MEDIA LITERACY AND DIGITAL SECURITY:

THE IMPORTANCE OF STAYING SAFE AND INFORMED

OAS | More rights for more people

CONTENT

CREDITS

Twitter Technical Team

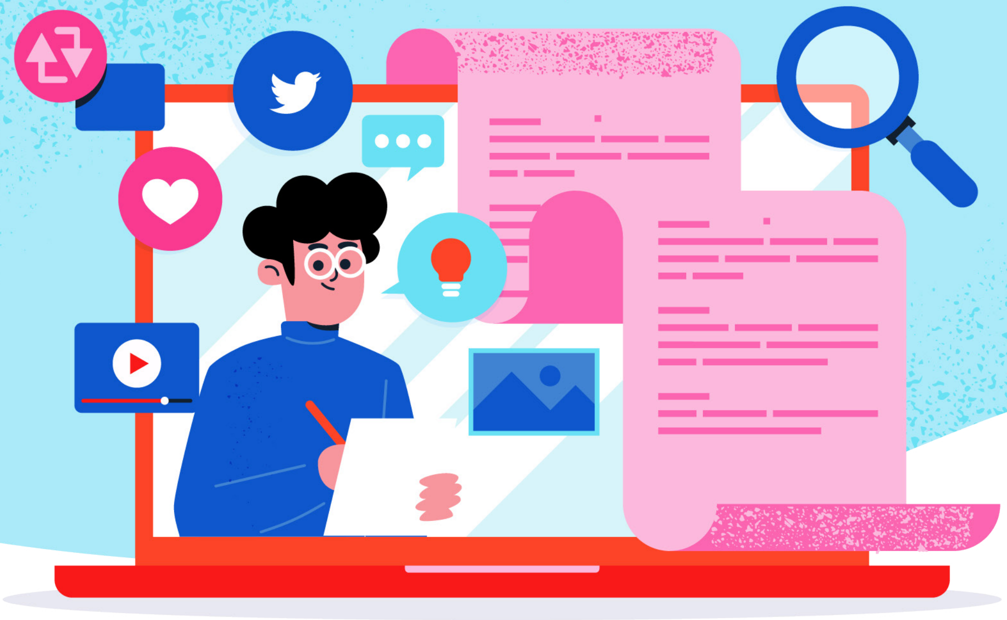
Andrea Pereira Palacios
Hugo Rodríguez Nicolat

OAS Technical Team

Alison August Treppel
Kerry-Ann Barrett
Gerardo De Icaza
Gonzalo Espáriz
Cristóbal Fernández
Mariana Jaramillo
Yerutí Méndez
Gabriela Montes de Oca
David Moreno
María Isabel-Rivero
Diego Subero
Katya Vera Morales

Design and Layout

Michelle Felguérez



INTRODUCTION

In the digital and social media age, information is abundantly and immediately available, allowing us to keep abreast of what is happening in the world instantaneously. The digitalization and transformation of everyday processes are both rapid and unstoppable.

Receiving and processing this wealth of information requires certain skills and an understanding of the media in which it circulates. It is important to know not only the origin, intention, or purpose of information consumed and published, but also the possible risks and the impact they may have on our environment.

Against this backdrop, **Twitter**, and the **Organization of American States (OAS)** have updated this publication on Media Literacy and Security, to provide tools and present best practices for monitoring, consuming, and sharing information, and recommendations for staying safe online, focusing, particularly, on Twitter. In this updated edition of the guide *"Media Literacy and Security: Twitter Best Practices,"*[1] published in September 2019, we identify additional phenomena related to disinformation and discuss the importance of literacy in democratic processes.

Since the launch of the first edition of this guide, the world has seen an exponential growth in internet activity. According to studies by the United Nations Economic Commission for Latin America and the Caribbean (ECLAC), advances in the use of communications networks and infrastructure that were expected to take years to materialize have occurred in just a few months since 2020[2]. ECLAC has also highlighted the need to develop digital skills as a key condition for taking advantage of the internet.

[1] Organization of American States and Twitter. (2019). Media Literacy and Digital Security: Twitter Best Practices. https://www.oas.org/en/sms/cicte/docs/20190913-DIGITAL-ENG-Alfabetismo-y-seguridad-digital-Twitter.pdf
[2] Economic Commission for Latin America and the Caribbean (ECLAC). Special Report COVID-19 (2020). Universalizing access to digital technologies to address the consequences of COVID-19. https://www.cepal.org/sites/default/files/publication/files/45939/S2000549_en.pdf

Both the OAS and Twitter have observed several changes in this scenario. In Latin America, over 30% of companies detected an increase in the number of cyberattacks in 2019 compared to previous years, although only 17% have cyber risk insurance[3]. This demonstrates the importance of continuing to foster awareness of the digital threats that exist in the region and how to combat them at all levels. As a public and open platform for the exchange of perspectives, ideas, and information, Twitter is constantly updating its rules, processes, tools, and technology to adapt to changes in society and the way people interact and share information on its service.

The first section of the publication focuses on defining media literacy, referring to the work by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the Inter-American Commission on Human Rights (IACHR). These approaches reflect the development of the term "literacy" and its relationship to technology, as well as regional considerations of media literacy. In addition, we have included a specific section prepared by the Department of Electoral Cooperation and Observation (DECO) of the OAS on the relationship between media literacy and democracy.

The second section is related to best practices in cybersecurity, presenting information on new cyber threats that have emerged since the first publication of this guide. Specific recommendations are also included in response to the increase in teleworking or remote work conditions and journalistic work in the region.

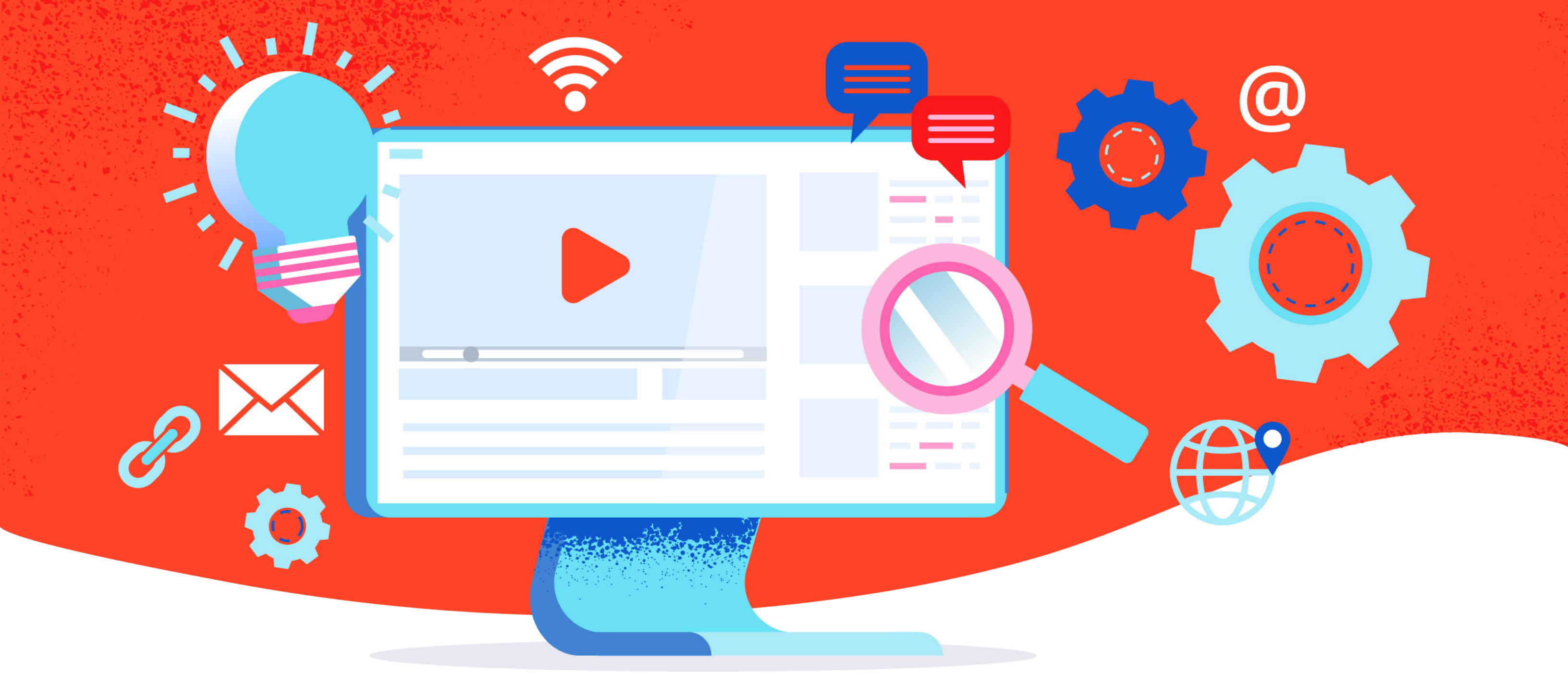
The third section contains expanded and updated information regarding information consumption on Twitter and tips on how to check its veracity. It also explains some tools available to better navigate the platform and find and verify information quickly and easily.

Finally, given the importance of Twitter as a communication tool, **the last section o**f the guide contains an update on the Twitter Rules and their enforcement, to provide information on the parameters governing the circulation of information and interactions on the platform. This section also explains Twitter's security tools and how to use them to have a more personalized and controlled experience on the platform.

Technology and the tools available for its use are constantly evolving, so we encourage everyone to be vigilant about product and policy updates that affect their digital media and social media interactions.

HAVING THE SKILLS TO USE THE INTERNET SAFELY IS ESSENTIAL TO COUNTER CYBERATTACKS AND OTHER DISINFORMATION MECHANISMS THAT ARE INCREASINGLY SOPHISTICATED AND COMPLEX.

[3] Marsh & Microsoft. (2020). Estado de Riesgo Cibernético en Latinoamérica en Tiempos del COVID-19 [Cyber Risk in Latin America during COVID-19]. https://coronavirus.marsh.com/mx/es/insights/research-and-briefings/report-cyber-risk-in-latin-america-in-times-of-covid19.html



MEDIA LITERACY

What is Media Literacy?

According to UNESCO, literacy is defined as

> a means of identification, understanding, interpretation, creation, and communication in an increasingly digital, text-mediated, information-rich and fast-changing world[4].

This definition, officially coined in recent years, considers digital skills to be a fundamental component of literacy, as opposed to previous notions that only consider the set of reading, writing, and numeracy skills.

The "Media and Information Literacy for Sustainable Development Goals" section of the 2016 UNESCO Yearbook contains the "Five Laws of Media and Information Literacy," which are:

[4] United Nations Educational, Scientific, and Cultural Organization (UNESCO). (2016). Literacy. https://en.unesco.org/themes/literacy

1. Information, communication, libraries, media, technology, the Internet and other sources of information providers fall into the same category. Neither is more relevant than the other nor should be treated as such.

2. Every citizen is a creator of information or knowledge and has a message. They must be empowered to access new information and express themselves.

3. Information, knowledge and messages are not always value neutral, or independent of biases. Any conceptualization, use and application of media literacy must make this truth transparent and understandable to all citizens.

4. Every citizen wishes to know and understand new information, knowledge and messages, as well as to communicate, and his/her rights should never be compromised.

5. Media literacy is a lived and dynamic experience and process. It is complete when it includes knowledge, skills and attitudes, when it covers access, evaluation, use, production and communication of information, media and technology content.

These concepts are key to understanding the importance of acquiring digital tools for the appropriate use of the internet by any person.

How do we achieve Media Literacy?

Acquiring digital tools directly affects our level of literacy. The UNESCO yearbook mentioned above sets out ten skills that should be developed to achieve media literacy[5]. These skills are:

1. Engage with information, media, and technology.

2. Be able to apply information and communications technology skills in order to process information and produce user-generated content.

3. Ethically and responsibly use information and communicate their understanding or newly created knowledge to an audience or readership in an appropriate form and medium.

4. Extract and organize information and media content.

5. Critically evaluate information and the content of media and other information providers, including those on the internet, in terms of authority, credibility, current purpose, and potential risks.

[5] Grizzle, A and Singh, J. (2016). In the MILID Yearbook 2016: Media and Information Literacy for the Sustainable Development Goals.

6. Locate and access relevant information and media content.
7. Synthesize or operate on the ideas abstracted from content.
8. Understand the conditions under which those functions can be fulfilled.
9. Understand the role and functions of media and other information providers, including those on the internet, in democratic societies and development.
10. Recognize and articulate a need for information and media.

Media Literacy to combat Disinformation

Considering the digital skills outlined in the previous section, a highly relevant issue related to media literacy is the existence of disinformation on the internet. In its Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts, the Office of the Special Rapporteur for Freedom of Expression (RFOE) of the IACHR defines disinformation, in practical and provisional terms, as the mass dissemination of false information:

a. with the intent to deceive the public; and
b. with the knowledge of its falsehood.

Although this phenomenon is not new, recent technological developments have allowed it to be reproduced at an accelerated rate, reaching a larger number of people and with consequences in various spheres of public life[6]. The existence of disinformation highlights the need to develop the ability to evaluate information critically; if information lacks authority, credibility, and purpose, and is spread in the same way, its impacts can be significant.

In its guide, the RFOE underscores the potential impact of disinformation in electoral contexts and outlines several recommendations for the different stakeholders involved in elections. These focus on helping stakeholders address disinformation issues and the potential side effects that could negatively affect human rights standards.

Some of these recommendations on disinformation in electoral processes[7], that focus largely on the need to contribute to media literacy and security include:

For States, including the different branches of government and electoral authorities:

- Avoid establishing regulatory frameworks that hold intermediaries responsible for content produced by third parties.
- Strengthen the legal frameworks for personal data protection.
- Remember the special responsibilities that senior public officials have in exercising their own freedom of expression.
- Carry out positive educational, training, and awareness-raising actions for citizens, to strengthen their ability to dismantle disinformation campaigns in electoral contexts.

[6] Organization of American States. (2019). Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts. https://www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf
[7] The complete list of recommendations is in the Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts, pp. 30-51.

For intermediary companies:

- Make transparent the criteria used to moderate, detect and prioritize content on platforms and ensure due process in content moderation.
- Support quality journalism and take other positive actions aimed at counteracting disinformation campaigns, including cooperating with electoral authorities and independent investigators.
- Respect and proactively comply with the protection of personal data.

For political parties:

- Avoid campaigns that use false information.
- Make election campaigns transparent.
- Respect and proactively comply with the protection of personal data.

For journalists and the media:

- Strengthen quality journalism against disinformation.

For fact-checkers:

- Standardize definitions of disinformation and strengthen regional networks.

For academic and research institutes:

- Expand empirical research on disinformation.

Throughout this guide, we discuss basic concepts on how to strengthen skills essential for achieving media literacy and combating disinformation and other phenomena that can undermine the safe use of digital technology.

IMPORTANCE OF MEDIA LITERACY FOR DEMOCRACY
ELECTORAL OBSERVATION MISSION OF THE OAS (DECO)

Technological platforms and social networks have amplified and created new modes of communication that have strengthened the interaction between representatives and citizens, complementing the more traditional forms of political participation. This has helped to produce an active citizenry involved in open debate about the ideas and interests that converge in the public sphere.

Today, people have various means at their disposal to express their opinions and positions and to become informed about events taking place around them, whether at the municipal, departmental, or national level, or about events in other countries and continents. There is also the opportunity to create content, which can be disseminated and shared with a wide audience. There is also more direct interaction with those who hold public office, which brings not only access but also oversight, thus subjecting the exercise of power to constant and immediate public scrutiny more than ever before.

It is now common to visit the social media accounts of government authorities to learn about the performance of their duties, or of candidates to learn about their proposals directly and to keep up to date on campaign activities. Political parties benefit from the ability to transmit their messages and political perspectives directly to citizens through the various media available, expanding their territorial scope to a national level. Civil society today has platforms with massive reach to publicize its actions and connect with people and citizens today have new forms of organization that make collective expression possible.

These factors allow the digital environment to be an important link in democratic processes and elections, contributing to the exercise of freedom of expression, free access to information, and freedom of association. It also promotes transparency and accountability. Going forward, the impact of digital platforms and other information technologies on our democracies will be ever increasing. The COVID-19 pandemic intensified this process. However, while there are important benefits to this, there are also several risks. The dynamics of the digital world exposes all people who use the internet to phenomena such as disinformation, personal data breaches, illegal activities, and the influence of external actors in domestic politics or electoral processes, which directly or indirectly undermine confidence in our democracies.

To fully use the opportunities offered by the digital world, and to know, understand, and protect oneself from the risks arising from this environment with which we are increasingly engaged, it is important that media literacy reaches everyone.

This is an essential element for strengthening democracy, and a necessary process so that all those involved in the public sphere, social actors, interest groups, institutions, the media, civil society, political parties, and citizens, will have the skills and opportunities to make use of the tools we have today to contribute to an active and responsible democracy through technology.

Learning to use technological tools, accessing information, distinguishing information from disinformation, evaluating it, being critical in its analysis, differentiating between sources, protecting data, and safeguarding privacy are some of the necessary conditions for bridging the digital divide, promoting safe interaction with technologies, and educating for a conscious and informed citizenry.

CYBERSECURITY AND DIGITAL SELF-CARE

THE INTERNET HAS BEEN A TOOL THAT HAS TRANSFORMED AND DEFINED COMMUNICATION IN THE 21ST CENTURY.

Its multiple uses have enabled individuals and organizations to connect and communicate with each other. In the wake of various events in recent years, such as the COVID-19 pandemic and the acceleration of digitalization processes around the world, an increasing number of people are using the internet to stay connected and share personal, professional, and social messages and information on numerous platforms.

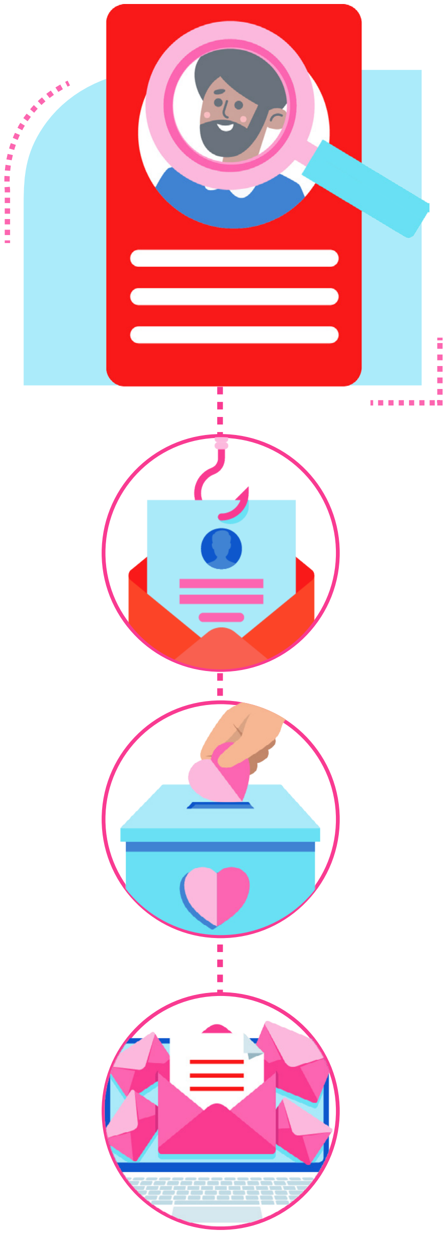
However, as these processes multiply, organizations and institutions worldwide have seen a significant rise in the level of exposure to online risks. This is mainly due to a lack of familiarity with the large-scale use of Information and Communication Technologies (ICT) and a general lack of knowledge of cyber threats and digital security tools. Limited cybersecurity skills and exposure to more online risks have created a scenario in which attackers have taken advantage of the digital "new normal" to exploit novel ways of attacking and accessing personal data (UNODC, 2020)[8].

With this scenario in mind, this section offers a variety of terms and descriptions to familiarize the general public with these forms of attack that have changed over time. It also includes tips to proactively mitigate and counteract them.

[8] Trend Micro. (2020). Developing Story: COVID-19 Used in Malicious Campaigns. https://www.trendmicro.com/vinfo/fr/security/news/cybercrime-and-digital-threats/coronavirus-used-in-spam-malware-file-names-and-malicious-domains

How to recognize Cyberattacks

Although cyberattacks are not new, it is important to be familiar with them to be able to report and counteract them.



Social Engineering

Characteristics: Social engineering is the use of non-technological methods to trick specific potential victims who have been previously investigated into sharing sensitive personal information, such as passwords or bank account details, almost voluntarily with a hacker[9].

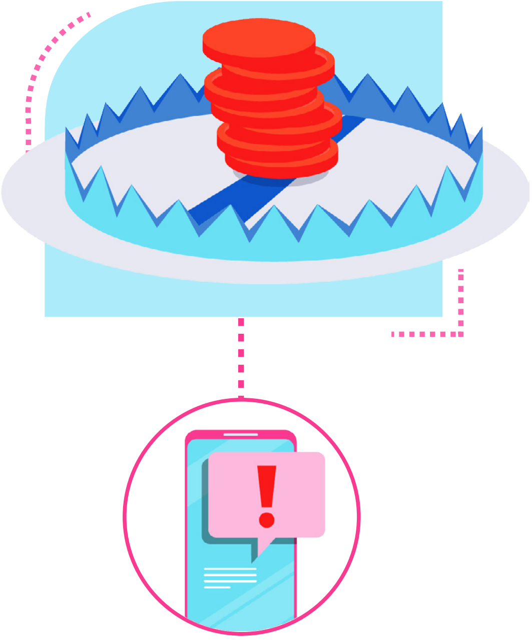
Examples include[10]:

Spear phishing: Through the personalization of emails and phishing messages, or by impersonating close contacts, recruiters, etc.
Example: a hacker who spoofs a bank's information and asks a person for private information to "unlock their account," or who impersonates a recruiter to request identity information to process a fake offer.

Pretexting: Through a captivating pretext or story, hackers attract the attention of a targeted person and engage them to do a specific action, such as donating to a fake campaign or providing sensitive personal information.
Example: emails promising money from a supposed inheritance, in which they seek to obtain bank account details.

Contact spamming: This consists of sending mass emails to a list of contacts of an account that has been hacked. These emails are sent from a known mailbox so as not to arouse suspicion, but the content of the emails will appear to recipients with shortened links or informal subject lines such as "Check this out." If the person clicks on it, malicious software will be installed that will continue the spam chain and may have negative consequences for their personal data.

[9] NortonLifeLock. What is Social Engineering? https://us.norton.com/internetsecurity-emerging-threats-what-is-social-engineering.html
[10] SoftwareLab.org. What is social engineering: Top 5 Types & Examples. https://softwarelab.org/what-is-social-engineering/



Fraud, internet scams, and phishing campaigns

Characteristics: These attacks are carried out on a massive scale through posts or messages on social networks that use unverified information on topics in the news cycle as a lure, persuading recipients to access fake websites, provide personal or banking data, or infect computer systems and electronic devices[11].

Example: phishing campaigns are common, in which cybercriminals impersonate international organizations and government and health authorities, offering information or support for supposed social programs[12], soliciting donations to respond to health emergencies or, given the exponential increase in online shopping, posing as a parcel or home delivery service.

Scams also use "smishing" campaigns, which offer free food, bonuses, medical products, discounts, free reloading services, or subscriptions to entertainment platforms.

Malware or installation of malicious software

Characteristics: Malware or malicious content infiltration uses information related to news topics as a lure to infiltrate electronic devices. In Latin America, a recurrent threat in recent years has been ransomware attacks against personal computers and cell phones, through which cybercriminals encrypt and "kidnap" victims' personal information and data, demanding a ransom to unlock the device or captive information and release the data[13]. Sometimes hackers may also make threats against their targets if they refuse to make the expected payment.

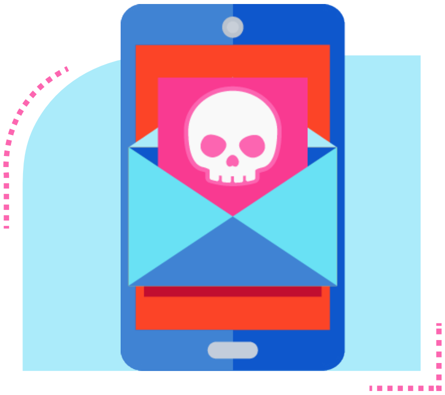
[11] Porter, Taryn. (2020). COVID-19 Scam Alters. Cybercrime Support Network. https://cybercrimesupport.org/covid-19-scam-alerts/

[12] World Health Organization. (2020). WHO reports a fivefold increase in cyberattacks, urges vigilance. https://www.who.int/news/item/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance; BBC News Mundo. (2020). Coronavirus: la advertencia de la OMS sobre los estafadores que están usando el nombre de la organización para robar dinero y datos [Coronavirus: WHO warning about scammers using the organization's name to steal money and data]. https://www.bbc.com/mundo/noticias-52009138

[13] We Live Security & ESET (2020). Threat Report. Q2 2020. https://www.welivesecurity.com/wp-content/uploads/2020/07/ESET_Threat_Report_Q22020.pdf

Creation of fake websites (spoofed domains)

Characteristics: Here, cybercriminals register domain names with keywords that are generally associated with current affairs or topics of public interest to confuse people. This is done to get people to visit these websites, which may offer products that promise to solve a specific problem with "miraculous" or quick solutions, but malicious software is inserted or some information is requested.



Mensajes de sextorsión falsos

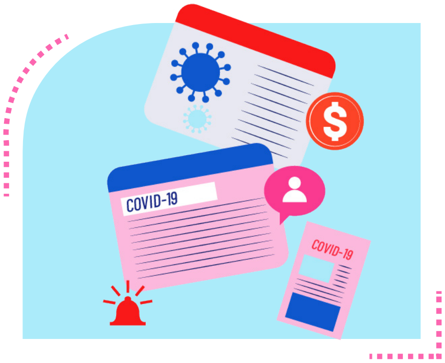
Characteristics: Through this type of attack, hackers send messages to people threatening to send their contacts intimate and compromising videos they have obtained by infiltrating their devices[14] or recorded while browsing sexual websites. This is usually accompanied by a notification to pay a certain amount of money to a digital wallet in exchange for preventing the hacker's threats from being carried out[15].

Ataques a través de las herramientas de trabajo remoto

Characteristics: The COVID-19 pandemic has caused many companies to have a branch office in every employee's home, exposing them to more online risks and exposing their workplace computer systems. In this environment, cybercriminals have identified vulnerabilities in software, networks, and remote work tools, targeting attacks to infiltrate corporate systems through the personal computers of employees.

[14] Duclkin, Paul. (2020). Dirty little secret extortion email threatens to give your family coronavirus. Sophos https://nakedsecurity.sophos.com/2020/03/19/dirty-little-secret-extortion-email-threatens-to-give-your-family-coronavirus/

[15] INCIBE & OSI. (2020). Detectada oleada de falsos correos de sextorsión o infección de COVID19 [Wave of fake sextortion or COVID19 infection emails detected]. https://www.osi.es/es/actualidad/avisos/2020/04/detectada-oleada-de-falsos-correos-de-sextorsion-o-infeccion-de-covid19



Spreading of false information and disinformation

Characteristics: The circulation of false, unverified information or conspiracy theories on the internet facilitates the execution of cyber scams and other cyberattacks[16]. This information can come from various sources, not only from fake accounts, trolls, or bots, but also from official accounts and close contacts, and circulates on the internet because, mostly, it is shared unthinkingly by the public[17].

Use of the Dark Web for criminal activities

Characteristics: There has been an increase in the use of the Dark Web for the sale of personal or corporate information and data obtained via ransomware and other malicious activities, including the sexual exploitation of children, the sale of personal data or email addresses, among other unlawful activities.

Practical Steps for Dealing with Cyberattacks

Given these threats, this section aims to share simple and easy steps to protect personal and corporate information, accounts, and data from cyberattacks. These recommendations are intended for the general public and some may not apply to public figures such as politicians, activists, or other actors whose social media practices are subject to greater scrutiny. The rights of expression, assembly, and protest must also be respected in the digital realm while ensuring safer internet practices.

[16] Stone, Jeff. (2020). How scammers use fake news articles to promote coronavirus 'cures' that only defraud victims. Cyberscoop. https://www.cyberscoop.com/coronavirus-cure-scam-social-media-riskiq/
[17] NewsGuard. https://www.newsguardtech.com/

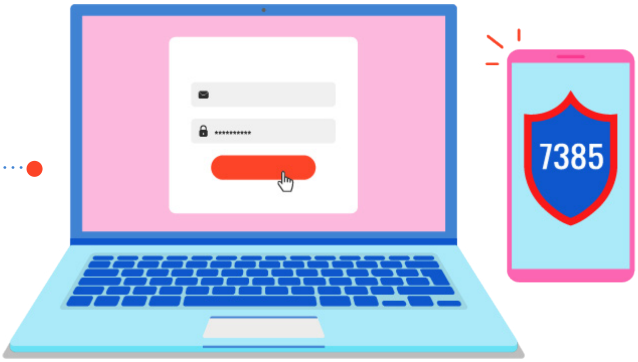
1. Review your Privacy Settings

Managing privacy settings on social networks and other personal accounts is one of the simplest ways people can control the security and privacy of their devices and data. Below are a few recommendations:

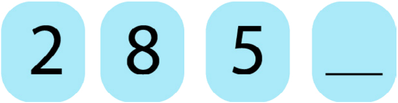
- Regularly check the “Privacy” section in the settings of your social media accounts, email, and internet-connected devices.
- Select who has the ability to view your social media activity (e.g. your Tweets and engagements).
- Check whether your profile is easily accessible or public to others and see how they can connect with you, either through friend requests, or by “following” you on Twitter.
- Review, understand, and determine how much personal information you post online. Remember not to put phone numbers, personal passwords, or sensitive information in your social media posts.
- Periodically monitor the security and login information for your accounts and check for suspicious activity and access by third party applications that may be accessing your personal data.
- Review the platform’s privacy policy to find out what data the services collect and with whom it is shared and select your preferences for these two items.

2. Set Up Two-Factor Authentication to log in to your Personal Accounts

Two-factor authentication provides an additional layer of security by requiring individuals to verify their identity with a second verification factor such as biometrics (fingerprint or face) or by providing a code, thus protecting against the risk of weak or compromised credentials.



Here are the two most popular ways to add this additional layer of security and the pros and cons of each:



Two-factor authentication method	Pros	Cons
 Text message Each time you log into an account through a new device, you will have to enter a multi-number code sent by text message to your phone or by phone call.	- It is a simple way to authenticate the user's identity, as it only requires a phone that can receive text messages. - It is accessible, since sometimes the code can also be sent via a call.	- In cases of identity theft or the loss or theft of your mobile device, another person could gain access to your personal information and log in to your accounts. - There is a practice, called "porting" or "sim swapping," which allows a criminal to exchange the user's sim card for an infected one to intercept two-factor authentication codes and gain access to personal accounts.
 Two-factor authentication app An authentication app is a standalone software application that is downloaded to a smart mobile device (tablet, iPad, etc.) or computer. It generates a random code that is entered after the credentials or sends a push notification (messages or alerts sent from a remote server to the device that has an app installed) to authenticate the person's identity.	- This functionality is available without an internet connection. - It is not susceptible to porting since it does not rely on a telephone chip. - The automatic notification version offers the additional benefit of being faster and easier to use. If the notification states that the approximate location is far from the person's home or office, notifications such as these are more likely to get their attention and encourage them to take the necessary action.	- If the app sends push notifications as an authentication method, an internet connection is required. - If you lose your phone or it is turned off, and you do not have copies of the code saved elsewhere, you will not be able to access the app.

3. Manage Personal Information in your Profile

When you create a social media account, all information disclosed in a profile is public by default, so anyone can access the content posted on that account. However, privacy needs and preferences vary from person to person. While some people prefer to have more exposure and thus be able to promote their content on social networks, others prefer to include limited or no information at all. To achieve greater protection, it is important to assess the extent to which you are willing to include personal information in a profile. Consider the following social media settings:

Username selection: The username is the "digital name" that a person chooses as a means of being identified online as an individual or organization. If you prefer not to be easily identified, you can use a pseudonym that may or may not be related to you. This name need not be consistent across all social networks and can be changed at any time by logging into your account settings.

Account images: You can personalize an account by including a profile picture. If you prefer not to be identified, we suggest that you choose an image in which you cannot be recognized and change it when necessary. Using the same image for all social networks makes it easier to identify the user on different platforms.

Location inclusion: When location services are enabled on a social networking platform, it allows the source of any online media activity to be tracked. Once this function is activated, it will remain active until it is disabled in the privacy settings. Even if you turn the location sharing feature on or off, your location could be discovered through the content or images you share.

Posting photos: Photographs and other media files contain information called Exif data, which details the location, device, date and time of capture, etc. Because of this, it is important to know the media content privacy policies of the sites you visit and where you share photos, and to constantly be aware of who you are sharing your posts and media content with18.

[18] Germain, Thomas. (2019). How a Photo's Hidden 'Exif' Data Exposes Your Personal Information. Consumer Reports.https://www.consumerreports.org/privacy/what-can-you-tell-from-photo-exif-data/

4. General Recommendations

Remote work or teleworking increases the risks associated with handling sensitive information in corporate and personal environments. Therefore, we recommend considering these tips that have been adapted to the "new normal" of teleworking and digital life[19]:

Keep your software up to date: The software of every device and app should be updated as often as possible. This provides the user with better security and a faster device response time. This can also provide protection against scams, viruses, Trojans, phishing attacks (impersonation), and other threats. These updates are likely to reset privacy preferences, so when making them, review the information you share with mobile apps and devices.

Use an antivirus program: Using antivirus software for portable devices connected to the internet serves as an initial scanner for any suspicious or malicious activity to which every web user is exposed. This can help monitor the delivery of notifications and provide an additional level of protection if you mistakenly click on suspicious links that may contain spam and different viruses.

Block and filter: The use of blocking, reporting, and filtering functions for emails, messages, and notifications allows platform services to remain secure and resilient. Whenever an account or post is blocked, it sends an important signal to the platforms about content or interactions that are undesirable to the individual, so that type of content is limited or blocked. You should not ignore suspicious content or content that violates a platform's user policies; it is better to report it continuously. If necessary, also report threats against your physical safety to law enforcement authorities[20].

Use a company laptop for remote work if possible and do not share information with other members of your household: Do not use your personal computer, as it may have fewer security controls than your company's hardware. If you can't avoid using personal equipment and have to use your own device, follow your organization's cybersecurity standards as closely as possible. Use the security software provided by your company, follow the company's data protection measures, and do not mix personal and work use.

Use designated VPNs and avoid public and free Wi-Fi networks: Using public Wi-Fi networks can put sensitive information at risk, such as passwords, bank details, and more. Similarly, if you work in a corporate environment, it is a good idea to use a designated VPN (Virtual Private Network) to keep your work assets secure while working from a remote location.

Use split networks: If you are working from home, we recommend that the personal network be accessible only to you, and that a specific network be created for guest use. If you have a router with VLAN capability, activate it and dedicate a VLAN for work purposes only.

[19] Roesler, Martin.(2020). Working From Home? Here's What You Need for a Secure Setup.Trend Micro. https://www.trendmicro.com/vinfo/ph/security/news/cybercrime-and-digital-threats/working-from-home-here-s-what-you-need-for-a-secure-setup

[20] U.S. Department of Homeland Security. (2019). Social Media Plan Guide: Science and Technology Directorate. https://www.dhs.gov/sites/default/files/publications/social_media_plan_guide_09_20_2019.pdf

Prepare a backup solution at home: Having backup options (e.g., hardware such as external hard drives or USB) is a key preventive measure in case you experience a data storage failure, such as loss of connectivity or server failure.

Recognize social engineering attacks: Several red flags may indicate a social engineering attack. The most common ones are[21]:

Generic language with mistakes: If the email comes from a safe and reliable source, the body should be written correctly, with proper spelling and grammar. Otherwise, it may be an attack. Another linguistic element that may signal an attempted attack are generic greetings and formulations. So, if an email starts with "Dear recipient" or "Dear user," be careful.

Unknown sender or sender with suspicious identification: If an email comes from an address that is a combination of random numbers and characters or is unknown to the recipient, it should go directly to the spam folder. However, sometimes, hackers may also have a legitimate email address, so it is still important to review the other warning signs in this section.

Sense of urgency: The criminals behind social engineering campaigns often try to scare victims into action by using anxiety-provoking phrases such as "send us your information immediately or your package will be discarded" or "if you don't update your profile now, we will close your account." Banks, parcel delivery companies, public institutions, and even internal departments usually communicate in a neutral and objective manner. So, if the message attempts to pressure the recipient to act quickly, it is probably a malicious and potentially dangerous scam.

[21] Eset. Social Engineering (in cybersecurity). https://www.eset.com/int/social-engineering-business/

Request for personal and private information: Institutions and even other departments in your own company will not normally request confidential information by email or telephone unless the contact was initiated by the employee.

These are just a few recommendations for how an individual or organization can proactively ensure high levels of cybersecurity on social networks and electronic devices. However, just as with literacy, it is the responsibility of each person to stay informed, review privacy settings, and continually update their cybersecurity measures to ensure that their important data and information is protected at all times.

Having this knowledge and putting these security measures into practice helps you to be better prepared to research, consume, and share information responsibly when online. The next section explores these areas in more depth for using Twitter as a tool to stay informed.

GENERAL RECOMMENDATIONS FOR JOURNALISTS

While the above sections contain key information for all individuals, those engaged in journalistic work, as well as others who use the internet for different activities, are not exempt from practicing good digital security habits to protect themselves in their work. For these people, the internet has also become a critical tool for them to carry out their work, especially at times of heightened public interest, such as during elections, when information about news events is fast moving. With that in mind, below are cybersecurity recommendations for specific instances, taken from the Covering elections: Journalist safety kit prepared by the Committee to Protect Journalists (CPJ)[22]:

I. Basic device preparedness

Before an assignment, the following are good practices to take into consideration:

- ☑ Back up your devices to a hard drive and remove any sensitive data from the device you are carrying.
- ☑ Log out of any accounts, apps and on all your browsers and clear your browsing history.
- ☑ Password protect all devices and set up your devices to remote wipe.
- ☑ Take as few devices with you as possible. If you have spare devices, take them instead of personal or work devices.

[22] Forbes, Jack. (2019). Covering elections: Journalist safety kit. Committee to Protect Journalists. https://cpj.org/2019/03/covering-elections-journalist-safety-kit/

II. Storing and protecting materials

It is important to have good protocols around the storing and securing of materials during election times. If a journalist is detained while covering elections, their devices may be taken and searched, which could have serious consequences for the journalist and their sources. The following steps can help protect you and your information:

- ☑ Review what information is stored on your devices, especially phones and computers. Anything that puts you at risk or contains sensitive information should be backed up and deleted.
- ☑ Review the content on your phones, including information stored on the phone (the hardware) as well as information stored in the cloud (Google Photos or iCloud).
- ☑ Check the content in messaging applications, such as WhatsApp. Save and then delete any information that creates a risk. Media workers should be aware that WhatsApp backs up all content to the cloud service linked to the account, for example iCloud or Google Drive.
- ☑ Regularly move material off your devices and save it on the back up option of your choice. This will ensure that if your devices are taken or stolen then you have a copy of the information.
- ☑ It is a good idea to encrypt any information that you back up. You can do that by encrypting your external hard drive or USB. You can also turn on encryption for your devices. Journalists should review the law in the country they are working in to ensure they are aware of the legalities around the use of encryption.
- ☑ If you suspect that you may be a target and that an adversary may want to steal your devices, including external hard drives, then keep your hard drive in a place other than your home.
- ☑ Put a PIN lock on all your devices. The longer the PIN, the more difficult it is to crack.
- ☑ Set up your phone or computer to remote wipe. This function allows you to erase your devices remotely, for example if authorities confiscate them.

SHARING AND CONSUMING INFORMATION ON TWITTER

Staying safe online is as essential to a positive and healthy internet experience as critical thinking and using information verification tools. Since Twitter works based on what is happening and what people are talking about now, it has become the main information platform for many people. However, with so much information available, it can sometimes be difficult to keep up with the conversation and verify the accuracy of the information being consumed. The following section provides recommendations, tools, and best practices for searching, organizing, sharing, and posting information on Twitter.

Verifying Information on Twitter

On Twitter, people can quickly find information and verify its accuracy. Being an open and public platform, there are several ways and tools for starting conversations with other people or doing a quick search of a hashtag[23] or key words, to assess the reliability of the information consumed on the platform.

When reading information, it is important to consider our own biases and opinions, as well as our personal reactions. Often, when we receive information we disagree with, we naturally ask ourselves certain questions or make certain comments that help to disprove the information. However, we generally omit this scrutiny when what we read confirms preconceived ideas. Because of this it is important to get into the habit of always asking yourself the who, what, where, when, why, and how of a piece of information before sharing, Retweeting[24], Quote Tweeting, or “liking” it.

[23] Hashtags (written with a # symbol) are used to index keywords or topics on Twitter. This function was created on Twitter, and allows people to easily follow topics they are interested in.
[24] A Retweet is a re-posting of a Tweet.

Who

- Who is the source? Do you know them?
- Is it a verified account?
- Who do they follow and who follows this account?
- Who wrote the article and what is their depth of knowledge on the matter?

What

- What did they say?
- What are their reasons for sharing this information?
- What type of article is it—information or opinion?
- What tone are they using? Is it, perhaps, intentionally false or a joke?
- What responses is this content getting? What are people saying on Twitter?

Where

- Where did it happen?
- Where was it said or published?
- Is it a reliable source?
- What is the URL or link to the website? Is it legitimate?
- What other media or people covered this news?

When

- When was it said?
- When was it published, and is it dated?

Why

- Why was the news published?
- Is it to generate traffic to the website or account?
- Is it intended to provoke action? If so, whose, and for what purpose?

How

- How is it written?
- Does it have excessive punctuation marks and capital letters to make it sensational?
- Does it have a misleading headline?
- Is it using hashtags unrelated to the topic to attract attention?
- Does it have a conspiratorial tone?

These seem like a lot of questions, but they can be answered in a matter of seconds, and there are Twitter tools that make this task easier and facilitate the safe and informed consumption of information on Twitter. The next section describes these tools and tips for their use.

Twitter Tools for Improving Information Consumption

Twitter continues to develop product updates to help people find information and learn the context around it quickly and easily. For example, to help people stay informed about important national and global events, reliable, high-quality information about events of significant public interest such as election information, natural disasters, or global crises such as COVID-19 may be displayed on the timeline, and you can choose whether you want to continue viewing that information or whether you are not interested in it.

To help you understand the context around a piece of information, when you want to Retweet (share) a Tweet that contains a link you have not opened from the platform, Twitter displays a notice recommending that you open the link to find out all the information before sharing it. This encourages people to ask critical questions about the information they are about to share.

The above are examples of products that reactively help people to be better informed. There are also Twitter tools and other platform design elements that help people proactively explore and organize the vast information available on the platform and provide additional context for its analysis and verification. Some are:

Trends

Trends exist to help people discover conversations taking place around them. They are determined automatically, considering different factors to identify topics that are popular at a given time, rather than topics that have been popular for a period or every day.

Trends are available in the ***Explore*** section of the Twitter app 🔍 , and in different places on twitter.com, such as the Home timeline, notifications, search results, and profile pages. By clicking or tapping on any trend, you will see the Twitter search results related to that trend, i.e., all Tweets that include that phrase or hashtag.

Trends for you vs. Trends from a geographic location

By default, Twitter displays *Trends For You*, which are automatically determined based on the accounts you follow, your interests, and your location. To view Trends by geographic area, on twitter.com or from within the app, you can click on the ***settings*** icon ⚙ and select Trends for specific locations. If you do not find the city or country you are looking for, it means there are not yet enough Tweets in that geographic area to create a Trends list. In these cases, you can search for local Tweets on any topic using Twitter's Advanced Search function.

Context in Trends

Along with some Trends, you can see:

- The approximate number of Tweets associated with that trend. The number of Tweets related to the Trends is just one of the factors involved in ranking and determining the Trends. That is why sometimes you can see that the number of Tweets about the first Trends is lower than the number of Tweets about the following ones.
- Customized contextual information, e.g., who in your network is tweeting about the trend.
- A category such as “Politics,” “Music,” or “Entertainment.” This is automatically selected based on what the Trending Tweets are about.
- Articles, which are automatically grouped together based on the trend conversation.

Search Results

Whenever you perform a search on Twitter, either from twitter.com or from the app, the results can be displayed according to when they were shared or the type of content. Each search is automatically organized into different tabs that give the option to view the Tweets:

Top

Most relevant results based on the interests of the account performing the search.

Latest

Search results in chronological order.

People

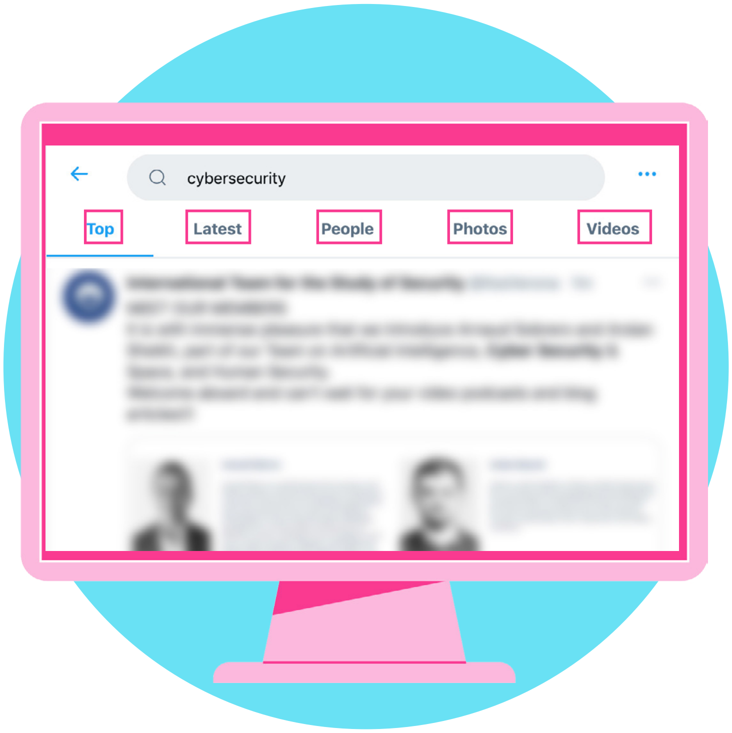
Results of the accounts that match the query.

Photos

Search results containing photos.

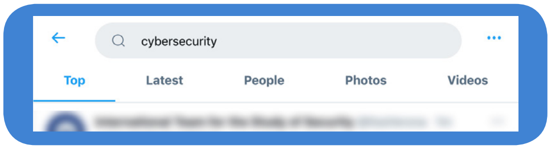
Videos

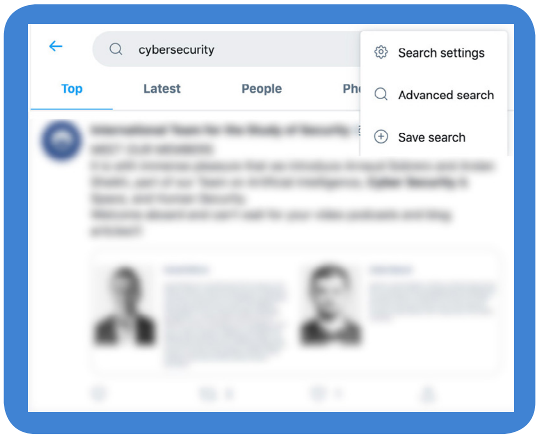
Search results containing video.

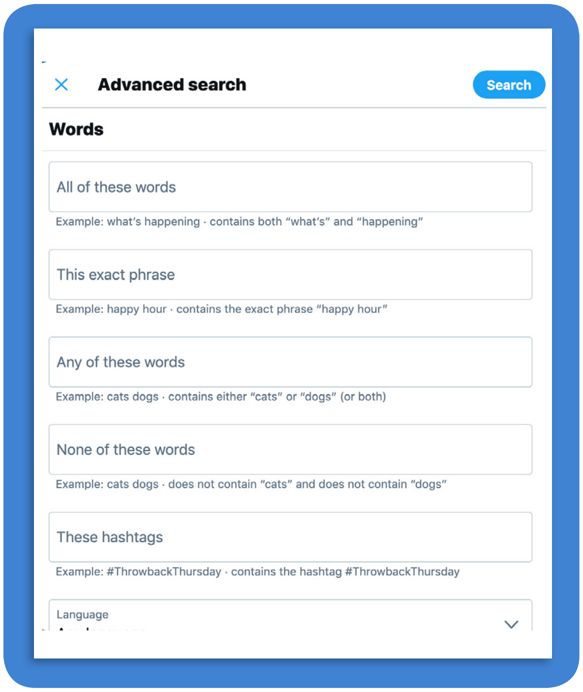


Advanced Search

If you want to search for specific information on twitter.com, you can do an advanced search. To use this option, enter your search into the search bar, then click *Advanced search* ⋯ . There you will find a set of fields to refine your search and find specific content more directly and quickly.







Words

All of these words

Example: what's happening · contains both "what's" and "happening"

This exact phrase

Example: happy hour · contains the exact phrase "happy hour"

Any of these words

Example: cats dogs · contains either "cats" or "dogs" (or both)

None of these words

Example: cats dogs · does not contain "cats" and does not contain "dogs"

These hashtags

Example: #ThrowbackThursday · contains the hashtag #ThrowbackThursday

Language

In the app, the search options are more limited, but, with certain search formulas, you can also refine information in the results. Using the phrase *Twitter app*, the following example show how terms can be used depending on the content you are looking for:

Twitter app

to search for content containing all the search terms, whether they are words, @account names, or hashtags. Here, the word *Twitter* and the word *app*.

Twitter -app

using a hyphen, or minus sign, to search for content that contains the word *Twitter* but not the word *app*. That is, to exclude what is placed after the hyphen.

"Twitter app"

to search for content containing the exact phrase in quotation marks, i.e., *"Twitter app."*

from:TwitterSafety

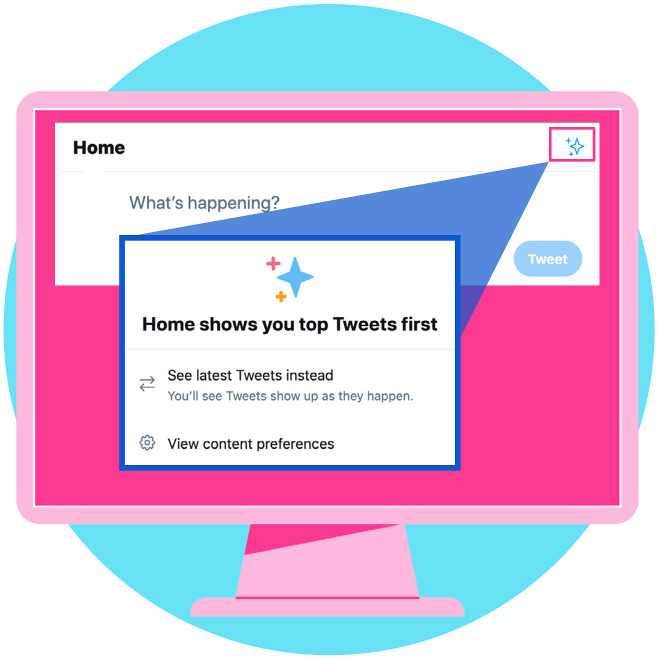
to search for content posted from a specific Twitter account. Here, *@TwitterSafety*.

Twitter OR app

to search for content containing one term or another. In this case, the terms *Twitter* or *app*, or both.

Home Timeline: “Top Tweets” vs. “Latest Tweets”

The twitter.com home page or app home page displays the most relevant Tweets at the top of the timeline by default. However, sometimes it is better to view Tweets in chronological order, i.e., view the most recent Tweets first. This depends not only on each person’s preferences, but also on the information sought. For example, during a sporting event or in emergency situations, it is more useful to view the most current information first.


Home
What’s happening?
Tweet
Home shows you top Tweets first
See latest Tweets instead
You’ll see Tweets show up as they happen.
View content preferences

This is why Twitter offers the possibility of easily and quickly changing the settings of the Home timeline between Top Tweets and Latest Tweets. To make this change, from twitter.com or from within the app, click on the icon ✨ in the upper right corner and select the option of your choice.

The default setting in Twitter is Top Tweets, so when you change the setting to Latest Tweets and you stop using Twitter for a while, the setting will automatically revert to Top Tweets.

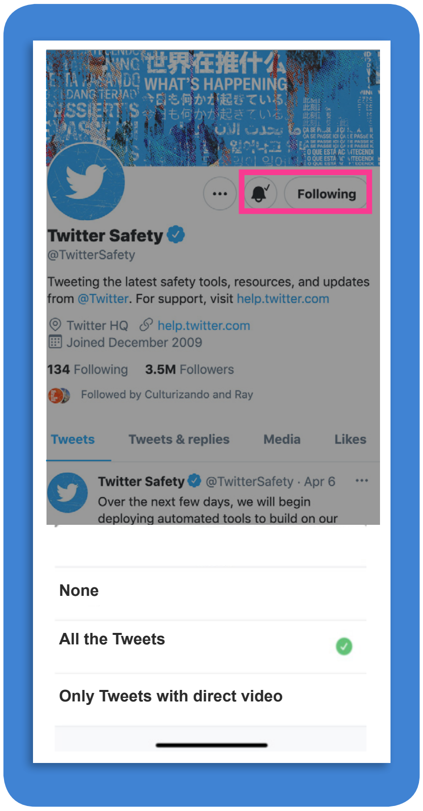
Account Notifications

Sometimes, you want or need to be aware of the content posted by specific accounts and this is what account notifications or push notifications are for. These notifications send the user a notification or alert when certain accounts post Tweets. You have the option of activating these notifications for all Tweets from an account, or only for Tweets containing live broadcasts. These can be enabled and disabled at any time.

To enable Notifications:

1. Make sure you are following the account for which you want to receive real-time notifications.
2. In the account profile, either on twitter.com or in the app, click on the ***Notifications*** icon 🔔 .
3. If you do this from twitter.com, both notifications will be activated: for all Tweets and for live videos. From the app, you can choose between two types of notifications: *All Tweets* or *Only Tweets* with live videos.

To **disable Notifications,** go back to the account profile, click on the highlighted ***Notifications*** icon 🔔 and select *None.*



Twitter Safety
@TwitterSafety

Tweeting the latest safety tools, resources, and updates from @Twitter. For support, visit help.twitter.com

Twitter HQ help.twitter.com
Joined December 2009

134 Following 3.5M Followers

Followed by Culturizando and Ray

Twitter Safety @TwitterSafety · Apr 6
Over the next few days, we will begin deploying automated tools to build on our

None

All the Tweets ✓

Only Tweets with direct video

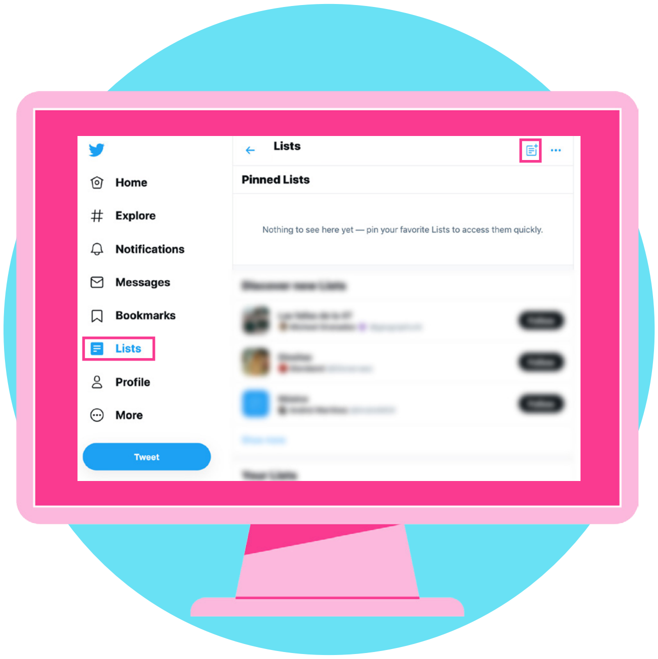
If you need to check which Notifications are active, you can do this from within the app at:

1. Your account's navigation bar
2. Select *Settings* and privacy.
3. Tap *Notifications*, then *Push Notifications.*
4. Tap *Tweets.*

Lists

A List is a filter that displays a customized Home timeline that only shows Tweets from accounts on that List. For example, you can sort the Home timeline by creating specific Lists of experts, journalists, comedians, authorities, services, etc. Some important features of Lists include:

- You can create your own Lists or subscribe to Lists created by others.
- You do not have to follow an account to add it to a List.
- Lists can be private, for personal monitoring, or public, for sharing information with others.
- If Lists are public, accounts added to the List will receive a notification to that effect.
- In the app, up to 5 Lists can be pinned to the Home screen for quick access.



Pinned Lists

Nothing to see here yet — pin your favorite Lists to access them quickly.

To create a List:

1. Click on *Lists* in the navigation bar.
2. Click on *Create New List.*
3. Choose a name for your List and write a short description. In this step, you need to specify if you want the List to be private (only the account owner can view and access it) or public (anyone can view and subscribe to the List).
4. Click on *Save List.*

To add people to a List:
To do this, it is not necessary to follow the accounts you want to add to the List.

1. Click on the ***more*** icon ··· in the profile of the account you want to add to the List.
2. Select *Add* or *remove from Lists.* A pop-up window will open displaying your created Lists or giving you the option to create a new one.
3. Click on the List(s) you would like to add the account to or uncheck the Lists you would like to remove the account from.

To edit or delete a List:

1. Click on *Lists* in the navigation bar.
2. Click or tap the List you would like to edit or delete from the Lists you have created.
3. Click or tap the *Edit* button to update the List details or to delete the List altogether.

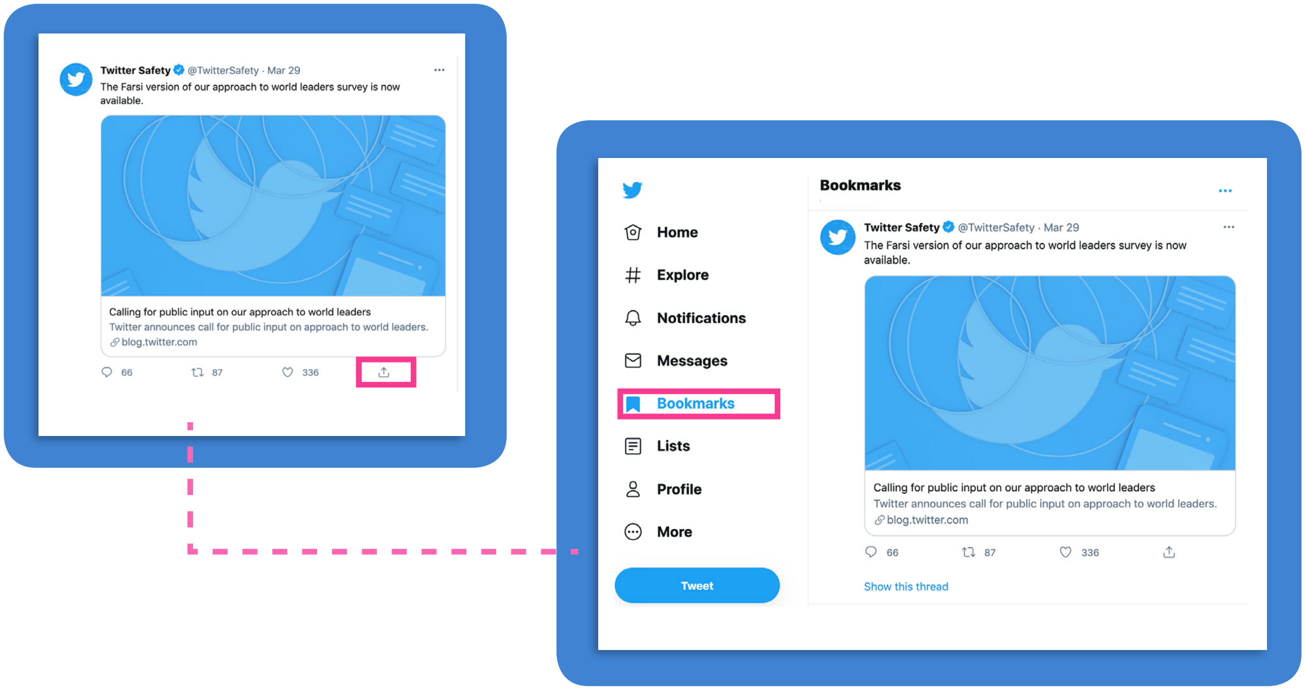
To follow other people's Lists:

1. Click on the ***more*** icon ··· in the account's profile.
2. Click or tap *Lists.*
3. Select the List you wish to follow.
4. On the List's page, you can click or tap *Follow* to follow the List. You can follow a List without having to follow the account that created it and without having to follow the individual accounts in that List.

Saved Tweets (Bookmarks)

From articles and threads to videos and GIFs, timelines are full of Tweets you don't always have time to explore right away, or that you want to save for later review or reference. That is why there are Twitter Bookmarks.

To bookmark a Tweet, tap on the ***share*** icon ⇪ below the Tweet to be saved and select *Add Tweet to Bookmarks*. When you want to find it, tap Bookmarks in your profile menu and you will find it there. Tweets can be removed from Bookmarks at any time and only the account's owner can view their Bookmarks.



Twitter Safety @TwitterSafety · Mar 29

The Farsi version of our approach to world leaders survey is now available.

Calling for public input on our approach to world leaders
Twitter announces call for public input on approach to world leaders.
blog.twitter.com

Using these tools properly is key to verifying the information found on Twitter but being on the platform involves much more than consuming information. It is also about sharing information and interacting with other people. To ensure that Twitter is a space where people can engage freely and safely, there are rules about what is and is not allowed on Twitter, as well as tools to help people control their experience on the platform. These guidelines are described in the section below.

TWITTER BEST PRACTICES FOR AUTHORITIES AND ORGANIZATIONS

Twitter is one of the fastest social media platforms for organizations, authorities, and experts to share relevant and accurate information with the largest number of people with minimal effort.

To use Twitter effectively, enhance the credibility of their online presence, and position themselves as trusted sources of information, organizations, entities, and authorities should have a consistent and interactive content plan that reflects their credibility and relevance. Some tips to make this possible:

- ☑ Your account profile is your calling card on Twitter, so it should reflect the most up-to-date information: name, bio, location, and website. The header and profile picture also help people to easily identify the account.
- ☑ Messages should always be concise, easily digestible, and conversational in tone rather than intended as a discourse. People go to Twitter to interact, ask questions, and share reactions. Basic interactions such as likes, Retweets, mentions, and replies can encourage conversations around your topics of interest.
- ☑ Media content is highly interactive and effective, but only if it is shared natively and not from other platforms. Media content also should be clearly related to the message to be shared, and in the case of videos, be extremely short—16 seconds is ideal.
- ☑ Relevant timing is key on Twitter. Engaging in the moment when events are happening, sharing reactions, and providing first-hand information increases the relevance and credibility of your presence on Twitter.
- ☑ Staying connected, sparking conversations, holding Q&A sessions—grouped around a hashtag—and live broadcasts are important for sharing your unique viewpoint of the moment directly with your followers to help them be better informed.

TWITTER SAFETY

Twitter is an open space of free expression, in which everyone can participate freely. This makes it a useful and relevant tool for sharing and finding up-to-date information quickly. To ensure that people feel safe to express diverse opinions and beliefs, there are rules for using the platform to protect the health of the public conversation and prevent voices from being silenced. This section explains the Twitter Rules, which are important guidelines to know what is and is not allowed on the platform. It also describes the tools available to customize and control each person's experience on the platform to make it as pleasant and productive as possible.

Twitter Rules

Twitter reflects the real conversations happening in the world, and sometimes that includes perspectives that some may find offensive, controversial, or bigoted. While Twitter is a participatory space where diverse opinions can be expressed, the platform does not tolerate behavior that uses harassment, intimidation, or fear to silence other people's voices. The platform's rules are intended to ensure that everyone can join in the public conversation freely and safely. These rules are divided into three main categories: **safety, privacy, and authenticity**:

Safety

☒ Violence: You may not threaten violence against a person or a group of persons. The glorification of violence is also prohibited.

☒ Terrorism or violent extremism: You may not threaten or promote terrorism or violent extremism.

☒ Child sexual exploitation: Twitter has zero tolerance for the sexual exploitation of children.

☒ Abuse/harassment: You may not engage in the targeted harassment of a person or incite others to do so. This includes wishing or hoping that someone will experience physical harm.

☒ Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.

☒ Suicide and self-harm: You may not promote or encourage suicide or self-harm.

☒ Sensitive media content, including graphic violence and adult content: You may not post media that is excessively gory or share violent or adult content within live video, profile headers, or banner images. Media depicting sexual violence or assault is also not permitted.

☒ Illegal or regulated goods or services: You may not use Twitter for any unlawful purpose or in furtherance of illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of regulated goods or services.

Privacy

☒ Private information: You may not publish or post other people's private information (such as telephone number and home address) without their authorization and permission. Twitter also prohibits threatening to expose private information or incentivizing others to do so.

☒ Non-consensual nudity: You may not post or share intimate photos or videos of someone that were produced or distributed without their consent.

Authenticity

☒ Platform manipulation and spam: You may not use Twitter’s services in a manner intended to artificially amplify or suppress information or engage in behavior that manipulates or disrupts people’s experience on Twitter.

☒ Election integrity: You may not use Twitter’s services for the purpose of manipulating or interfering in elections. This includes posting or sharing content that may suppress voter participation or mislead people about when, where, or how to vote.

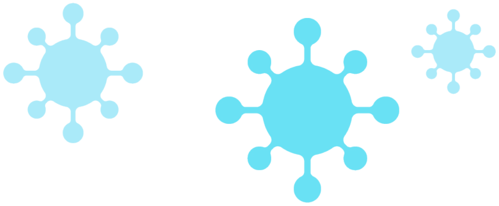
☒ Impersonation: You may not pose as other individuals, groups, or organizations in a manner that attempts to, or succeeds in, confusing, misleading, or communicating a misconception to others.

☒ Synthetic and manipulated media: You may not deceptively promote synthetic or manipulated media that are likely to cause harm. In addition, Tweets containing synthetic and manipulated media may be labeled to help people understand their authenticity and to provide additional context.

☒ Copyright and trademark: You may not violate others’ intellectual property rights, including copyright and trademark.

More detailed information on the Twitter Rules, and the most current version of them, is available in the Twitter Help Center or at: https://twitter.com/rules.

The Twitter Rules are constantly evolving and may change from time to time to support the goal of fostering healthy and constructive public conversation. The global COVID-19 pandemic prompted a rule change to broaden the definition of harm and address content directly contrary to the instructions of authoritative sources of global and local public health information. These updates are done transparently and communicatively. For example, during the changes made in the wake of COVID-19, Twitter kept people up to date through its blog Coronavirus: Staying safe and informed on Twitter.



Enforcement of the Twitter Rules

When Twitter takes action to enforce its rules, it may do so either on a specific piece of content (e.g., an individual Tweet or Direct Message) or on an account. It may also employ a combination of these options. In enforcing the rules, Twitter starts from the assumption that people do not intentionally break the rules. Therefore, unless a violation is so egregious that it forces the platform to suspend an account immediately, efforts are first made to educate individuals about the Twitter Rules. By doing this, Twitter gives people who use the platform the opportunity to correct their behavior by showing them the Tweet(s) that violate the rules, explaining which rule was broken, and requiring that the content be removed before the person can tweet again. If a person repeatedly violates the rules, enforcement action becomes more serious.

Tweet-level enforcement

Action is taken at the Tweet level to ensure that an otherwise healthy account that made a mistake and violated the rules is not treated too harshly. These actions may include:

Limiting Tweet visibility

This makes content less visible on Twitter, in search results, replies, and on timelines.

Requiring Tweet removal

If it is determined that the Tweet violated the Twitter Rules, the violator is required to remove it before they can Tweet again. In these cases, the person receives an email notification identifying the Tweet(s) that violate the rules. The Tweet in question must be deleted or the individual can appeal the decision if they believe Twitter made an error.

Hiding a violating Tweet while awaiting its removal

In the interim period between when Twitter takes enforcement action and the person removes the Tweet, that Tweet is hidden from public view and the original content is replaced with a notice stating that the Tweet is no longer available because it violated the Twitter Rules. This notice will remain visible for 14 days after the Tweet is removed.

Notice of public interest exception

In very specific cases where it is determined that it is in the public interest for a Tweet that violates the Twitter Rules to remain accessible on the platform, the Tweet is placed behind a notice explaining the exception and giving you the option to view the Tweet if you wish.

When this notice is applied, measures are also taken to reduce the visibility of the Tweet, such as:

- ✓ Engagements with the Tweet (replies, Retweets, and likes) will be turned off.
- ✓ No engagement counts on the Tweet (e.g. number of likes, replies) will be shown.
- ✓ Any previous replies will not be viewable within the Tweet details.
- ✓ The Tweet will no longer be available in:
    - The Home timeline under *Top Tweets*
    - Search results
    - Recommendations and Notifications
    - The *Explore* tab

The following section explains Twitter notices and their meaning.

Direct Message-level action

Stopping conversations between a reported violator and the reporter's account

In a private Direct Message conversation, when a participant reports the other person, Twitter prevents the violator from continuing to send messages to the person who reported them. The conversation is also removed from the reporter's inbox and can only be resumed if the reporter continues to send Direct Messages to the violator.

Placing a Direct Message behind a notice

In a group Direct Message conversation, the violating message may be placed behind a notice to ensure no one else in the group can see it again.

Account-level enforcement

Twitter takes action at the account level if it determines that a person has violated the Twitter Rules in a particularly egregious way or has repeatedly violated them even after receiving notifications from Twitter.

Requiring media or profile edits

If an account's profile or media content violates the rules, the account may be temporarily suspended and Twitter, in addition to informing the account owner, may require the violator to edit the content or information in their profile to come into compliance.

Placing an account in read-only mode

If an otherwise problem-free account has an episode of abusive behavior, Twitter can temporarily modify their settings to be read-only, limiting their ability to Tweet, Retweet, or Like content for a set amount of time. The person will still be able to view their Home timeline and send Direct Messages to their followers.

When an account is in read-only mode, others will still be able to see and engage with it. The duration of this enforcement action can range from 12 hours to 7 days, depending on the nature of the violation.

Verifying account ownership

To ensure that people do not abuse the anonymity that Twitter offers, Twitter sometimes requires account owners to verify their authenticity with a phone number or email address. This also helps to identify and take action against violators who are operating multiple accounts for abusive purposes.

Permanent suspension

This is Twitter's most serious enforcement action. Permanently suspending an account will remove it from global view, and the violator will not be allowed to create new accounts. When Twitter permanently suspends an account, it informs the person that they have been suspended for abuse-related violations and explains to them which policy or policies they have violated, and which content was in violation.

Appealing these actions

Reported persons or violators may appeal any of these actions if they believe that Twitter has made a mistake. They can do this through the platform interface or by sending a report using the appeal form.

Reporting Violations of the Twitter Rules

If you find content on Twitter that you believe violates the platform's rules, the best thing to do is to report it. When reporting, remember that the context you can provide is very important. Likewise, remember that not all content that some consider offensive or bigoted is necessarily in violation of the Twitter Rules.

In determining whether to take action on a report, Twitter teams consider several factors, including:

- Whether the behavior is directed at an individual, a group, or a protected class of people.
- Whether the reporter is the object of the abuse or a witness.
- Whether the reported person has a history of violating the platform's rules.
- The seriousness of the violation.
- Whether the content is a topic of legitimate public interest.

At help.twitter.com/forms you can find the direct forms to report any violation of the Twitter Rules. There are also direct options from twitter.com and from the app to report Tweets, accounts, or Direct Messages.

To report an account:

1. Open the profile you wish to report.
2. Select the ***more*** icon ••• .
3. Select *Report @username* and then select the type of violation you would like to report.
4. Depending on your selection, the platform will ask you for additional information about the issue you are reporting.

To report a Tweet:

1. Navigate to the Tweet you would like to report.
2. Tap the ***more*** icon ⋯ located at the top of the Tweet.
3. Select *Report Tweet.*
4. Depending on your selection, the platform will ask you for additional information about the issue you are reporting.

To report a Direct Message:

1. Click into the Direct Message conversation and find the message you would like to report.
2. Click on the ***more*** icon ⋯.
3. Select *Report @username.*
4. Depending on your selection, the platform will ask you for additional information about the issue you are reporting.

Notices on Twitter and their Meaning

For some of the actions outlined above, Twitter places a notice on the account or Tweet to provide more context on the enforcement actions taken. It is important to understand the meaning of these notices to know the difference between, for example, an account that has been deactivated by the person or an account that has been temporarily suspended; or between a Tweet that has been deleted by the original author and a Tweet that has violated the platform's rules. Some of the notices you may see on Twitter include:

> This Tweet may include sensitive content.

> The following media includes potentially sensitive content. **Change settings**

Notice of sensitive media content, not suitable for minors or that includes graphic violence. In this case, you are informed that if you choose to click on the notice, you will be viewing sensitive media content.

Notice of content in the public interest. This is for very specific cases where a Tweet that violates the Twitter Rules is determined to be of public interest and will therefore remain accessible on the platform. The Tweet is placed behind a notice explaining the exception and giving you the option to view the Tweet if you wish.

> This Tweet violated the Twitter Rules about [specific rule]. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. **Learn more**

> This Tweet violated the Twitter Rules. **Learn more**

Notice for a deleted Tweet that violated the rules. If a Tweet that violates the rules has not yet been deleted by the person who tweeted it, it is hidden behind a notice and the account remains blocked until the author deletes the Tweet. Once the Tweet is deleted, the notice remains on the platform for 14 more days.

Notice for a Tweet from a suspended account. Tweets from an account suspended for violations of the Twitter Rules appear hidden behind a notice with this information.

> This Tweet is from a suspended account. **Learn more**

> You reported this Tweet. **View**

Notice for a reported Tweet. When you report a Tweet, you will see the Tweet behind a notice to that effect and may choose whether you want to see the content again.

Notice for a Tweet from a blocked or muted account. If you muted or blocked one or more accounts and someone else shares their Tweets, Twitter will hide the content of those Tweets behind a notice and give you the option to click through and view it.

> This Tweet is from an account you muted. **View**

If you muted words or hashtags, you will see a similar notice:

> This Tweet includes a word you muted. **View**

> This account owner limits who can view their Tweets. **Learn more**

Notice for a Tweet with limited visibility. This notice appears when a Tweet is unavailable to view if, for example:
It is a Tweet from a protected account, meaning only people who follow it can view its content or it is a Tweet from an account that has blocked you.

> This Tweet is unavailable. **Learn more**

The Tweet was deleted by its author.

> This Tweet is from an account that no longer exists. **Learn more**

The Tweet is from a deactivated account.

Notice for an account that must verify ownership. When an account owner is required to verify their authenticity with a phone number or email address, the account is temporarily restricted until the requested information is provided.

> Caution: This account is temporarily restricted. **Yes, view profile**

> This account doesn't exist. Try searching for another.

Notice of deactivated account. Account owners can deactivate their account at any time. When an account owner deactivates their account, the page will be rendered as unavailable.

Notice of permanent suspension. If an account has been suspended for violating the Twitter Rules, this information is displayed on the account in question.

> Account suspended. Twitter suspends accounts which violate the Twitter Rules.

Twitter Transparency Report

Following its principle of transparency and under the belief that the free exchange of information can have a positive impact, Twitter has published a biannual transparency report since 2012. This report covers information and takedown requests that Twitter receives from law enforcement, copyright and trademark infringement notices, enforcement of the Twitter Rules, and instances of platform manipulation, including information operations that Twitter has determined to be supported by States.

Control your Experience on Twitter

Twitter is a place to share ideas and information, connect with your communities, and see the world around you. To protect that experience, Twitter provides tools designed to help you personalize and control what you see and what others can see about you, so you can express yourself on Twitter with confidence.

Notifications Filter

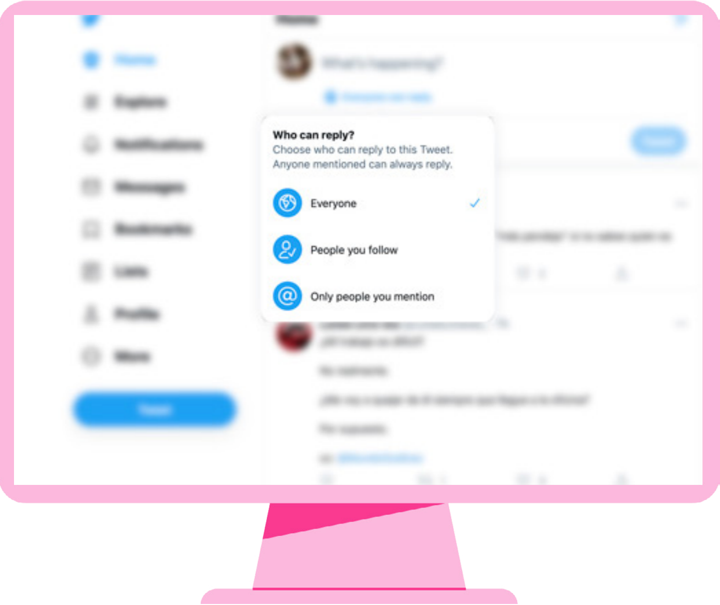
The Notifications timeline shows your interactions with other Twitter accounts, such as mentions, Likes, Retweets, and who has started following you. If you receive unwanted replies or mentions, you can filter the types of notifications that you receive. You have three options in your notifications settings to filter the notifications you receive: quality filter, muted words, and advanced filters.

- **Quality filter:** filters lower-quality content from appearing in your notifications (for example, duplicate Tweets or content that appears to be automated) but does not filter notifications from people you follow or accounts you have recently interacted with.
- **Muted words:** mutes notifications that include specific words and phrases that you would not like to see in the notifications.
- **Advanced filters:** allow you to disable notifications for certain types of accounts, like accounts that do not follow each other, or use the default Twitter profile picture, or do not have an email or phone number on record to confirm their authenticity.

Replies Control

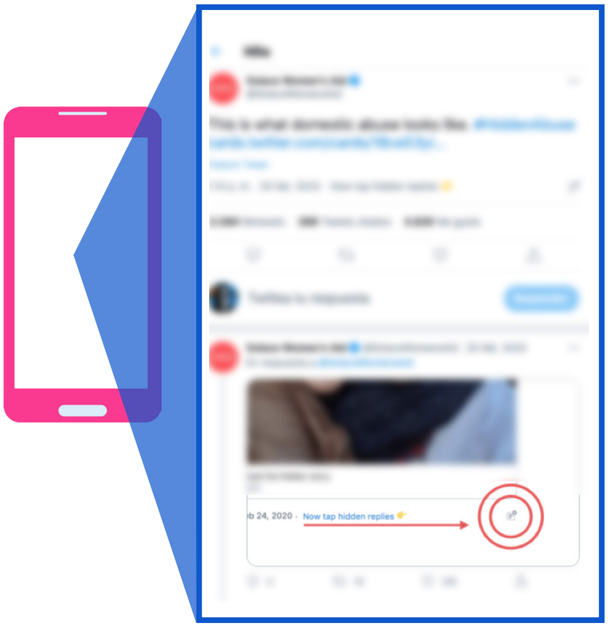
When you write a Tweet you can choose who can reply to it. By default, the *Everyone can reply* option is selected at the bottom left in the Tweet compose box. By clicking or tapping on this option before posting a Tweet, you can choose who can reply to that specific Tweet. The options are: *Everyone, People you follow,* or *Only people you mention* in that Tweet.

People will see if you limited who can reply to a Tweet and restrictions cannot be changed once a Tweet has been posted.



Who can reply?
Choose who can reply to this Tweet.
Anyone mentioned can always reply.

Hidden Replies



To help you remain in control of the conversations you have started, Twitter has a tool that allows the Tweet author to hide specific replies to their content. Everyone can still access hidden replies through the ***hidden reply*** icon, which shows up in the original Tweet when there are hidden replies. The Tweet author can hide or unhide a reply at any time and the author of the reply will not be notified.

To hide a reply, tap or click on the ***more*** icon ••• of the Tweet you want to hide, select *Hide reply* and confirm. To view your hidden replies, tap or click on the ***hidden reply*** icon, which will be available in the bottom-right of your original Tweet.

Muting

You can mute accounts, words, or conversations:

Mute an account. Muting an account on Twitter means that the Tweets from that account will not appear on your Home timeline. Muted accounts do not receive any notification that they have been muted. Notifications of mentions from these accounts will remain active and Direct Messages can still be exchanged. You can also mute accounts that you do not follow to hide their Tweets from your Notifications timeline.

To mute an account, tap the ***more*** icon ••• in a Tweet and click on *Mute*. To unmute an account, visit the Twitter profile of the muted account, and click on the account's ***more*** icon ••• and then click on *Unmute @username* to unmute it.

Mute words, phrases, usernames, emojis, or hashtags. The mute option will prevent these Tweets from appearing in your Notifications tab, push notifications, email notifications, Home timeline, and replies, although they will still be visible in search results. To add or remove items from your mute list:

1. Click on the *more* icon in the navigation bar and select *Settings and privacy*.
2. Click on *Privacy and safety.*
3. Click on *Mute and block.*
4. Click on *Muted words* and then on *Add.*
5. Enter the words or hashtags you would like to mute, one at a time.
6. Select *Home timeline* if you wish to mute the word or phrase from your Home timeline or *Notifications* if you wish to mute the word or phrase from your Notifications.
7. Specify *From anyone* or *From people you don't follow.*
8. Under *Mute timing* choose between *Forever, 24 hours from now, 7 days from now,* or *30 days from now.*
9. Click *Save.*

Mute conversations allows you to stop receiving notifications for a particular conversation. When you mute a conversation, you will not get any new notifications about that conversation. You will, however, still see Tweets from the conversation in your timeline and when you click into the original Tweet. To mute a conversation:

1. Click on the ***more*** icon ••• of any Tweet or reply in the conversation you would like to mute.
2. Click or tap *Mute this conversation.*
3. Tap or click to confirm.

Blocking

Blocking a Twitter account prevents that account from interacting with the account that blocked it. Blocking can be useful to control unwanted interactions from accounts you are not interested in interacting with. Accounts you have blocked cannot see your Tweets, interactions, who you follow, or who follows you if they are logged in on Twitter. You will not receive notifications from blocked accounts and their Tweets will not appear in your timeline. The person managing an account you have blocked may notice that they have been blocked if they try to visit your profile or follow your account, but they will not receive notifications alerting them of the block.

To access this option, tap the ***more*** icon ⋯ in a Tweet from the account you wish to block or from its profile and click *Block*. To unblock an account, visit its profile on Twitter, click on the *Block* button and confirm that you wish to unblock the account.

FINAL CONSIDERATIONS

The content presented in this document, developed by the **OAS** and **Twitter**, is intended to inform and raise awareness about managing, consuming, and sharing information online with a focus on social networks, particularly Twitter. This guide is designed to help everyone, including journalists, government authorities, and organizations, gain a better understanding of the importance of media literacy and cybersecurity.

The rise of digital activities has highlighted preexisting vulnerabilities in the digital space. The proliferation of cyberattacks and the digitalization of countless everyday processes reaffirm the need to improve media literacy and awareness of good cybersecurity practices. This edition of the guide provides a fresh look at tools and best practices for consuming information and content safely and responsibly.

Technological platforms and social networks have created new modes of communication, expanding possibilities for political engagement by allowing the digital environment to play a role in democratic processes. Media literacy is essential for the strengthening of democracy, since it is a tool that contributes to mass participation and facilitates active and responsible citizen engagement. Literacy also counteracts phenomena such as disinformation and the interference of external actors in domestic politics, among other elements that directly or indirectly affect and influence democratic processes.

Throughout the different sections, information related to cybersecurity and digital self-care has been compiled and updated to present the new threats and tools arising from the changing environment and the expansion of teleworking. The guide also includes specific recommendations for consuming information on Twitter, updates on the rules for its use, and essential tools to enhance the user's experience on the platform.

GIVEN THE CONSTANTLY EVOLVING NATURE OF MEDIA LITERACY AND SECURITY, AND THE COMPLEXITY OF ONLINE CRIME, IT IS VITALLY IMPORTANT TO STAY INFORMED AND UP TO DATE ON THE PRODUCTS AND POLICIES THAT AFFECT DIGITAL MEDIA AND SOCIAL MEDIA INTERACTIONS.

The intensive use of digital technologies in the world will remain a part of everyday life, so cybersecurity and media literacy, applied by every individual, are critical to ensuring that we can safely capitalize on the benefit of connectivity and the availability of information, to provide an environment that offers greater possibilities for development, as well as for social welfare and the strengthening of democracy.

References

BBC News Mundo. (2020). *Coronavirus: la advertencia de la OMS sobre los estafadores que están usando el nombre de la organización para robar dinero y datos.* https://www.bbc.com/mundo/noticias-52009138

Duclkin, Paul. (2020). *Dirty little secret extortion email threatens to give your family coronavirus. Sophos* https://nakedsecurity.sophos.com/2020/03/19/dirty-little-secret-extortion-email-threatens-to-give-your-family-coronavirus/

Eset. *Social Engineering (in cybersecurity).* https://www.eset.com/int/social-engineering-business/

Forbes, Jack. (2019). Covering elections: Journalist safety kit. Committee to Protect Journalists. https://cpj.org/2019/03/covering-elections-journalist-safety-kit/

Germain, Thomas. (2019). *How a Photo's Hidden 'Exif' Data Exposes Your Personal Information.* Consumer Reports. https://www.consumerreports.org/privacy/what-can-you-tell-from-photo-exif-data/

INCIBE y OSI. (2020). *Detectada oleada de falsos correos de sextorsión o infección de COVID19* [Wave of fake sextortion or COVID-19 infection emails detected]. https://www.osi.es/es/actualidad/avisos/2020/04/detectada-oleada-de-falsos-correos-de-sextorsion-o-infeccion-de-covid19

Marsh y Microsoft. (2020). *Estado de Riesgo Cibernético en Latinoamérica en Tiempo del COVID-19* [Cyber Risk in Latin America during COVID-19]. https://coronavirus.marsh.com/mx/es/insights/research-and-briefings/report-cyber-risk-in-latin-america-in-times-of-covid19.html

Media and Information Literacy for the Sustainable Development Goals. Grizzle, A and Singh, J. (2016). In the MILID Yearbook 2016.

News and Media Literacy: What is Media Literacy, Common Sense Media: https://www.commonsensemedia.org/news-and-media-literacy/what-is-digital-literacy (consultado el 18 de agosto de 2019).

NortonLifeLock. *What is Social Engineering?.* https://lam.norton.com/Internetsecurity-emerging-threats-what-is-social-engineering.html

World Health Organization. (2020). *WHO reports a fivefold increase in cyber attacks, urges vigilance.* https://www.who.int/news/item/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance

United Nations Economic Commission for Latin America and the Caribbean. Special Report on COVID-19 (2020). *Universalizing access to digital technologies to address the consequences of COVID-19.* https://repositorio.cepal.org/bitstream/handle/11362/45939/5/S2000549_en.pdf

United Nations Educational, Scientific, and Cultural Organization (UNESCO). (2016). *Literacy.* https://en.unesco.org/themes/literacy

Organización de los Estados Americanos y Twitter. (2019). *Alfabetismo y Seguridad Digital: Mejores Prácticas en el Uso de Twitter.* https://www.oas.org/es/sms/cicte/docs/20190913-DIGITAL-Alfabetismo-y-seguridad-digital-Twitter.pdf

Organization of American States & Twitter. (2019). *Media Literacy and Digital Security: Twitter Best Practices.* https://www.oas.org/en/sms/cicte/docs/20190913-DIGITAL-ENG-Alfabetismo-y-seguridad-digital-Twitter.pdf

Organization of American States. (2019). *Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts.* https://www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf

Porter, Taryn. (2020). COVID-19 *Scam Alters. Cybercrime Support Network.* https://cybercrimesupport.org/covid-19-scam-alerts/

Roesler, Martin.(2020). *Working From Home? Here's What You Need for a Secure Setup.* Trend Micro. https://www.trendmicro.com/vinfo/ph/security/news/cybercrime-and-digital-threats/working-from-home-here-s-what-you-need-for-a-secure-setup

SoftwareLab.org. *What is social engineering: Top 5 Types & Examples.* https://softwarelab.org/what-is-social-engineering/

Stone, Jeff. (2020). *How scammers use fake news articles to promote coronavirus 'cures' that only defraud victims.* Cyberscoop. https://www.cyberscoop.com/coronavirus-cure-scam-social-media-riskiq/

Trend Micro.(2020). *Developing Story: COVID-19 Used in Malicious Campaigns.* https://www.trendmicro.com/vinfo/fr/security/news/cybercrime-and-digital-threats/coronavirus-used-in-spam-malware-file-names-and-malicious-domains

U.S. Department of Homeland Security. (2019). *Social Media Plan Guide: Science and Technology Directorate.* https://www.dhs.gov/sites/default/files/publications/social_media_plan_guide_09_20_2019.pdf

We Live Security & ESET (2020). *Threat Report. Q2 2020.* https://www.welivesecurity.com/wp-content/uploads/2020/07/ESET_Threat_Report_Q22020.pdf

MEDIA LITERACY AND DIGITAL SECURITY:

THE IMPORTANCE OF STAYING SAFE AND INFORMED

OAS

More rights
for more people