MIT Technology Review
Publicado por TEC

The Hidden Cyber Risks of Well-Intentioned Regulations

What you will find in this article:

– Tensions between scanning and the protection of personal data – Increase in privacy violations with the removal of encryption
– False accusations and reputational damage in the use of scanning technologies

by **Stuart Madnick** August 12, 2025

 Download

Data managers are increasingly being pressured to operate in a complex and sometimes contradictory regulatory environment where laws and technologies intertwine in ways that can have unintended consequences for companies. Regulations that require the use of technologies to scan harmful content on digital platforms, for example, introduce vulnerabilities that weaken encryption and compromise the privacy and security of all consumer data.

Across sectors and governments, the consensus is growing that data privacy is not just a technical issue, but a fundamental digital right. In response, leaders have been pushing for more robust regulatory frameworks and stronger protection mechanisms to protect consumer data, strengthen trust, and combat illegal activities. But sometimes, these two goals conflict. So, what to do?

Tensions between scanning and protecting personal data

A clear example of the possibility that well-intentioned laws lead to unintended consequences is in the global efforts to eliminate child pornography online. This is almost a consensus. As a result, governments around the world have been proposing or implementing regulations that require tech companies to seek out and report cases of child sexual abuse (known as CSAM) to law enforcement.

In Brazil, for example, online child sexual abuse material is combated through a set of laws and initiatives focused on protecting minors and combating online exploitation. Brazilian legislation criminalizes the production, possession and dissemination of CSAM, with penalties provided for in the Penal Code. In addition, laws such as the Artificial Intelligence Law and the General Data Protection Law (LGPD) help to ensure the responsible use of technologies and the protection of personal data, which is relevant in the context of CSAM detection and prevention.

Internationally, there are other initiatives such as the **Earn It Act and STOP CSAM in the** United States, which propose to hold digital platforms accountable for the distribution of CSAM by their users. There is also the **UK's Online Safety Act**, **Australia's Online Safety Act** and a **European Union** proposal that would require digital platforms to scan not only images, but also texts, to detect attempts to groom minors, seeking to prevent future abuse.

According *to the International Centre for Missing and Exploited Children*, since 2006, **156 countries** have improved or created laws against CSAM.

These efforts are clearly well-intentioned. However, what many may not realize is that these regulations can have severe side effects on the security and privacy of millions of citizens, with important implications for companies that handle personal data.

One of the central points is end-to-end encryption, a technology that encrypts messages in such a way that only the sender and receiver can decode them. Not even the digital platform can read the content. This technology is essential to ensure the confidentiality of communications, the privacy of users, and freedom of expression. In

addition, it protects data even if it is intercepted by cybercriminals. In the US, the use of apps like Signal, which use end-to-end encryption, has received a lot of media attention. The technology works perfectly, the problems usually occur in misuse, and not in the technology itself.

But if digital platforms are held accountable for illegal photos shared in their environments, they will need a way to bypass encryption so they can see the real message — such as through a "master key" or a "backdoor." A master key is like those old ones that open any door, or like the "fireman's key" present in many elevators, which allows you to take control. A backdoor, on the other hand, is a special, non-publicly disclosed method that likewise allows someone to take control of the system. Digital platforms would have to develop and modify their existing systems to offer these capabilities. And this poses serious risks to all owners of private information.

According to our research, there are risks that can arise and lead to unintended and harmful consequences for citizens and the digital platforms that manage their personal data. These risks to the business include potential litigation costs, due to increased chances of false accusations and data leaks, as well as damage to the company's reputation and customer trust. Other risks involve increased cybersecurity costs, as well as ethical responsibility for contributing to the erosion of trust in confidentiality, freedom of expression, and the credibility of digital economies and businesses.

Increase in privacy violations

The whole purpose of end-to-end encryption is to ensure that no one, not even the digital platform, can read the message other than the sender and receiver. To perform CSAM (child sexual abuse material)

scanning, this privacy protection needs to be removed. At the very least, suspicious photos – regardless of whether they are later found to be legitimate or not – selected by the scanning system should be reviewed by people, possibly several, who have not been authorized by the content owners. This is an immediate violation of privacy.

In addition, there is no guarantee about who will see what, since the security provided by end-to-end encryption ceases to exist. There have been cases of company employees with access to personal information who have taken advantage of the situation. One example involved Samantha de Jong (also known as "Barbie"), a 28-year-old Dutch reality TV star, who was urgently admitted to the **Haga Hospital** in the Netherlands. An insider revealed that several hospital employees abused their position by repeatedly accessing the celebrity's medical record. A detailed investigation conducted by the hospital confirmed that more than 85 employees violated the patient's privacy by illegally accessing her medical history through the internal system called *Chipsoft.*

Situations like this occur all over the world. In Australia, it was recently discovered that there were several cases where police forces and other authorities failed to demonstrate that they had followed the law when exercising the power to access data and metadata last year. The number of cases like this is unknown, as they are often not discovered or disclosed. You have no way of knowing who is accessing your private information and why it is being read.

False accusations and reputational damage

Automated detection of illegal photos is not perfect. For example, a Facebook **study** of 150 accounts reported to authorities for alleged

child sexual abuse material found that 75% of those accounts were incorrectly flagged. Another **study** pointed out that if the scan was done only on WhatsApp – where about 4.5 billion images are shared per day – more than a million images per day would be incorrectly identified as child sexual abuse material. Based on the recommendations of an automated scanning system like this, innocent people could be subjected to police investigations and other consequences. In two real-life cases, parents — one in San Francisco and the other in Houston — had young children with genital infections and took photos of the area at the request of health care providers. They were reported to the authorities, underwent a 10-month investigation, and had some of their digital accounts deleted. The stress caused by this type of accusation can be terrifying.

Abuse of power by the government and erosion of freedom of expression

While the stated purpose of these regulations is to search for child pornography material, since governments have the ability to decrypt all of your information, they will be able to use it for any purpose they choose—such as identifying government critics. In this way, they will have unprecedented levels of access to track the personal life and activities of any user, without offering these individuals any choice or requirement of a court order. This is a global concern, as highlighted by the **Freedom Online Coalition**:

"An increasing number of governments have abused digital technologies to restrict access to information and the exercise of human rights and fundamental freedoms. These actions often target journalists, human rights defenders, activists, workers and union

leaders, members of the political opposition, or any other individual perceived as a dissident or critic."

Increased vulnerability to cyberattacks

As digital platform operators increasingly centralize the storage of user data and facilitate the sharing of information between them, cyber threats continue to grow. Encryption works as an essential protection, ensuring data privacy and security by making compromised information unreadable to unauthorized parties. Even in the event of a breach, encrypted data remains protected, reducing the risks associated with cyberattacks and unauthorized access.

Any technology or procedure created by platforms to replace security mechanisms and allow authorities access to information can be stolen and exploited by cybercriminals. A recent breach exploited ***backdoors*** that telecommunications companies had created to legally share information with **law enforcement**. Another example is **EternalBlue**, a cyberattack tool developed by the NSA to accumulate and weaponize digital security vulnerabilities. In 2017, the software was stolen by cybercriminals, who began using it in attacks on high-profile targets, such as the city of Baltimore. The same type of theft of security override mechanisms can happen with CSAM scanning technology.

The above concerns could even be considered acceptable if there was a guarantee that all child sexual abuse would be eradicated. However, as mentioned earlier, recognition techniques are not foolproof. Not only do they mistakenly flag legal content, but they also often fail to identify illegal content. In fact, there are ingenious techniques that offenders can utilize to disguise their illegal content. One study

showed that small modifications to images caused 99.9% of them to go unnoticed by the most widely used scanning algorithms.

While the goal of preventing child sexual abuse is undeniably worthy, policymakers need to carefully weigh the benefits and consequences of any new rules or laws. It is essential to assess the risks generated by the large-scale scanning of private content and the impact on individuals, their privacy and security, as well as on society as a whole.

Rather than creating mechanisms that circumvent the protections of end-to-end encryption, it may be more effective to strengthen existing laws and regulations. This includes increasing prison sentences for offenders, imposing fines on platforms that do not respond adequately to reports, expanding public awareness campaigns, and widely publicizing Child Sexual Abuse Material (CSAM) reporting channels. By recognizing that most of these activities do not occur in isolation, we can all contribute to prevention—after all, as the saying goes, "If you see something, say something."

Crypto strategies

Policymakers around the world have expressed support for encryption technologies as a means of preserving the inviolability of citizens' personal data. Through these technologies, individuals' personal data is protected, thus functioning as a fundamental foundation to ensure freedom of expression and privacy in day-to-day transactions. The U.S. Department of Justice's Office of Public Affairs has stated that it supports "... robust encryption, which plays a crucial role in

protecting personal data, privacy, intellectual property, trade secrets, and cybersecurity." ⊠

Commonly, end-to-end encryption turns user content — such as an image, video, or text message — into a meaningless string of characters (usually through "encryption keys," also known as public keys), which can later be decrypted back to its original, "readable" form (through "decryption keys," also known as private keys). Cryptography, which uses mathematical cryptography (in the technical sense), has already been proven to be extremely difficult — essentially impossible — to crack. End-to-end encryption is one of the strongest forms of protection precisely because the decryption keys stay only with the users, allowing only the intended recipients to be able to access the encrypted content. However, if a master key or backdoor is created, it defeats the end-to-end encryption. The large-scale implementation of private content scanning would require digital platforms' scanning algorithms to have access to all user content in unencrypted form in order to scan it for illicit material. This could occur on the platforms' servers, through decryption of the content with a master key (server-side scanning), or directly on users' devices, prior to encryption, through a backdoor (on-device scanning).

Figure 1 illustrates how the introduction of scanning technologies would circumvent common encryption methods by creating a master key or a backdoor. In this example, the encryption takes place within the confines of the platform itself. If scanning technologies were to operate within these limits, communications would need to be decrypted within the platform so that they could be scanned for illicit content, which would lead to potential unintended consequences and risks already discussed earlier.

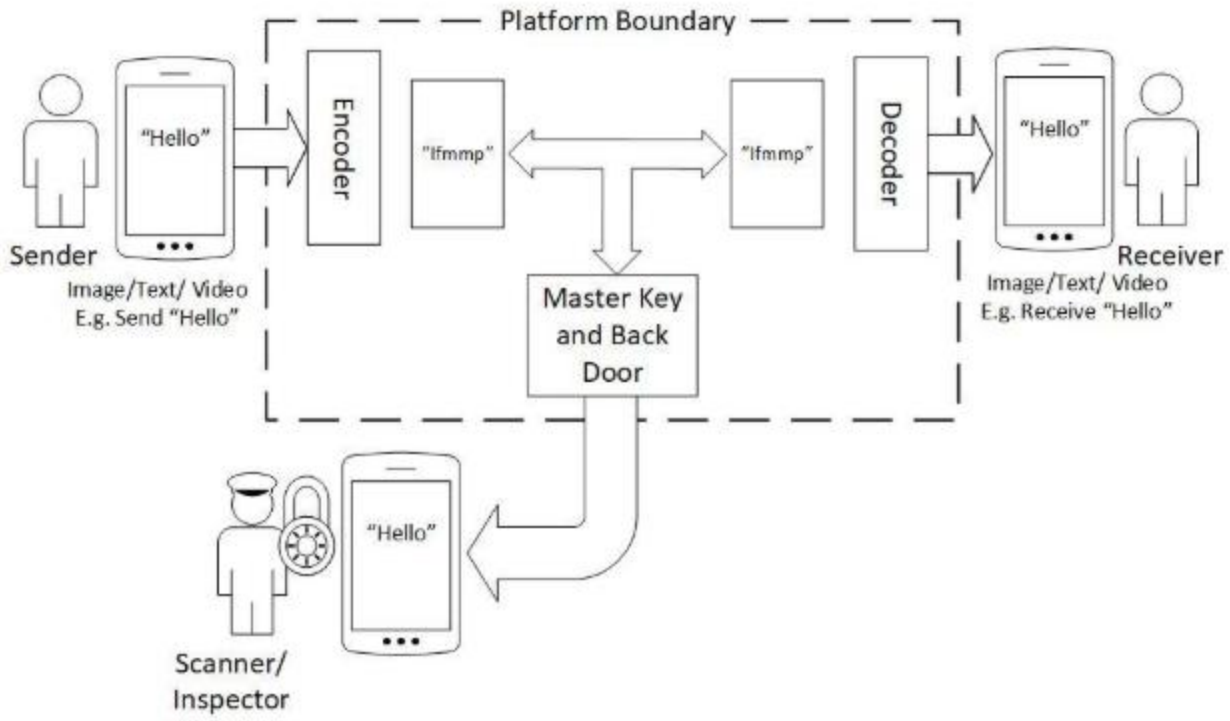

Figure 1: End-to-end encryption with introduction of master keys or backdoors

There is a relatively simple way for CSAM (Child Sexual Abuse Material) abusers to circumvent this type of scanning, as illustrated in Figure 2. In this method, a dual form of encryption is employed. First, the communication is encrypted while still within the limits of the user's device, before being sent to the platform. There are several publicly available — and often free — encryption packages. When the encrypted message arrives on the platform, it is re-encrypted, but it cannot be fully decrypted through the master keys or backdoors introduced by content scanning technologies, as it had already been encrypted before reaching the platform. When the message is sent to the recipient, it is partially decrypted when it leaves the platform, and then completely decrypted in a readable communication as soon as it reaches the recipient's device.

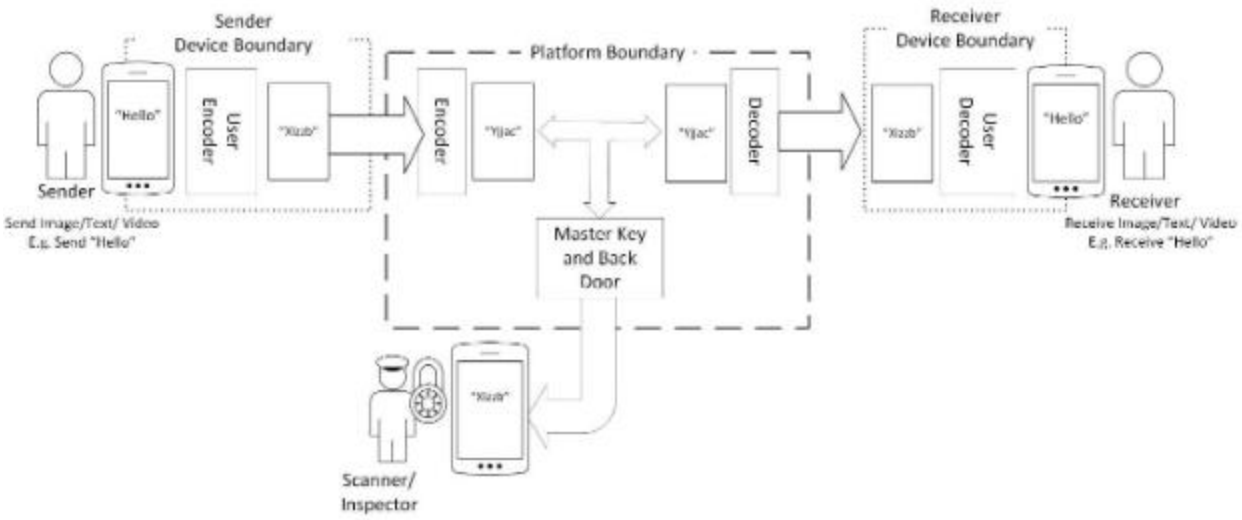

Figure 2: Double encryption on users' devices and platforms

How the Double Encryption Strategy Can Be Used to Prevent Damage from CSAM Scans

While the approach presented in Figure 2 was introduced to illustrate how abusers could use double encryption to avoid detection, it can also be employed by legitimate data users as a way to protect themselves from the potential damage caused by CSAM scans, as discussed earlier. It is important to understand that cybercriminals often target entities that hold sensitive information, as this data has high value in the underground market.

In addition, the adoption of user-controlled encryption approaches has also been considered in a relevant way as a means of protecting data security in the future in the face of emerging technologies, such as quantum computing.

Final thoughts and recommendations

Our research raises important questions about how organizations can protect personal data against a backdrop of increasing challenges such as regulatory restrictions, privacy concerns, and security vulnerabilities.

To remain resilient and secure in the face of future risks, organizations must review their data encryption management strategies. Many still treat encryption as a secondary measure, applying it only when required by regulations. However, leaders must consider encryption approaches at every stage of the data lifecycle, from collection to analysis. As a first step, managers should understand the existing and proposed regulatory obligations that affect crypto systems. From this, strategies can be developed that make it clear how personal data is shared securely. Organizations that proactively communicate their data security and encryption policies—and demonstrate their robustness—can turn privacy into a competitive advantage.

AUTHORS

Dr. Stuart E. Madnick is John Norris Maguire Professor (1960) of Information Technology at the MIT Sloan School of Management, Professor of Systems Engineering at the MIT School of Engineering, and Founding Director of the MIT Sloan Cybersecurity Center (CAMS).

Dr. Daniel Gozman is an Associate Professor at the University of Sydney Business School (Australia) and an Honorary Fellow at the Henley Business School of the University of Reading (UK).

ACKNOWLEDGMENTS

This research was supported, in part, by resources from members of the Cybersecurity at MIT Sloan (CAMS) consortium.

Content scanning and privacy at risk

Companies are under pressure to scan digital content in the name of security, but these requirements can compromise end-to-end encryption and open the door to privacy breaches.

Conflict between regulation and data protection

Laws against online child abuse, such as the Earn It Act and EU proposals, require access to encrypted messages. This can force platforms to implement backdoors and skeleton keys, opening loopholes for cyberattacks and abuses of power.

Risks for businesses and consumers

In addition to potential false positives and reputational damage, organizations can face litigation, loss of consumer trust, and increased information security costs.

Encryption as a right and strategy

Experts from MIT and international universities reinforce: encryption must be treated as a pillar of digital security. Adopting robust strategies from data collection to analysis can turn privacy into a competitive advantage.

"*Collaborative article with **Daniel Gozman**, Associate Professor at The University of Sydney Business School*

By: Stuart Madnick
Stuart Madnick Professor of Information Technologies, MIT Sloan School of Management and Founding Director of Cybersecurity at MIT Sloan (CAMS)."