

January 2024
version 1

DISINFORMATION ON FACEBOOK: RESEARCH AND CONTENT MODERATION POLICIES



TABLE OF CONTENTS

INTRODUCTION	3
PLATFORM ORGANISATION	4
INVESTIGATIONS ON THE PLATFORM	10
HOW TO FLAG CONTENT ON THE PLATFORM AND ITS ENFORCEMENT	12
RELEVANT CASES ON HOW THE PLATFORM IS USED IN DISINFORMATION CAMPAIGNS	14

Author: **Maria Giovanna Sessa**, EU DisinfoLab

Reviewer: **Aleksandra Atanasova**, Reset

The opinions expressed are those of the authors and do not necessarily reflect the position of EU DisinfoLab. This factsheet does not represent an endorsement by EU DisinfoLab of any organisation.

Welcome to a comprehensive guide on how Facebook operates and how it can be employed for misinformative purposes. A different document will be compiled for Instagram as a different product of Meta. This resource is designed to empower the community engaged in countering disinformation on this platform. Information is updated as of 11 January 2024.

INTRODUCTION

This document shows how Facebook operates and can be exploited for disinformative purposes. It is designed as a resource for the counter-disinformation community to learn more about this platform.

The technical document is structured as follows: (I) First, the organisation of Facebook chats is presented; (II) Then, some tools on how to investigate the platform are introduced; (III) The subject of how to report content is addressed, including an overview of policy aspects related to enforcement. (IV) Finally, a repository of studies tackling Facebook use in disinformation campaigns is provided.

Facebook, founded by Mark Zuckerberg in 2004, has become one of the world's largest social media platforms, with around [2.9 billion](#) monthly active users worldwide and [259 million](#) monthly active users in the EU. Being a very large online platform (VLOP), Facebook is bound to additional transparency reporting requirements under Articles 15, 24, and 42 of the European Union's Digital Services Act (DSA). In view of this, it recently published its first [Transparency Report](#) for the period from 25 April 2023 to 30 September 2023.

Given its magnitude and longevity, it plays a significant role in the dissemination of information and, unfortunately, disinformation. On a positive note, Facebook has also made great efforts to counter the dissemination of false, misleading, and manipulated content with relevant policies.

The general requirements to join the platform are to be at least 13 years old, provide a real name, date of birth, gender, a valid email address or mobile phone number to verify and secure the account, and a password.

PLATFORM ORGANISATION

Understanding the structure of Facebook is crucial for analysis. Facebook's organisation encompasses profiles, groups, and pages, where disinformation often takes root and spreads.

PROFILES

- Personal **profiles** are the foundation of Facebook, where individuals share content with friends and followers starting from personal information, such as their name, profile picture, cover photo, hometown, workplace, or education history. Profiles may be **public**, where anyone can see the content, or **private**, with restricted access – while they also allow for intermediate levels of visibility.
- The **newsfeed** is a place where friends' posts, updates from pages and groups, as well as sponsored content in line with what the platform considers in line with the user's interests.
- The **timeline** is a user's record of all the posts they have shared and interacted with or the tagged content.
- Status and **posts** are often used interchangeably. However, Facebook uses the term 'post' to describe something the users share on their newsfeed, whether in the form of text, images, videos, links, or locations.
- The people the user either sent a request or has accepted a request from are known as **friends**. In contrast to various other social networking platforms, Facebook's friend system mandates mutual friendship between you and the other individual. By turning on [professional mode](#) on their profile, users can have unlimited followers. Moreover, [this](#) allows more monetisation options.
- Users automatically follow their friends. If a user sends someone a friend request that allows everybody to [follow](#) them, they will automatically follow their posts.

REACTIONS, COMMENTS, SHARES, AND MORE

Users can interact with posts and content via several features. The most [common](#) are:

- **Reactions:** Besides the traditional 'like' button, users can choose from a set of predetermined reactions, i.e., 'like', 'love', 'care', 'haha', 'wow', 'sad', and 'angry'.
- **Comments:** These are text-based responses that users can leave on a post.
- **Shares:** A shared post gets reposted on a user's timeline, making it visible to a tailored audience, including friends and followers.

Moreover, there are many other actions for users to act and interact on the platform, such as:

- [Answer a question:](#) Users can choose a predetermined question to answer or create their own question to answer and share it with their Facebook friends.
- [Ask for recommendations:](#) Users can ask their Facebook friends for recommendations on a topic.
- [Check-in:](#) A function to let a user's friends know where they are, for example, in a specific city or place.

- [Feeling/activity](#): A function allowing users to express their current mood or what they are doing at a given moment. These can also be used to complete a post, choosing from a list of predefined feelings (e.g., “feeling happy”) or activities (e.g. “watching a movie”).
- [Life event](#): Users can post about significant moments and milestones in their lives (e.g., graduating, buying a home, getting married, etc.).
- [Lists](#): Users can create Facebook lists as additional organisational tools for categorising their connections. Through list functionalities, updates can be shared with precision, targeting specific groups like co-workers or close friends. Users can also monitor updates from designated groups through those lists.
- [Poll](#): Users can ask their Facebook friends what their opinion is on a certain topic.
- [Reel](#): A video format complete with music, audio, and other options. Public reels from Instagram can be seen if the creator allows it. Like the overall Facebook algorithm, the Facebook Reels algorithm is powered by AI and machine learning, analysing interactions to determine interesting content for users.
- [Stories](#): A feature that allows users to share temporary photos and videos that disappear after 24 hours (unless the users save or highlight them). Friends and followers can view stories, and depending on one’s privacy settings, it is possible to reach and comment. Users can access insights on how the Stories are performing in terms of view counts and interactions.
- [Tag friends/event](#): Friends, Pages, and events can be tagged in a user’s status or post.

GROUPS

- [Groups](#) are online communities formed around shared interests or beliefs, where members can post content and resources to engage in discussions. Groups can be public (anyone can search and join them), private (anyone can search them, but joining requires approval), or secret (non-searchable and can be joined by invitation only), with varying levels of access control.
- Group administrators can exert control over content and membership. They can establish and enforce rules to ensure discussions remain focused and respectful.

PAGES

- [Pages](#) are public profiles originally designed for businesses, organisations, public figures, and other entities to create a presence on the social media platform and engage with their audience. Unlike personal profiles, Pages have an unlimited number of followers and likes. Followers can engage with the content created and shared by Page administrators by liking, sharing, and commenting. In addition, there is a reviews and recommendations section for feedback and rating.
- Page administrators can customise the Page profile by adding information such as the page’s name, category, contact details, website, a brief description, a profile picture, and a cover photo. Administrators can assign different roles (e.g., admin, editor, moderator, etc.) to individuals or team members to manage the Page. Page administrators can access insights and analytics (e.g., post reach, engagement, audience demographics, etc.).

- The [Page Transparency section](#) is a crucial tool for accountability. It includes the creation date, primary locations where the Page is managed if there are multiple admins, the number of people who manage it in each country, previous name changes, Page merges (on or after 6 September 2018), the confirmed (i.e., through Meta Verification) businesses or organisations that have claimed ownership of or have been granted access to help manage the Page, if it belongs to a state-controlled media organisation, if it is currently running ads, and if the page has run ads on political issues in the past. However, such information is not always available for unclear reasons.
- Pages can connect to [Facebook for Business](#) and run ad campaigns.

VERIFICATION

- The [verified badge](#) signifies that Facebook has validated a Page or profile as the genuine representation of the individual, public figure, or brand it portrays.
- It is now possible for profiles to subscribe to [Meta Verified](#).

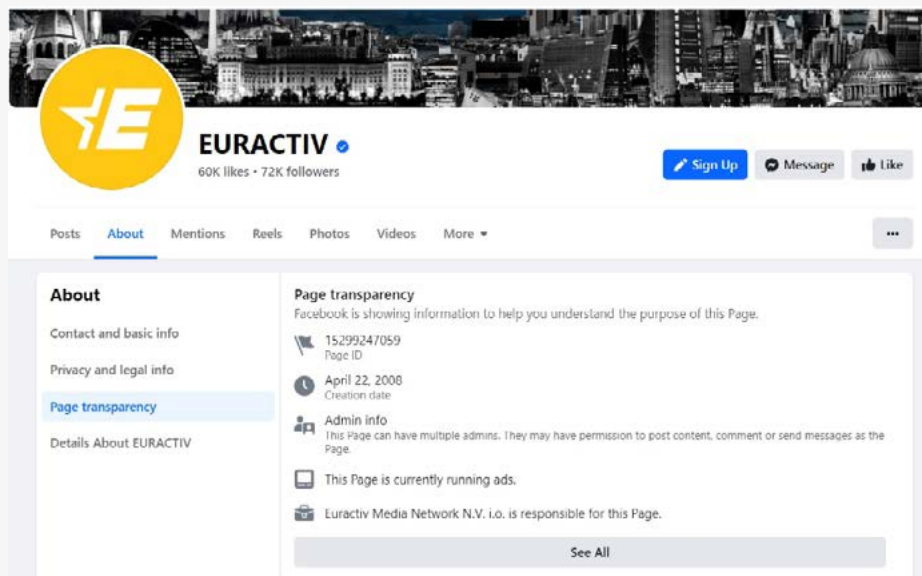


Figure 1. What a page transparency section looks like

MORE FEATURES

Ads

- Businesses can create [ads](#) on Facebook, choosing the type of audience they would like to reach, which is profiled based on the personal information shared on the platform and one's activities on Facebook and Instagram.
- “Meta adheres to the framework established by the Digital Advertising Alliance, European Interactive Digital Advertising Alliance and Digital Advertising Alliance of Canada, a coalition of organisations that develops privacy standards for online advertising”.
- Users can express their [ad preferences](#), but also opt-out from ads personalisation based on activities on and off Facebook. It is always possible to click on “[why am I seeing this ad](#)” to know why certain

ads are displayed. These insights are gathered through machine learning models.

- To address concerns from EU regulations about Meta’s ad targeting and data collection practices, the platform is launching a [paid subscription](#) that will remove ads from Facebook.

Ad Library

- The [Facebook Ad Library](#) is a feature that provides a publicly accessible repository of all ads that are currently active or have been active on Facebook or Instagram in order to promote transparency and accountability in advertising.
- Ad categories comprise housing, employment, credit, and “issues, elections or politics”. In the latter case, ads must complete Meta’s [authorisation process](#). However, the category is based on the self-declaration of the advertiser, potentially letting users see who is behind these ads, how much they have spent, and the intended audience by demographics (e.g., age, gender, location). It is unclear whether users can circumvent these requirements by providing inaccurate information.

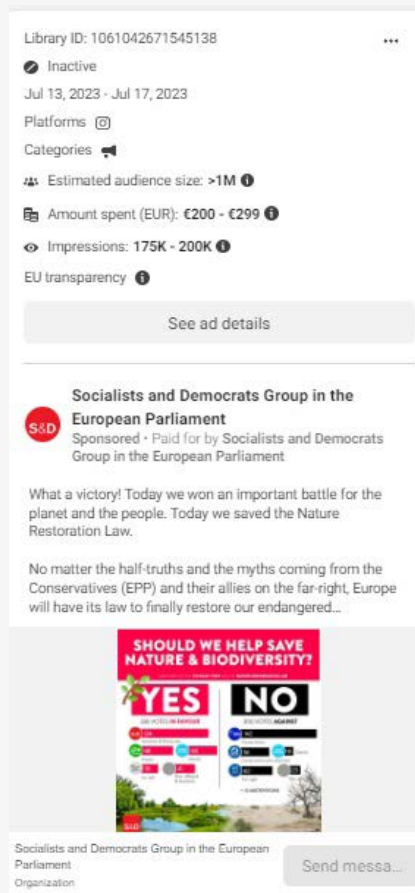


Figure 2. An inactive “issues, elections or politics” ad on the Ad Library

Chat and Messenger

- Facebook Chat is a built-in chat functionality on the main Facebook website or app. Being integrated into the Facebook interface, it does not require a separate app.
- Facebook Messenger is a standalone app that offers additional features, such as group chats, stickers, and games. It also allows users to communicate with people who are on their friends list by using their phone numbers.
- When users [message](#) their friends privately, the message gets delivered to their Facebook messages inbox and, if the person uses [Messenger](#), also to their Messenger app. Once viewed, the message will be marked as seen. Pro tip: by switching to airplane mode, it is possible to read the message without it getting marked as seen.

Dating

- [Facebook Dating](#) is a feature that allows users to create dating profiles and connect with other users who are also looking for romantic relationships. It is designed to help users find potential matches based on common interests, events, and groups they are part of on Facebook.
- First, users need to set up a separate dating profile that is not visible to their Facebook friends, which includes personal information and photos. The algorithm recommends potential matches based on the various information provided, with whom users can start text-based conversations.

Events

- Both individual users and businesses can choose to promote [events](#). These events can receive RSVPs, generate guest lists, and communicate essential details such as the event's date, time, and location. Events can take place in a virtual setting via a video platform or occur in person at a physical venue.

Fundraisers and donations

- Fundraisers (nonprofit) and donations (one-time or recurring) are [features](#) that enable users to create and contribute to charitable causes and support nonprofit organisations. They are designed to make it easier for individuals and organisations to raise funds for important causes and for users to donate to those causes directly through the Facebook platform.
- Facebook has different country-specific payment processing fees for donations.

Live

- [Facebook Live](#) is a feature of the social media platform Facebook that allows users to broadcast live videos to their followers and friends in real-time. Due to its design, the feature is expected to provide better user reach than other options, thanks to the pop-ups that urge users to join. Viewers can tune in to the live broadcast in real-time and interact with the broadcaster (e.g., through comments or reactions). After the live broadcast is finished, the video is saved to the broadcaster's profile or page so it can be accessed later as any other video on the platform.

Marketplace

- [Facebook Marketplace](#) is an online platform serving as a virtual location-based marketplace where individuals and businesses can buy and sell a wide variety of items, products, and services within

their local communities. The platform is free and does not charge any fees. Users can browse through various categories, use the search bar, filter and sort listings based on various criteria, and message sellers to ask questions, negotiate, or arrange a meeting.

Gaming

- [Facebook Gaming](#) is both a feature and an app. On the one hand, it is a gaming-related feature in the main platform, allowing users to stream and watch live gameplay, interact with gaming communities, and participate in gaming groups. On the other hand, it is also the name of a standalone mobile app designed specifically for gaming.

Meta Pay

- [Meta Pay](#) (previously known as Facebook Pay) is used to shop, donate, and send money across Meta-owned platforms and apps, providing a secure way to make online financial transactions within the company's ecosystem.

APPS AND FACEBOOK MOBILE APPS

Facebook apps are standalone applications or software developed by Facebook or third-party developers that serve specific purposes and can be installed separately on devices or platforms. Everyone is familiar with apps such as Instagram and WhatsApp; in addition, Facebook Messenger and Facebook Gaming have already been mentioned. There are many other separate apps owned by Meta (e.g., Facebook Analytics, Facebook Business Suite, Facebook Workplace, etc.), which lie beyond the scope of the present document focusing on Facebook.

METaverse

- Facebook's [Metaverse](#) is not an app or a feature but rather a concept and vision for the future of the Internet and social interaction beyond the current two-dimensional online experience. It represents an expansive, interconnected, and immersive digital space where people can interact, work, play, and create in a virtual environment through technologies like augmented reality (AR) and virtual reality (VR).

INVESTIGATIONS ON THE PLATFORM

FACEBOOK API

- The [Facebook API](#) is a tool that allows developers and app users to access a wide range of data from the Facebook platform; access levels apply. For example, accessible data comprises user profile information (e.g., name, location, lists of friends and followers, etc.); posts' reactions and comments (including associated texts, images, and videos); pages and groups (including the name, description, and members); events (including the date, time, location, and attendees); photos and videos (including the captions, tags, and associated location); data related to Facebook advertising campaigns, Marketplace, gaming interactions, and more.
- The [Facebook Graph API](#) is a specific API for interacting with the “social graphs”, i.e., a representation of the relationship between users.
- Data availability depends on the permissions granted by the user in order to ensure their privacy and security. Additionally, data is available through the API only to verified developers or with specific permissions. Developers are required to adhere to Facebook's Data Usage Policies, which outline how data can be collected, used, and shared. These policies emphasise the importance of protecting user data and respecting user privacy.

CROWDTANGLE

- [CrowdTangle](#) is a social media analytics tool (acquired by and integrated into Facebook) that provides valuable insights and data related to social media content, reach, engagement, and performance. To access CrowdTangle, organisations need to be vetted by Meta and registered.
- The tool allows users to monitor and identify trending and viral content through detailed metrics (e.g., reactions, shares, comments, etc.) and data-driven insights into the demographics and behaviour of the audience engaging with specific content. Users can create custom dashboards to track specific posts, pages, or hashtags. This allows for a tailored and focused analysis of social media data.
- It is worth noting that Meta is progressively dismantling CrowdTangle in countertendency with the [data access requirements](#) established by Article 40 of the Digital Services Act.

DATA SCRAPING

- Data scraping is the process of using automated tools or scripts to extract information from Facebook, including public profiles, posts, or pages. To efficiently scrape data from Facebook or other online sources, researchers and data analysts often rely on third-party tools or software. These tools can streamline the data collection process and allow users to extract, organise, and analyse the data more effectively.
- When the process entails [unauthorised scraping](#), it raises important ethical considerations regarding privacy (for instance, for content that users have not made public or consented to be collected), compliance with the platform's Terms of Service, data usage for legitimate research purposes, and other legal considerations (e.g., copyright laws, intellectual property rights, and data protection laws).

RESEARCH TOOLS

- “[Meta Content Library](#) is a web-based, controlled-access environment and does not allow data to be exported for analysis.” A deeper analysis of public content can be performed using the Content Library API. Both tools provide almost real-time public content from Facebook on posts, Pages, groups, and events.
- [Meta Ad Library](#) is a comprehensive, searchable, and exportable database (depending on file size) for ad transparency. People can use the Ad Library to get more information about the ads they see across Meta technologies. Only active ads on Meta are searchable, while ads about social issues, elections or politics are supposed to be visible whether they are active or inactive.
- The Transparency Centre also has a section called “[other research and datasets](#)”, mostly accessible to academic researchers.
- Meta publishes quarterly [Adversarial Threat Reports](#), which provide summaries and analyses of emerging cybersecurity concerns and adversarial activities targeting Facebook and Instagram. As these reports provide threat indicators, they are a useful resource to grasp insights on the platform.

HOW TO FLAG CONTENT ON THE PLATFORM AND ITS ENFORCEMENT

HOW TO REPORT CONTENT

On most content, users will see a menu represented by three dots (ellipsis) in the top right corner of the post, photo or comment they wish to report. In addition, it is also possible to report users. A dropdown menu appears, from which it is possible to click on “report post” (if the content belongs to a group, the user can also decide to “report post to group admins”). To report a post, it is possible to choose from a set of options, although Facebook warns, “If someone is in immediate danger, get help before reporting to Facebook. Don’t wait”. These options (called “problems”) are eating disorder, false information, harassment, hate speech, nudity, report as unlawful, spam, unauthorised sales, violence, terrorism, suicide or self-injury, or something else (being the same for Android, iOS app, MacOS, and Windows desktop). For each category, a few subcategories are listed, allowing the user to narrow the issue further, as displayed below. For example, as displayed below, “false information” can be about health, politics, social issues, or “something else”.

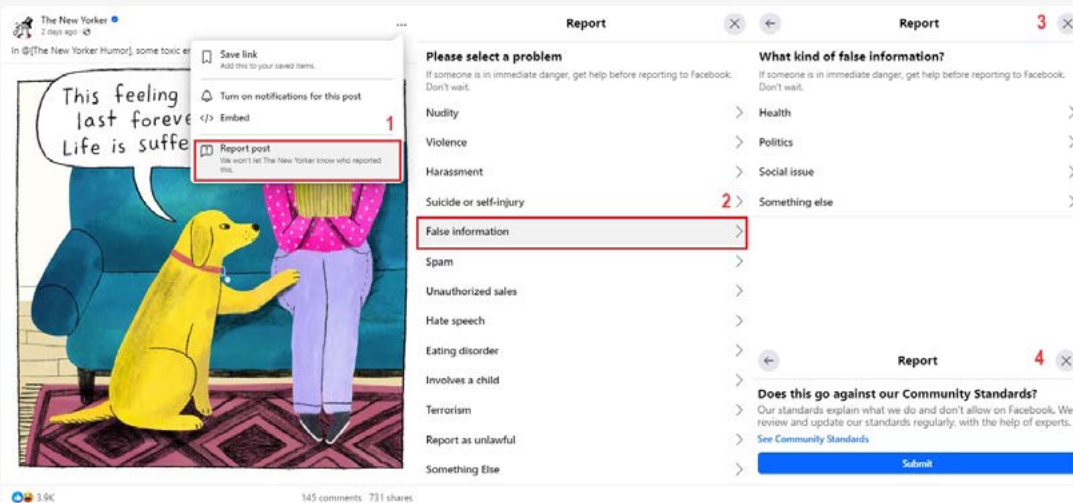


Figure 3. How to report content on Facebook

Content that goes against the platform’s Community Standards is removed using [automated technology](#) (via machine learning models) or a review team, but the number of times something is reported does not determine whether it is removed. Users usually have the chance to [request a review](#) of decisions they disagree with. An ulterior appeal can be made to the [Oversight Board](#).

Besides removal, Facebook has a [strike system](#) applied to offending accounts. Repeated violations that lead to further [restrictions](#) (e.g., removal from recommendations or demonetisation) and eventually the account being disabled. Nevertheless, if enough users report an account, it can be [shadowbanned](#).

PLATFORM POLICIES AGAINST DISINFORMATION

- Facebook’s [Community Standards](#) are guidelines that define what is allowed and prohibited on the platform, which comprises the previous list of content that can be reported. The main categories are violence and criminal behaviour, safety, objectionable content, integrity and authenticity, and intellectual property.

- The Community Standards have a section about [misinformation](#) outlining the rationale for removal – i.e., “the risk of imminent physical harm” and “content that is likely to directly contribute to interference with the functioning of political processes and certain highly deceptive manipulated media”. The types of misinformation Facebook removes are: (I) physical harm or violence; (II) harmful health misinformation; (III) voter or census interference; (IV) manipulated media.
- Misinformation that is debunked and in violation of the platform’s Community Standards is also classified as “unacceptable content” in Meta’s [Advertising Standards](#).
- Certain topics led to the development of specific policies. EU DisinfoLab analysed Facebook policies around [AI-related and manipulated](#), [electoral](#), [climate change](#), and [health](#) misinformation.
- The platform prohibits [Coordinated Inauthentic Behaviour \(CIB\)](#), and some related takedowns are reported in Meta’s quarterly Adversarial Threat Reports.
- Another helpful resource is the [Help Centre](#), a repository of information on a wide range of topics related to using the platform, including how-to guides, FAQs, technical support, and articles addressing different aspects of Facebook usage. For example, the Help Centre explains that Facebook works with [third-party independent](#) fact-checkers and is committed to reducing the distribution of false information.

RELEVANT CASES ON HOW THE PLATFORM IS USED IN DISINFORMATION CAMPAIGNS

This final section is by no means exhaustive but simply wishes to convey with a few examples the extent of Facebook's use to deceive and mislead.

- (2023) A [Chinese influence operation](#) positively promoting China and the Xinjiang province, where massive human rights abuses against the Uyghur minority are being committed, was taken down by Meta.
- (2022) An ongoing Russian-based influence operation network labelled [Doppelganger](#), operating in Europe since at least May 2022, used multiple clones of authentic media to spread pro-Kremlin disinformation and propaganda about the Ukraine war. Various Facebook Pages and fake profiles have actively amplified this operation.
- (2020) Facebook was a battleground for COVID-19 and vaccine-related disinformation, as [many studies show](#).
- (2020) Disinformation overflowed the [2020 U.S. elections](#), with a delayed response from Facebook to counter it. Besides QAnon and "Stop the Steal" conspiracies, electoral disinformation heavily targeted [voters of colour](#).
- (2019) Russian groups interfered with the [2019 EU elections](#), targeting Eurosceptic demographics on social media, including Facebook.
- (2019) During the [Yellow Vests](#)' protests in France, disinformation was widely circulated on Facebook, contributing to unrest and confusion.
- (2018) The [Cambridge Analytica](#) case concerned the unauthorised access and exploitation of personal data from Facebook users, implicating the homonymous data firm and Facebook. The scandal came to light in 2018 when it was revealed that Cambridge Analytica had harvested the personal information of up to 87 million Facebook users without their consent, using the data to provide analytical assistance to the 2016 presidential campaign of Donald Trump and Ted Cruz.