within existing civil and constitutional standards. Volume and speed often cannot be attributed to a single individual; these phenomena are not dependent on the willingness or the intention of the content generator themselves. A piece of content may go viral inadvertently to the author, or it may not go viral despite the author's best intentions. Virality is achieved through the concurrence of conduct by third parties who often act in uncoordinated and unrelated ways. Since this does not represent individual conduct, nor can it be effectively attributed, causality cannot be established. Thus, legal principles dictate that such content cannot generate liability the way that publishing, editing, or selling a specific work can.

The Challenge of Virality

Courts, legislators, and advocates have been unable to properly characterize volume, speed, and accessibility and deal with them. Instead, they have been looking for ways to factor these features into existing law without addressing the underlying shortcomings of the existing legal frameworks. It is not unusual to find court decisions or bills of law tweaking and tuning legal standards to account for these factors, usually addressing them indirectly and opaquely.

Moreover, what has been popularized as harmful but legal content often refers to content that, at an individual level, does not amount to illegal or redressable damage but, at a collective level, may be deemed unjust or unfair for targeted victims given the speed at which the content spreads and the amount of views it garners. The emergence of this messy hybrid and contradictory category of speech stemmed from the difficulty of cataloging content that was potentially harmful to vulnerable groups, democracy, public health, safety, and economics, among others, but failed to reach the threshold for illegal or legally reprehensible content. In many instances, legitimate content, when aggregated with other similar content and/or made available to wider audiences, becomes problematic. There are many current examples of this trend—where the content is not illegal by itself but can allegedly produce societal harms when aggregated—including anti-vaccine movements, COVID-19 disinformation, and electoral disinformation.

One critical example relates to harassment on the internet. The accumulated level of gender violence against women is often manifest, immediate, visible, and yet hard to address from a policy or legal perspective. In these situations, individual pieces of content may not produce significant harm, nor do they amount to legally reprehensible content. But the aggregate effect can be problematic. These questions were recently addressed before a district court in Amsterdam for a case related to [the online harassment of a Dutch newspaper columnist](#). In this case, C. Gargard, a columnist at the Dutch newspaper *NRC*, had received 7,600

misogynist, discriminatory, and distressing messages. Gargard identified and presented 200 of them as evidence. The prosecutor filed charges against twenty-four defendants who were subsequently convicted by the court for a variety of criminal offenses, including incitement of assault, incitement of murder and manslaughter, incitement of discrimination, and incitement of defamation. The court found that the volume of messages increased the risk of harm for these crimes. The defendants argued that they had no intent to commit any of these crimes and their content individually did not amount to violations of the law. The Court determined that given the broader context where the statements were issued, these 24 defendants were liable under different crimes/misdemeanors depending on each case. In isolation, it is probable that each individual case would not meet the three-part test for discrimination (legality, necessity, and proportionality) under [Article 10 of the European Convention on Human Rights on freedom of expression](#). But, were the convictions proportional to the level of online harassment experienced by the plaintiff?

Unsettling Solutions

Given the difficulties of these issues, multiple stakeholders have turned to intermediaries for potential solutions. New bills of law are introduced daily around the globe to deal with internet companies' liability for harmful but legal content or awful but lawful content. The European Union's [Digital Services Act](#) (DSA) is probably the most recent law enacted in this area. The DSA mandates that companies must assess the risks that their services could generate and take steps to mitigate their impact. While this may make sense for speech deemed illegal in the EU (which is one of the many categories that the law asks companies to assess and mitigate), when it comes to awful but lawful or harmful but legal content, the questions that the Amsterdam district court faced will sooner or later come up for companies to decide. Should volume, speed, and permanence be considered risky per se? What would that mean for the internet as we know it or for the democratic advancements that the internet facilitated?

The [International Covenant on Civil and Political Rights](#) (ICCPR) expressly states that every person has a right to seek and receive information and ideas of all kinds across borders. International tribunals have found that limitations on the dissemination of information amount to restrictions of free expression. Under these benchmarks, laws like the DSA that mandate restrictions to dissemination or publication need to be specific in identifying the types of content to be restricted and must also ensure that these limitations are necessary and proportionate to the contemplated harm. Lawful but awful content does not reach this threshold. This raises a question: should the ICCPR standards be changed? And if so, what are the implications?

Modern legal systems did not anticipate legal conduct becoming illegal based on the number of viewers or the number of people who shared that same statement. Nor did they consider as relevant whether the statement would be accessible through time. As Yale Law School professor Robert Post [explained](#):

The scale of the internet produces forms of harm that may best be characterized as stochastic. Previously we asked whether particular speech acts might *cause* particular harms. The internet has rendered this kind of question almost obsolete. Speech that is simultaneously distributed to billions of persons may produce harm in ways that cannot meaningfully be conceptualized through the lens of discreet causality. We will need instead to think in terms of statistical probability of harm. Yet at present we lack any legal framework capable of assessing such stochastic harms in ways that will not drastically over-regulate speech.

There are two ways to interpret Post's assertion. One possible interpretation is that given the volume and the speed of harms produced, legal systems cannot effectively address them through ex post causality analysis. This is the argument that Stanford Law School's Evelyn Douek made in a [2022 article](#): "When the pursuit of formalism stands in the way of achieving other governance goals, like speed of decision making or responsiveness to prevailing social conditions, it will harm rather than enhance legitimacy and perceptions of accountability and effectiveness." She was discussing private content moderation, however, rather than state restrictions to speech. Another interpretation is that causality in some or many of these cases cannot be factually established because it is not the content that gives rise to the harm but rather the speed, volume, and permanence.

Conclusion

State limitations based on speed, accessibility, and volume are incompatible with current freedom of expression standards globally. This may be the reason why so many recent laws and bills to address freedom of expression online fail to comply with international human rights law. If freedom of expression is the key to democratic societies, this right should protect not only popular opinions but unpopular and even shocking statements. Further attempts to address harms raised by speed, volume, and permanence should first acknowledge the existing gaps. A number of questions could be raised thereafter: What would be the implications of changing existing global standards for freedom of expression? Could these new harms be legally addressed without undermining the philosophical premises underlying the right to freedom of expression and its importance in democratic societies? And what other models could be brought in to serve as frameworks for new, human rights-respecting solutions?

This essay reflects a larger research agenda I am pursuing on the topic of online harms and their compatibility with freedom of expression principles. As I get deeper into the research, I am convinced that if volume, speed, and accessibility are to be considered independent of sources of harm, the way forward to achieve this within the existing human rights framework lies in the framing of these questions and issues. The first step is to attain clarity about the incentives and trade-offs in adopting one solution over another. It is important that the legal community acknowledge the gaps in modern legal systems about how to deal with these new phenomena, particularly if judges and courts aspire to continue incorporating human rights principles into their decisions.

The Stalled Machines of Transparency Reporting

Jan Rydzak

On April 25, 2023, Twitter published a brief blog post innocuously titled “[An update on Twitter Transparency Reporting](#).” It opened with a capitulation: under its new management, the company would break from its [decade-long history](#) of publishing full reports on content restrictions and demands for user data and “review [its] approach to transparency reporting.” As a consolation, the uncredited post offered a stripped-down table detailing the volume of content restricted for violating each of Twitter’s policies in the first half of 2022. This was the last reporting period before Elon Musk took over the company and subsequently eliminated much of the company’s staff, including its [entire human rights team](#).

Twitter’s cursory announcement was curiously timed. The post appeared on the same day the European Commission [designated](#) Twitter a Very Large Online Platform (VLOP) under the European Union’s Digital Services Act (DSA),¹ which [obligates](#) such platforms to produce transparency reports on their content moderation efforts every six months. Two days later, Rest of World, a tech-focused media outlet, reported that Twitter under Musk had fully complied with [more than 80 percent](#) of governments’ demands for content removal. That data came from systems that had continued to automatically log information in the [Lumen database](#), a Harvard-hosted repository that has long worked with major platforms to document such demands. This represented a sharp jump from Twitter’s 50 percent compliance rate a year prior and struck a jarring note for a platform whose new owner began his tenure under maxims such as “[transparency is the key to trust](#).”

Figure 1. The End of Twitter’s Transparency Updates



The landing page of Twitter’s [Transparency Center](#), as of May 15, 2023, showing a series of updates between 2018 and 2021 that subsequently come to an abrupt halt.

Stagnant Waters

Twitter’s transparency turnaround stood in clear contrast to more than a decade of slow but steady [progress](#) in tech companies’ openness about how they handled content and pressure from governments. The presence of a tech leader who took pride in shirking his own commitments, agreements, and obligations also dealt a blow to the broad movement to hold tech giants to a high standard on these issues. Indeed, the usual release date of Twitter’s full transparency report had come and gone, and months of silence had elapsed, before the April disclosures were published. Observers [warned](#) that the platform risked becoming the antithesis of its former self, morphing from a [pioneer of transparency](#) into a purveyor of opacity for others to emulate.² Musk himself made no attempt to continue championing the resistance against unwarranted government-ordered content removals that had previously led to Twitter [challenging governments](#) in court.

Yet, while the former “bird app” (now “X”) offered the most glaring departure from precedent, progress had quietly stalled among many of its peers. In a 2022 assessment, Ranking Digital Rights [found](#) that none of the fourteen digital platforms it scrutinized had divulged any new information on how they treated government censorship demands in more than a year. Worse, alongside this lack of meaningful progress, even the existing data on the topic had [deteriorated overall](#) due to Twitter’s removal of figures on “unofficial” demands from governments. Opacity and aggregation remained pervasive.

Figure 2. Meta’s Report on Government-Requested Content Restrictions



Meta’s [report on content restrictions](#), as of May 15, 2023, showing a clear “pandemic surge” in government demands, beginning in the first quarter of 2020 and continuing in every subsequent reporting period. The systematic surge of pressure from governments is a recurring trend across platforms and adds urgency to calls for greater transparency.

The pace at which companies are introducing transparency reporting appears to be slackening as well. Between mid-2021 and mid-2022, several companies—including BitChute, Clubhouse, OnlyFans, Yubo, and even popular Chinese video app [Kuaishou](#)—published their inaugural reports on how they enforced their standards. The Global Internet Forum to Counter Terrorism (GIFCT), a tech industry initiative that aims to increase coordination on identifying and moderating extremist content, [reported](#) in 2022 that four companies had started releasing enforcement data as a step toward joining the group. But by July 2023, the stream of new entrants in transparency reporting had dried up, with Microsoft-owned gaming platform Xbox as the most notable exception.

Recent academic studies have articulated a similar rift between growing expectations of transparency and companies’ lethargic pace of adjustment. In [one illustration](#), Aleksandra Urman and Mykola Makhortykh, showed that no Big Tech platform comes [anywhere near](#) full compliance with the [Santa Clara Principles 2.0](#), a set of recommendations that are arguably the closest proxy for a common set of content governance principles.

Silver Linings

Companies’ reluctance to be fully candid about their content governance practices has become a trigger for more assertive forms of pressure, including from grassroots initiatives and massive-scale regulatory action.

Shareholders are a vocal group of actors on this front. In the past year, both Apple and Amazon have faced petitions launched by their own shareholders to come clean on government censorship demands. Yet, these campaigns have resulted in differing outcomes.

The iPhone manufacturer, which had shared selected data on the subject in previous years, [promised](#) a dedicated App Store transparency report that was ultimately [released](#) in May 2023. Among other revelations, the report laid bare an extreme jurisdictional skew: out of 1,500 apps removed in 2022 in response to government requests, 97 percent had been taken down in mainland China.

Amazon, meanwhile, has entrenched itself in its secrecy. The dominant titan of the global e-commerce industry has long revealed [nothing](#) about how it responds to pressure from governments to suppress content or tailor its local iterations according to their preferences.³ The [purging of reviews](#) for a book authored by Chinese President Xi Jinping and the censoring of [search results](#) for LGBTQ content in the United Arab Emirates provide glimpses regarding Amazon's content governance. But the company's policies generally remain a mystery. A [2023 effort](#) by investors to force more policy transparency from Amazon through a stand-alone shareholder proposal won [10 percent of shareholders' votes](#) at the firm's annual meeting. Ahead of the vote, the company [argued](#) that it had no reason to publish detailed data on content restrictions because, among other things, it had "chosen to offer a very broad range of viewpoints, including content or products that may conflict with our stated positions."

Although the Amazon proposal did not persuade the majority of the company's investors, it marked the first time that a petition on government censorship demands reached the voting stage at an e-commerce company. This has raised the visibility of content governance and censorship as issues of broad concern for tech companies rather than as flash points for social media platforms. It also sets the stage for government regulation.

The EU's DSA, which came into force in November 2022, is the strongest incentive to date for global tech giants to set the wheels of transparency reporting in motion. Most of the services that received the status of VLOPs alongside Twitter already had broad reporting on content moderation in place. But four did not: AliExpress, Amazon, Booking.com, and Zalando. Along with [fifteen of their peers](#), these platforms were required to enact several of the DSA's enhanced requirements by August 2023 or risk hefty fines. Regular reporting on content moderation and policy enforcement is one of those requirements.

In September 2023, the European Commission provided a first look at how seriously platforms treated enforcement when it launched a [DSA Transparency Database](#), an open repository to which the largest platforms are meant to submit statements explaining each content moderation decision that affects users in the EU. In its first day, the Transparency Database accumulated more than 5 million entries. More than half of these were submitted by TikTok. Twitter ("X") submitted 23,000, all but one of which pertained to the protection of minors.

The DSA's transformation from theory to reality offers hope for a resurrection of dormant transparency reporting and renewed efforts toward increasing its quality. The inclusion of online marketplaces and platforms outside the constellation of social networking tools also sends a strong regulatory signal: as long as a major platform moderates content, it should

deliver a public account of its efforts regardless of what corner of the tech industry it occupies. The DSA should also serve as a jolt for platforms to apply a higher standard of transparency outside the EU. Civil society groups should enhance their vigilance and call out major imbalances between companies' transparency in the EU and elsewhere, lest the transparency gap continue to widen.

There are advancements in other corners of the tech world as well. In 2020 and 2021, South Africa's MTN and Mexico's América Móvil became the first telecommunications companies in their respective regions to issue dedicated transparency reports. [The former](#) illustrated the natural tensions that such revelations generate: in MTN's first year of reporting, authorities in Iran were responsible for the vast majority of the demands that the company disclosed, but this data was [removed](#) from subsequent transparency reports, as was the original iteration in its entirety.

Nonetheless, the granularity of MTN's report sets an example for other industry players in Africa and beyond. Similarly, companies and divisions in other sections of the tech industry are making first forays into the field, often under the banner of trust and safety. In online gaming, Xbox has led the charge, [revealing](#) more than 7 million instances of content removal and account suspension in the first half of 2022 alone.

No Guarantee

But the progress in industries beyond social media is tenuous, and the state of play among larger platforms is both precarious and prone to reversal. The mass layoffs that began sweeping through the tech world in 2022 took an [extreme toll](#) on trust and safety teams, ethics departments, task forces on integrity and misinformation, and the programs they created and supported. Content moderators worldwide, with little power and inherently precarious contracts, have seen their livelihoods [slashed](#) in the name of efficiency.

With a human workforce too small to handle endless waves of content, tech companies have increasingly turned to AI, buoyed most recently by the media attention sparked by OpenAI's launch of ChatGPT in November 2022. A month earlier, Spotify [acquired](#) AI-powered content moderation company Kinzen but has yet to publish any enforcement data. For Xbox, the acquisition of a similar tool prompted a [ninefold increase](#) in the number of enforcement actions in the quarter immediately following the purchase, particularly against accounts deemed inauthentic. Enforcement surges may ultimately improve safety on the platform. But the sheer volume of new takedowns warrants more disclosure about how the platform addresses the risk of overenforcement, including whether the takedowns came with a parallel surge in appeals.

The trade-offs between human and automated labor also come into play here, particularly as tech companies [continue](#) to lay off their workforce. In July 2023, Shopify was reportedly [laying off](#) its customer support representatives en masse and replacing them with a chatbot.

Moves like this ratchet up the tension between investors' demands for cost-saving measures and the long-term threats to human rights that emerge from automated tools that are not adequately supervised or held in check.

Transparency is a tug-of-war whose outcome is not preordained. Positive momentum can quickly fizzle out, grind to a halt, or reverse course. Behind every reported statistic lie instances of potential human rights harms and legitimate questions about how well platforms protect their users. Technological advances in artificial intelligence have triggered [apocalyptic visions](#) that offer a compelling distraction from the damage AI can inflict today. But digital rights researchers and advocates should continue to keep a spotlight firmly trained on established vehicles of digital repression. The guardrails around them can crumble easily.

Notes

- 1 Under the DSA, a platform must report at least 45 million monthly active users to qualify as a VLOP or Very Large Online Search Engine (VLOSE). VLOPs and VLOSEs face the strictest reporting requirements.
- 2 Such widespread and deliberate disruptions to the flow of disclosures, though still hypothetical for now, would mirror the apparent degradation of quality among social media platforms that author Cory Doctorow memorably dubbed "[enshittification](#)."
- 3 This eschews not only the norms followed by other major U.S. tech companies but also the progress made by its closest competitors in democratic settings. Like many large tech companies in the United States, Amazon does publish regular [reports](#) on government requests for user data, but it trails its peers on transparency regarding content and account suppression, including on behalf of governments. EBay, Amazon's distant runner-up in the United States, publishes [detailed data](#) about its content moderation mechanisms, including a regulatory portal through which government actors request the removal of listings. Mercado Libre, the largest e-commerce platform in Latin America, stops short of discussing requests from state actors but [far outpaces](#) Amazon in its reporting on content deemed to violate its rules. China's online marketplaces, some of which have gained a strong foothold internationally, remain largely silent on government demands.



The UN Global Digital Compact Must be Multistakeholder and Inclusive

Irene Poetranto

The internet's early days were marked with optimism that it would be a "[liberation technology](#)" that would foster democratic norms, facilitate political mobilization, and bolster the impact of civil society. As internet access has proliferated, however, there has been [a persistent decline](#) in internet freedom globally for more than a decade. This downturn has been driven by increasing state-sponsored online attacks on free expression, such as through the use of [internet shutdowns](#), [internet filtering systems](#), and sophisticated [commercial spyware](#) against civil society. Those advocating for the rights of vulnerable populations, such as ethnic, religious, and sexual minorities, and those living in conflict-affected regions have been especially targeted.

Considering these trends, the United Nations' effort to create a [Global Digital Compact](#) (GDC) has a crucial role to play in ensuring that the internet remains open, secure, and inclusive for all, allowing for engagement in public debate, full participation in the digital economy, and the achievement of the UN Sustainable Development Goals (SDGs).

What Is the GDC?

The GDC stems from UN Secretary General António Guterres's 2021 [Our Common Agenda](#) report, which aims to create "shared principles for an open, free, and secure digital future for all" to be agreed at the 2024 Summit of the Future. The GDC's purpose is to establish an agreement on how best to address complex digital issues, including "connecting the unconnected, avoiding fragmentation of the Internet," and protecting digital rights. The Our Common Agenda report proposes twelve actions that are designed to accelerate achieving the SDGs—one of which is to "apply human rights online."

Despite this acknowledgment and the [consensus by](#) the UN Human Rights Council that “the same rights that people have offline must also be protected online,” UN member states have continued to tamper with or disrupt networks globally, including by filtering and shutting down the internet. These disruptions have caused billions of dollars in [economic losses](#), harmed the globally interconnected internet infrastructure, and hampered international collaboration. Digital attacks against civil society in particular have not only become [more frequent](#) but also more sophisticated, particularly through the use of [powerful](#) surveillance technologies. Yet, civil society plays a key role in ensuring that the voices of marginalized people and those in underrepresented countries and regions are included in the development and deployment of digital technologies.

Challenges to Achieving the GDC

Governments of underrepresented countries, such as Indonesia, Kenya, and South Africa, as well as various civil society groups, are increasingly participating in global debates on cybersecurity. An example is the UN Open-Ended Working Group (OEWG), which is a process that began in 2019 that seeks to establish rules for responsible state behavior in cyberspace. The first OEWG [included the](#) participation of almost all underrepresented countries; many regional organizations, such as the Organization for Security and Cooperation in Europe, the African Union (AU), the Organization of American States, and the European Union; and over 100 nongovernmental organizations (NGOs) that attended the meetings as observers.

Yet, the OEWG has faced challenges. Subsequent OEWG meetings involved far less NGO participation due to objections from China and Russia. Substantive progress in the OEWG on pressing concerns, such as data sovereignty and the application of international law to cyberspace, has also stalled as a result of geopolitical tensions between Western democracies and China and Russia. Furthermore, many participants registered [concerns about](#) how best to achieve cybersecurity goals while safeguarding sovereignty and the principle of noninterference in the domestic affairs of other states (a point emphasized by more recently decolonized countries, such as India and Indonesia). Even so, the OEWG successfully adopted a [consensus report](#) in 2021, which despite criticism over its [broad and vague](#) language, recognized the growing digital threat landscape and the importance of strengthening collaboration with civil society and others to address such threats.

The OEWG process has shown that countries desire to discuss the security of cyberspace under the auspices of the UN. In this vein, the GDC can facilitate building common ground for concrete policies and actions, and foster dialogue to deepen trust that will improve upon the OEWG. For example, while the OEWG has been criticized for implementing a highly politicized process that has increasingly excluded civil society from its deliberations, the GDC [seeks to involve](#) many stakeholders, including “governments, the United Nations system, the private sector (including tech companies), civil society, grass-roots organizations, academia, and individuals, including youth.” To ensure that such participation occurs in a

collaborative and inclusive manner, the GDC should prioritize the principle of multistakeholderism, while incorporating mechanisms that [allow for](#) bottom-to-top input, particularly from people in developing countries. Such mechanisms would help member states and the global community hear not only from those who have benefited from different forms of digital technologies but also from those who have been harmed by them, especially those [who are](#) most marginalized. In other words, the GDC has [the potential to “re-energize”](#) the multistakeholder model through specific commitments and concrete actions that are conducted in an inclusive manner.

Conclusion

Although disagreements among states are likely to persist, the GDC can help identify specific goals that are [of mutual interest](#) among countries that approach cybersecurity and internet regulation differently and therefore enable cooperation. Deepening collaboration and trust would [help convince](#) states that have not decided how the internet should be regulated about the value of adopting inclusive and multistakeholder governance, rather than pursuing the state-dominated model advocated by China and Russia. For instance, in their respective submissions to the GDC, the [United States](#) and [China and the Group of Seventy-Seven \(G77\)](#), a coalition of developing countries, both agreed that the internet and digital technologies are crucial to achieving the SDGs and that there is a need to safeguard the security and interoperability of the internet to avoid its fragmentation.

Given this concurrence, the GDC should elevate the shared principle of an open, distributed, and interconnected internet and create specific avenues for collaboration—both with UN agencies and beyond (for example, with regional civil society and groupings such as the AU)—that will protect the complex ecosystem of devices and infrastructure that make the internet accessible. The inclusion of such a pledge would make tangible policy decisions possible, and result in a more meaningful and action oriented GDC.



For AI, Does Democracy or Development Come First?

H. Akin Unver

The increasingly widespread adoption of AI has revealed a stark divide: while developed nations curate dialogues around ethical AI and democratic technological frameworks, emerging economies are often caught in a turbulent pathway of rapid modernization and authoritarian misuse. Some of the literature on [democratic development](#) cynically contends that for developing nations and emerging economies, ethical AI and its democratic use are a luxury, eclipsed by the pressing urge to leverage AI for technological, economic, and arguably, political gain. As AI burgeons into a pervasive force across global landscapes, a discussion has emerged: Does responsible and democratic AI use catalyze development, or is development a precursor, enabling nations to then adapt AI usage to democratic norms?

Some of the prominent scholars of democratization and development [posited](#) that democratic structures, underscored by inclusive institutions, provide a fertile ground upon which development can robustly bloom. This perspective argues that situations where citizens have a voice and institutions are accountable and inclusive foster policies that support broader public interests, which, in turn, propel economic development.

However, [a significant body of literature](#) takes a contrary view, contending that economic development creates conditions conducive for democratization by enhancing education, creating a middle class, and fostering a civil society that demands democratic governance. Another group of democratization and development scholars propose that wealthier and more developed societies are more likely to sustain democratic institutions. [They argue that](#) economic prosperity engenders an educated and economically stable middle class, which starts to demand greater political participation, accountability, and transparency from governing bodies. Countries such as South Korea, Taiwan, Singapore, and Hong Kong are often [cited](#) to underscore the development-induces-democracy perspective.

[Some have argued](#) that utilizing AI within a democratically ethical framework can, indeed, catalyze development. Democratically governed AI systems promote transparency, accountability, and inclusivity, ensuring that technology is leveraged for the collective good and potentially minimizing socioeconomic disparities. Democratic AI usage implies that technology is deployed transparently, with mechanisms in place to mitigate bias and ensure equitable access and benefits. In such a scenario, AI can enhance various sectors inclusively, from healthcare and education to agriculture and public services, steering a nation toward comprehensive development.

Conversely, AI adoption without invoking democratic norms and regulatory frameworks can lead to ethical and societal pitfalls. Governments can wield AI tools for mass surveillance, information manipulation, and to accentuate existing social and economic disparities, especially when deployed without adequate ethical and democratic checks and balances. As detailed by Carnegie’s [AI Global Surveillance Technology index](#), an increasing number of nations—in fact, the majority of governments—utilize AI-facilitated surveillance and data analytics to monitor citizens and potentially quell dissent, exemplifying the darker underbelly of unbridled AI adoption devoid of democratic governance. Instead of nurturing democratization, AI can be an instrument to fortify authoritarian proclivities.

Nations that prioritize democratization prior to or in tandem with AI adoption may theoretically cultivate a more ethically aligned technological landscape. Democratic norms, such as transparency, accountability, and citizen participation, when ingrained within AI development and deployment, can ensure that governments and private companies use technology in a manner that aligns with societal welfare and ethical tenets. In this scenario, regulatory frameworks, policies, and societal norms that uphold democratic principles guide AI applications.

However, this trajectory, largely defined by stringent regulatory frameworks, policies, and societal norms, demands [meticulous scrutiny](#) of AI applications, ensuring unbiased algorithms and equitable access to technology. The bureaucratic and ethical oversight, while safeguarding societal welfare and ethical tenets, may potentially decelerate the pace of AI innovation and implementation. The bureaucratic obstacles involved in ensuring compliance with democratic norms could inhibit the quick, agile deployment of AI technologies occurring in less regulated environments. In this scenario, the ethical high road, punctuated by rigorous democratic frameworks, could risk [technological and economic stagnation](#), as other nations, unencumbered by such considerations, sprint ahead in the global AI race. Consequently, nations that prioritize ethics and democratization might find themselves grappling with the unsettling reality of being globally outshone by counterparts that have embraced a path of rapid, albeit ethically questionable, AI deployment. Slower development, in turn, may lead to lower job creation, slower infrastructure development, and slower growth, paving the way for [public support for authoritarian practices](#).

The interconnectedness of AI adoption, development, and democratization reveals a multifaceted narrative, particularly complex for developing nations straddling the demands of rapid technological advancement and ethical governance. The exigency of development often propels these countries toward a trajectory where AI adoption may sideline democratic norms, not out of disregard for ethical use but for pragmatic reasons: [to ensure global competitiveness](#) and to [satisfy internal developmental pressures](#). While unbridled AI adoption holds the promise for developmental leaps, it also risks becoming a tool for authoritarian consolidation, mass surveillance, and the exacerbation of socioeconomic disparities. For nations navigating this intricate pathway, an imperative emerges for a nuanced, contextually apt approach toward intertwining democratic governance with AI development and use. Thus, the future necessitates an inclusive global dialogue and cooperative platform that acknowledges the varied socioeconomic, cultural, and political landscapes of nations, fostering an environment where knowledge, resources, and collaborative efforts toward responsible AI use can coalesce.

Navigating the entwined trajectories of AI, development, and democracy uncovers a realm where scientific scholarship and policymaking intertwine yet diverge—and where consensus remains elusive. The discourse among academics and practitioners continues to grapple with an unyielding question: is the democratization of AI an imperative or a luxury in the pathway toward rapid technological and economic development? The inconclusiveness emanating [from scholarly debates](#) underscores that the link between faster technological development and the prerequisites of [political liberalization and democratic governance](#) remains shrouded in ambiguity.

Consequently, Western nations need to recalibrate their diplomatic and international development dialogues. Rather than endorsing a prescriptive model that underscores political liberalization as a quintessential precursor for sustainable AI development—a luxury for developing nations—it is preferable to adopt a nuanced stance that emphatically prioritizes developmental outcomes. The emphasis should traverse beyond democratization to recognize and validate diverse pathways toward development, acknowledging that nations particularly in the Global South may sculpt their developmental and technological narratives divergently, shaped by their unique sociopolitical and cultural tapestries. Moving forward, it is crucial to foster a multilateral, inclusive dialogue that emboldens a collective pursuit toward ethical AI development, while respecting and appreciating varied developmental paradigms. Through this lens of mutual respect, collaborative spirit, and acknowledgment of distinct pathways, the global community can navigate the future of AI, development, and democracy, aligning technological progress with an array of governance frameworks.



What Matters More for U.S. and EU Tech Export Controls: Human Rights or Geoeconomics?

Brian Kot

In recent years, both the United States and the European Union (EU) have modernized their export control regulations, particularly those related to digital technology. Yet, experts disagree on the purpose and intent of these changes.

Scholar Cindy Whang, for instance, [argues](#) that the United States has centered its export control policies around the concept of geoeconomics, or an increasing “securitization of economic policy and economization of strategic policy.” The EU, on the other hand, has modernized its export control regime to position human rights as a central pillar of its dual-use export controls. According to Whang, the infusion of geoeconomics in U.S. national security and the EU’s incorporation of human rights considerations are “made mutually exclusive of each other” and detrimental to unifying export controls. Another scholar, Olga Hrynkyiv, [contends](#) that U.S. export controls conceptualize national security as “at least in part intertwined with its economic policy concerns and the race for technological supremacy,” whereas the EU “emphasizes the human security dimension of its export control[s]” and “seems to be more resilient and better able to defend its values without resorting to excesses of securitization.”

Yet, recent developments in the respective export control policies of the United States and the EU undermine this dichotomized narrative—human rights and geoeconomic concerns are prominent in both jurisdictions’ export regulations. While the EU has undoubtedly been a vanguard in using export controls to advance fundamental rights, the United States has also taken substantial steps to enshrine human rights principles in its export control policies. Likewise, while the United States has shown a far greater willingness than the EU to use export restrictions to secure its geoeconomic interests, exemplified by its curtailment

of certain advanced technologies to China, the EU's sweeping sanctions against Russia demonstrate the bloc's willingness to use export controls to protect its geoeconomics interests as well.

Export Controls for Human Rights

The EU is often portrayed as a “[key norm-setter](#)” when it comes to controlling the export of surveillance technologies. The bloc began advocating for stronger regulations of cyber-surveillance exports between 2009 and 2011, when journalists and researchers [revealed](#) that European companies had provided surveillance technologies to governments in Bahrain, Egypt, Iran, and Syria. Cognizant that EU-based firms had been major exporters of surveillance technology, the EU attempted to get its house in order. In 2016, the European Commission [kick-started](#) a process of reforming the EU's regulations on dual-use exports—technologies that can be used for both civilian and military purposes. The process culminated in a [new regulation](#) in 2021. In addition to explicitly referencing human rights as a consideration in its export controls, the new regulation created several mechanisms so that EU regulators could control a wider range of dual-use items, enhance accountability from exporters, and increase transparency about how member states were implementing export control rules. While certain human rights protections proposed by the European Commission were ultimately [scaled down](#) due to pressure from member states, the resulting political compromise still represented a modest but steady stride toward enshrining the protection of fundamental rights as a normative pillar of EU's dual-use export control regime.

Compared to the EU, the United States is often portrayed as a [laggard](#) in regulating digital technologies domestically. But when it comes to denying authoritarians access to American technology, the United States has taken a more active role. In 2018, then president Donald Trump signed the Export Control Reform Act (ECRA), which expanded the U.S. Department of Commerce's ability to impose unilateral controls on items not covered under existing international agreements (notably the Wassenaar list). ECRA affirmed that “carry[ing] out the foreign policy of the United States, including the protection of human rights and the promotion of democracy” is one of the act's [policy purposes](#). In October 2020, U.S. officials amended the Export Administration Regulations (EAR)—a set of export guidelines administered by the Department of Commerce's Bureau of Industry and Security (BIS)—to enhance “[consideration of human rights concerns](#) when reviewing almost all license applications for items on the Commerce Control List.” In March 2023, the BIS further amended the EAR to “[explicitly confirm](#) that the foreign policy interest of protecting human rights worldwide is a basis for adding entities to the Entity List” (emphasis added).

The United States has taken concrete actions in furtherance of these policies. For instance, regulators have taken aggressive steps to tackle China's “[tech-enhanced authoritarianism](#),” adopting wide-ranging sanctions against Chinese entities found to be materially supporting the Chinese Communist Party's brutal atrocities against Uyghurs in the western Xinjiang region. According to the Uyghur Human Rights Project, between October 2019 and March

2023, the BIS put a total of [seventy-four Chinese government and business entities](#) on its entity list, preventing them from accessing U.S. technology.

The United States has also stepped up its global leadership to curb the proliferation of digital surveillance. For example, in combating the spread of spyware, the U.S. government not only imposed sanctions on major vendors, such as the [NSO Group, Cytrox, and Intellexa](#), but President Joe Biden also signed an [executive order](#) in March 2023 limiting federal agencies' use of commercial spyware. During the second Summit for Democracy, the United States led a group of twenty-four countries to launch a voluntary [code of conduct](#) for enhancing export controls to prevent human rights violations. The nonbinding document will not be a silver bullet for ending the flow of goods and technologies to human rights abusers, but it reflects a policy shift that human rights ought to be taken into account in the conduct of international trade.

Export Controls for Geoeconomics

This is not to say that protecting human rights is the primary goal around which the U.S. export control system is designed. One of the ECRA's national security policy goals is to maintain the United States' leadership in the science, technology, engineering, and manufacturing sectors. In particular, the ECRA focuses on controlling the exports of “emerging” and “foundational” technologies, which are deemed critical to innovation and U.S. competitiveness in global markets. Hence, under the ECRA, U.S. dual-use export controls are no longer just about mitigating military risks. They are redesigned to secure, as U.S. National Security Adviser Jake Sullivan put it, “[as large of a lead as possible](#)” in U.S. technological and, hence, economic leadership.

U.S. export control measures against China exemplify this geoeconomic approach. On October 7, 2022, the BIS issued a [new rule](#) significantly expanding the scope of export restrictions against Chinese actors. Under this rule, export controls are not tailored to Chinese entities found to be aiding and abetting repression in China or even entities found to have ties with the Chinese military. Instead, the controls aim to set back China's AI development by limiting its access to Western technologies across at least [four layers of the AI development stack](#): high-end chips, chip design software, semiconductor manufacturing equipment, and components to build manufacturing equipment. While the BIS notice does reference China's use of high-end chips for committing human rights abuses, undermining China's technology base is likely the overriding concern, since the restrictions seem disproportional for purely addressing human rights issues.

European allies similarly demonstrate a growing incorporation of geoeconomics in their export restrictions policies. In March 2023, after months of U.S. persuasion, the Netherlands—home to ASML, the only company in the world that can produce highly sophisticated chip-making machines—announced new export controls on semiconductor technology. In a letter to parliament, trade minister Liesje Schreinemacher outlined the

three strategic goals of the new controls: 1) preventing Dutch goods from contributing to undesirable end use; 2) preventing unwanted strategic dependencies in the long term; and 3) maintaining Dutch technological leadership. The first principle falls within the purview of the EU's policy of incorporating human rights in its export controls, but the second and especially the third justifications move beyond pure human rights or traditional national security logics—they [echo](#) the United States' rationale of widening the technological gap vis-à-vis adversaries.

While pressure from the U.S. government likely influenced the Dutch government's decision, this does not undercut the latter's motivations. The Hague's move to curb tech exports coincides with an EU-led initiative to securitize economic and trade relations with geopolitical rivals. The bloc unveiled the [European Economic Security Strategy](#) in June 2023, acknowledging the need to “[complete] traditional approaches to national security with new measures to safeguard our economic security” and advocating for “a comprehensive strategic approach to economic security, de-risking and promoting [the EU's] technological edge in critical sectors.” The new strategy signifies a shift in EU policy, challenging the bloc's modus operandi of separating trade matters from national security issues. As one expert [describes](#) it, “The EU's division between trade instruments controlled by the European Commission and security instruments controlled by member states is increasingly inadequate in the face of technological and industrial rivalry where economic security and national security are intertwined.”

To be sure, analysts have pointed out that the United States and the EU are “[only partially aligned](#)” on the appropriate depth of technology decoupling from China.” But differences on China should not mask the bloc's newfound drive to better integrate economic and security strategies. According to experts Matthias Matthijs and Sophie Meunier, in recent years the EU has undergone a “[geoeconomics revolution](#)” that has transformed Brussels into “a major macroeconomic and geoeconomic actor in its own right.” The question is not *whether* but *from whom* and *to what degree* the EU is hoping to achieve economic security.

When dealing with an adversary that it deems sufficiently threatening, the EU has shown an ability to muster the political will and regulatory capacity to protect its strategic interests. Russia's invasion of Ukraine illustrates that point. Moscow's aggression has forced Europe to securitize economics on an unprecedented scale. European countries have applied robust sanctions and trade restrictions not just on cutting-edge technology and dual-use goods but also on commodities and luxury goods. The list of banned products “is designed to [maximize the negative impact](#) of the sanctions for the Russian economy.” Interestingly, China's complicity in Russia's war efforts prompted the EU to extend trade restrictions against Beijing. In June 2023, Brussels [imposed sanctions](#) on a handful of Chinese entities for supporting Russia's war machine, reflecting the EU's increased willingness to wield economic tools to defend its geopolitical interests.

Conclusion

When it comes to regulating the export of digital technology, the United States is often described as self-interested, relying on geoeconomics calculi, whereas the EU is perceived to be largely driven by normative ideals and human rights principles. In fact, both jurisdictions have strengthened their legal basis for using export controls to pursue a broad set of foreign policy objectives—including geoeconomics objectives as well as human rights goals. While the substantive provisions differ, both jurisdictions have elevated human rights language in their export control regimes, and both have prioritized geoeconomics in enacting high-tech trade restrictions.



About the Authors

Mahsa Alimardani has been working within civil society for over a decade on projects related to access to the internet, especially in Iran. She is a senior researcher with the international freedom of expression organization ARTICLE19, working on digital rights within the Middle East and North Africa region. She is also a PhD candidate at the University of Oxford's Oxford Internet Institute, where she is finishing a thesis on Iran's internet controls.

Afef Abrougui has more than ten years of experience researching and writing about technology and human rights. She is the owner of Fair Tech, a consultancy based in The Hague whose mission is to protect human rights in the digital space.

Arindrajit Basu is a PhD candidate at Leiden University and a nonresident fellow at the Centre for Internet & Society, India.

Luca Belli is a professor of digital governance and regulation at Fundação Getulio Vargas (FGV) Law School, Rio de Janeiro, where he directs the Center for Technology and Society (CTS-FGV) and the [CyberBRICS](#) project.

Agustina Del Campo heads the Center for Studies on Freedom of Expression and Access to Information (CELE) at the University of Palermo in Argentina.

Steven Feldstein is a senior fellow in Carnegie's Democracy, Conflict, and Governance Program, where he focuses on issues of democracy and technology, human rights, and U.S. foreign policy.

Iginio Gagliardone teaches at Wits University and is the author of *China, Africa, and the Future of the Internet* (Zed Books, 2019), *The Politics of Technology in Africa* (Cambridge University Press, 2016), and *Countering Online Hate Speech* (UNESCO Publishing, 2015).

Brian (Chun Hey) Kot was a research assistant in Carnegie's Democracy, Conflict, and Governance Program.

Irene Poetranto is a senior researcher for the Citizen Lab at the Munk School of Global Affairs, University of Toronto. She is also a doctoral candidate in the Department of Political Science, University of Toronto. Her research interests include internet governance and the politics of internet regulation. Southeast Asia is her geographic area of focus. Irene obtained her master's degree in political science and Asia Pacific studies from the University of Toronto and her bachelor's degree in political science from the University of British Columbia.

Jan Rydzak is the digital transformation lead at the [World Benchmarking Alliance](#). He oversees the [Digital Inclusion Benchmark](#), which evaluates technology companies on issues ranging from [ethical AI](#) to expanding access to technology for women and girls. His previous work in civil society and academia focused on corporate accountability and the impact of internet shutdowns on human rights and collective action.

Janjira Sombatpoonsiri is a member of the Carnegie Digital Democracy Network. She is an assistant professor at the Institute of Asian Studies, Chulalongkorn University; a research fellow at the German Institute for Global and Area Studies; and a regional manager for the Digital Society Project.

H. Akin Unver is an associate professor of international relations at Ozyegin University in Istanbul and the coauthor of "Democratization, State Capacity and Developmental Correlates of International Artificial Intelligence Trade," forthcoming in *Democratization*.



Carnegie Endowment for International Peace

The Carnegie Endowment for International Peace is a unique global network of policy research centers around the world. Our mission, dating back more than a century, is to advance peace through analysis and development of fresh policy ideas and direct engagement and collaboration with decisionmakers in government, business, and civil society. Working together, our centers bring the inestimable benefit of multiple national viewpoints to bilateral, regional, and global issues.

Democracy, Conflict, and Governance Program

The Carnegie Democracy, Conflict, and Governance Program rigorously analyzes the global state of democracy, conflict, and governance, the interrelationship among them, and international efforts to strengthen democracy and governance, reduce violence, and stabilize conflict.



[CarnegieEndowment.org](https://www.carnegieendowment.org)