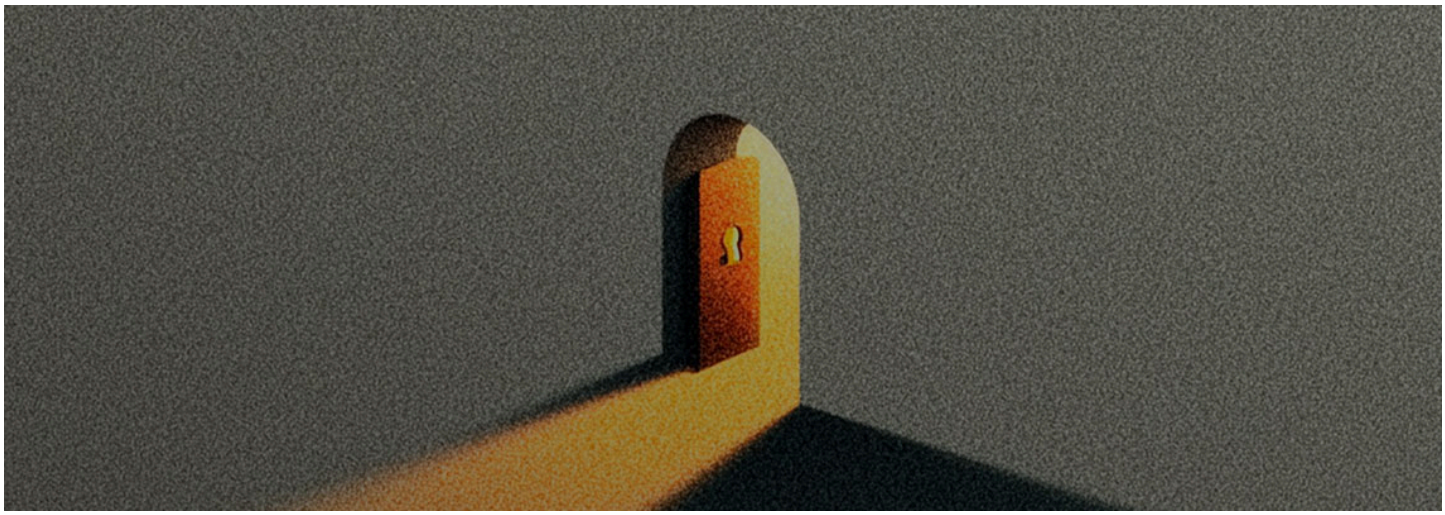
MIT Technology Review
Publicado por TEC



HUMANS AND TECHNOLOGY

The Hidden Cyber Risks of Well-Intentioned Regulations

by **Stuart Madnick Daniel Gozman** August 12, 2025

Download

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

Across industries and governments, there's a growing consensus that data privacy isn't just a technical issue, but a fundamental digital right. In response, leaders have pushed for more robust regulatory frameworks and stronger safeguards to protect consumer data, build trust, and combat illegal activity. But sometimes these two goals conflict. So, what should we do?

Tensions between scanning and personal data protection

A clear example of the potential for well-intentioned laws to lead to unintended consequences is the global efforts to eliminate child pornography online. This is almost a consensus. Thus, governments around the world have been proposing or implementing regulations requiring technology companies to seek out and

MIT Technology Review
Publicado por TEC

focused on protecting minors and combating online exploitation. Brazilian law criminalizes the production, possession, and dissemination of CSAM, with penalties provided for in the Penal Code. Furthermore, laws such as the Artificial Intelligence Law and the General Data Protection Law (LGPD) help 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 the STOP CSAM** in the United States, which propose holding 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 text 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 could have severe side effects on the security and privacy of millions of citizens, with significant implications for companies that handle personal data.

One of the key points is end-to-end encryption, a technology that encodes messages so that only the sender and recipient can decode them. Not even the digital platform can read the content. This technology is essential for ensuring the confidentiality of communications, user privacy, and freedom of expression. Furthermore, it protects data even if intercepted by cybercriminals. In the US, the use of apps like Signal, which use end-to-end encryption, has received considerable media attention. The technology works perfectly; problems usually arise from misuse, not the technology itself.

But if digital platforms are held accountable for illegal photos shared within 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 the old-fashioned key that opens any door, or the "fireman's key" found in many elevators, allowing you to take control. A backdoor, on the other hand, is a special method, not publicly disclosed, that similarly 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 for 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 business risks include potential litigation costs due to the increased likelihood of false accusations and data leaks, as well as damage to the company's reputation and customer trust. Other risks include increased cybersecurity costs, as well as the ethical responsibility of contributing to the erosion of trust in confidentiality, freedom of expression, and the credibility of digital economies and companies.

read the message beyond the sender and recipient. To scan for CSAM (child sexual abuse material), this privacy protection must be removed. At the very least, suspicious photos—regardless of whether they are later determined to be legitimate—selected by the scanning system must be reviewed by people, possibly multiple people, who have not been authorized by the content owners. This constitutes an immediate violation of privacy.

Furthermore, there's no guarantee of who will see what, since the security provided by end-to-end encryption no longer exists. There have been cases of company employees with access to personal information who took 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 **Haga Hospital** in the Netherlands. A whistleblower revealed that several hospital employees abused their position by repeatedly accessing the celebrity's medical records. A detailed investigation conducted by the hospital confirmed that more than 85 employees violated the patient's privacy by illegally accessing her medical records 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 in which police forces and other authorities failed to demonstrate compliance with the law when exercising their power to access data and metadata last year. The number of such cases is unknown, as they often go undiscovered and unreported. 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 isn't perfect. For example, a Facebook **study of 150 accounts reported to authorities for alleged child sexual abuse material found that 75% of these accounts were incorrectly flagged. Another study** found that if the scan were performed solely on WhatsApp—where approximately 4.5 billion images are shared daily—more than a million images per day would be incorrectly identified as child sexual abuse material. Based on the recommendations of such an automated scanning system, 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 professionals. They were reported to 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.

Government abuse of power and erosion of freedom of expression

While the stated purpose of these regulations is to search for child pornography, once governments have the ability to decrypt all of your information, they can use it for any purpose they choose—such as identifying government critics. This will give them unprecedented levels of access to track the personal lives and activities of any user, without offering them any choice or requiring a court order. This is a global concern, as highlighted by the **Freedom Online Coalition** :

"A growing 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

MIT Technology Review
Publicado por TEC

Increased vulnerability to cyberattacks

As digital platform operators increasingly centralize user data storage and facilitate information sharing among users, cyberthreats continue to grow. Encryption serves as an essential safeguard, 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 override security mechanisms and allow authorities to access information can be stolen and exploited by cybercriminals. A recent breach exploited ***backdoors*** that telecommunications companies had created to legally share information with **authorities** . 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 then used it in attacks on high-profile targets, such as the city of Baltimore. The same type of theft of security override mechanisms can occur with CSAM scanning technology.

The above concerns might even be considered acceptable if there were a guarantee that all child sexual abuse would be eradicated. However, as mentioned previously, recognition techniques are not infallible. Not only do they erroneously flag legal content, but they also frequently fail to identify illegal content. In fact, there are ingenious techniques that offenders can use to disguise their illegal content. One study showed that small modifications to images resulted in 99.9% of them going undetected by the most commonly used scanning algorithms.

While the goal of preventing child sexual abuse is undeniably worthwhile, policymakers need to carefully weigh the benefits and consequences of any new rules or laws. It's crucial to assess the risks posed by 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 fail to 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."

Encryption strategies

Formuladores de políticas em todo o mundo têm expressado apoio às tecnologias de criptografia como um meio de preservar a inviolabilidade dos dados pessoais dos cidadãos. Por meio dessas tecnologias, os dados pessoais dos indivíduos são protegidos, funcionando assim como um alicerce fundamental para garantir a liberdade de expressão e a privacidade nas transações do dia a dia. O Escritório de Assuntos Públicos do Departamento de Justiça dos Estados Unidos declarou que apoia "... criptografia robusta, que desempenha um papel crucial na proteção de dados pessoais, privacidade, propriedade intelectual, segredos comerciais e cibersegurança."

volta à sua forma original e "legível" (por meio de "chaves de descriptografia", também conhecidas como chaves privadas). A criptografia, que utiliza a criptografia matemática (no sentido técnico), já foi comprovada como extremamente difícil — essencialmente impossível — de ser quebrada. A criptografia de ponta a ponta é uma das formas mais fortes de proteção justamente porque as chaves de descriptografia ficam somente com os usuários, permitindo que apenas os destinatários pretendidos consigam acessar o conteúdo criptografado. No entanto, se uma chave-mestra ou uma porta dos fundos for criada, isso anula a criptografia de ponta a ponta. A implementação em larga escala da varredura de conteúdo privado exigiria que os algoritmos de varredura das plataformas digitais tivessem acesso a todo o conteúdo do usuário em forma não criptografada, a fim de escaneá-lo em busca de material ilícito. Isso poderia ocorrer nos servidores das plataformas, por meio da descriptografia do conteúdo com uma chave-mestra (varredura no servidor), ou diretamente nos dispositivos dos usuários, antes da criptografia, por meio de uma porta dos fundos (varredura no dispositivo).

A Figura 1 ilustra como a introdução de tecnologias de varredura contornaria os métodos comuns de criptografia por meio da criação de uma chave-mestra ou de uma porta dos fundos. Neste exemplo, a criptografia ocorre dentro dos limites da própria plataforma. Se as tecnologias de varredura passassem a operar dentro desses limites, as comunicações precisariam ser descriptografadas dentro da plataforma para que pudessem ser escaneadas em busca de conteúdo ilícito, o que levaria a potenciais consequências não intencionais e riscos já discutidos anteriormente.

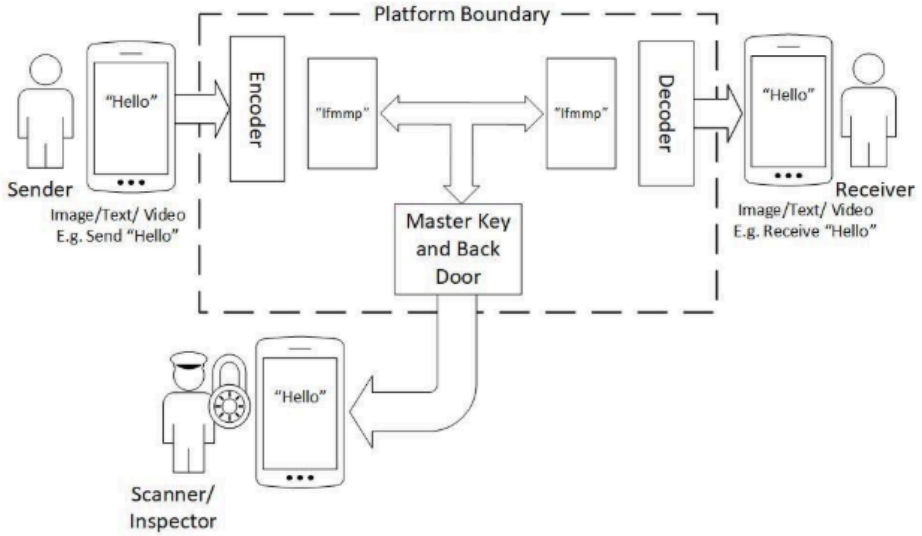


Figura 1: Criptografia de ponta a ponta com introdução de chaves-mestras ou portas dos fundos

Há uma maneira relativamente simples de abusadores de CSAM (Material de Abuso Sexual Infantil, na sigla em inglês) contornarem esse tipo de varredura, como ilustrado na Figura 2. Nesse método, emprega-se uma forma dupla de criptografia. Primeiramente, a comunicação é criptografada ainda dentro dos limites do dispositivo do usuário, antes de ser enviada à plataforma. Existem diversos pacotes de criptografia disponíveis publicamente — e, muitas vezes, gratuitos. Quando a mensagem criptografada chega à plataforma, ela é criptografada novamente, mas não pode ser totalmente descriptografada por meio das chaves-mestras ou portas dos fundos introduzidas pelas tecnologias de varredura de conteúdo, já que já

MIT Technology Review
Publicado por TEC

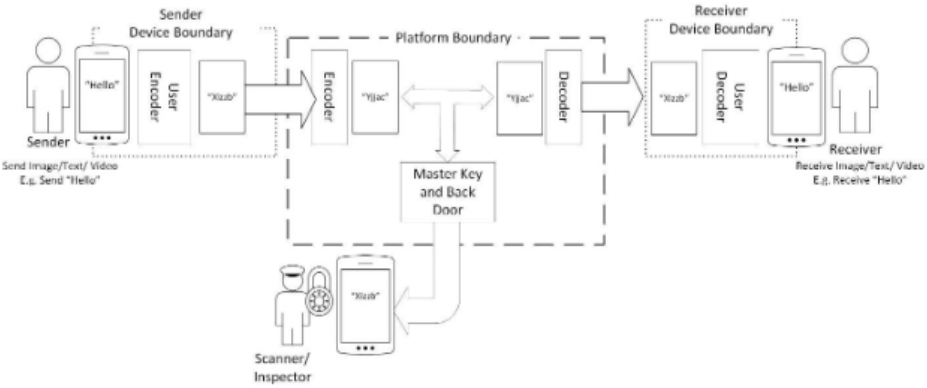


Figura 2: Dupla criptografia em dispositivos dos usuários e nas plataformas

Como a estratégia de dupla criptografia pode ser usada para evitar danos causados por varreduras de CSAM

Embora a abordagem apresentada na Figura 2 tenha sido introduzida para ilustrar como abusadores poderiam utilizar a dupla criptografia para evitar a detecção, ela também pode ser empregada por usuários legítimos de dados como uma forma de se proteger dos danos potenciais provocados pelas varreduras de CSAM, conforme discutido anteriormente. É importante compreender que cibercriminosos frequentemente têm como alvo entidades que detêm informações sensíveis, uma vez que esses dados possuem alto valor no mercado clandestino.

Além disso, a adoção de abordagens de criptografia controladas pelo usuário também vem sendo considerada de forma relevante como um meio de proteger a segurança dos dados no futuro frente a tecnologias emergentes, como a computação quânticai.

Considerações finais e recomendações

Nossa pesquisa levanta questões importantes sobre como as organizações podem proteger os dados pessoais em um cenário de desafios crescentes, como restrições regulatórias, preocupações com a privacidade e vulnerabilidades de segurança.

Para se manterem resilientes e seguras diante dos riscos futuros, as organizações devem revisar suas estratégias de gerenciamento de criptografia de dados. Muitas ainda tratam a criptografia como uma medida secundária, aplicando-a apenas quando exigido por regulamentações. No entanto, os líderes devem considerar abordagens de criptografia em todas as etapas do ciclo de vida dos dados, da coleta à análise. Como primeiro passo, os gestores devem compreender as obrigações regulatórias existentes e propostas que afetam os sistemas de criptografia. A partir disso, podem ser desenvolvidas estratégias que deixem claro como os dados pessoais são compartilhados de forma segura. Organizações que comunicam proativamente suas políticas de segurança de dados e criptografia — e demonstram sua robustez — podem transformar a privacidade em uma vantagem competitiva.

AUTORES

MIT Technology Review
Publicado por TEC

Dr. Daniel Gozman é Professor Associado na University of Sydney Business School (Austrália) e Pesquisador Honorário (Honorary Fellow) na Henley Business School da University of Reading (Reino Unido).

AGRADECIMENTOS

Esta pesquisa foi apoiada, em parte, por recursos provenientes dos membros do consórcio Cybersecurity at MIT Sloan (CAMS).

🔐 Escaneamento de conteúdo e privacidade em risco

Empresas estão sob pressão para escanear conteúdos digitais em nome da segurança, mas essas exigências podem comprometer a criptografia de ponta a ponta e abrir portas para violações de privacidade.

⚖️ Conflito entre regulação e proteção de dados

Leis contra o abuso infantil online, como o Earn It Act e propostas da UE, exigem acesso a mensagens criptografadas. Isso pode forçar plataformas a implementar backdoors e chaves-mestras, abrindo brechas para ciberataques e abusos de poder.

📉 Riscos para empresas e consumidores

Além de possíveis falsos positivos e danos à reputação, organizações podem enfrentar litígios, perda de confiança do consumidor e aumento nos custos com segurança da informação.

🔒 Criptografia como direito e estratégia

Especialistas do MIT e universidades internacionais reforçam: a criptografia deve ser tratada como pilar da segurança digital. Adotar estratégias robustas desde a coleta até a análise dos dados pode transformar a privacidade numa vantagem competitiva.

Autor

Stuart Madnick Daniel Gozman

Compartilhar