

## Factsheet

### Content Moderation and Freedom of Expression

*Prepared by Erik Tuchtfeld, Head of the humanet3 research group, Max Planck Institute for Comparative Public Law and International Law*

This Factsheet complements our Special Collection paper [Case law on content moderation and freedom of expression](#). It contains a selection of the most important decisions by national courts on content moderation. Content moderation is understood here in a broad sense: it does not only refer to the (non-) removal of content by intermediaries, such as social media platforms, but also includes moderation measures taken by public officials who administer social media pages, and cases in which governments try to force a social media platform to moderate in a specific manner. Just like the Collection, this Factsheet is structured taking into consideration the nature of the parties involved (user v. intermediary, user v. public official, state v. intermediary), and the content of the claim (removal or reinstatement of content). Within every section, key arguments or conclusions were highlighted for each case. For more detailed information about them, please refer to the corresponding footnote or online case analysis.

### Claims against intermediaries to reinstate content or accounts

[Cox. v. Twitter](#) 2019 (District Court for the District of South Carolina, Charleston Division, USA)<sup>1</sup>

---

[5-6] First, the undersigned agrees with the Defendant’s contention that it is immune from Plaintiff’s claims to the extent Plaintiff seeks to impose liability on the Defendant for declining to publish content created by the Plaintiff, as such claims are barred by Section 230(c) of the CDA.

[8] However, even assuming Plaintiff’s theory of liability is based on breach of contract, he has still failed to state a claim, as the “contract” Plaintiff had with the Defendant clearly provides that the Defendant reserves the right to remove content that it deems to have violated the User Agreement, including content constituting unlawful conduct or harassment.

---

<sup>1</sup> In this case, a Twitter user whose account was suspended after publishing a tweet criticizing Islam sued Twitter, seeking monetary damages and injunctive relief. The Court determined that Twitter qualified as publisher according to Section 230 of the Communications Decency Act and was immune against liability claims. Thus, it dismissed the claimant’s action.



# Global Freedom of Expression

## COLUMBIA UNIVERSITY

[FAN v. Facebook](#) 2019 (District Court for the Northern District of California, San Jose Division, USA)<sup>2</sup>

[13] It is “undisputed that the First Amendment of the United States Constitution only applies to government actors; it does not apply to private corporations or persons.”

[17-18] This Court has previously held that “private entities who creat[e] their own . . . social media website and make decisions about whether and how to regulate content that has been uploaded on that website” have not engaged in “public functions that were traditionally exclusively reserved to the state” [...] Thus, by operating its social media website, Facebook has not engaged in any functions exclusively reserved to the government.

[18-21] Plaintiffs argue that Facebook, a private actor, was a willful participant in joint action with the government because Facebook provided the government with information for the government’s investigation of Russian interference with the 2016 presidential election [...] Plaintiffs do not allege that the government played any role in shutting down FAN’s Facebook page or blocking FAN’s access to its Facebook account. Thus, Plaintiffs failed to allege any state action “directly by or jointly conceived, facilitated, or performed by the” government that relates to the deletion of FAN’s Facebook page or restriction on FAN’s access to its Facebook account.

[The Case on Facebook’s Terms of Service](#) 2021 (Federal Court, Germany)<sup>3</sup>

---

[59] The defendant is not bound by Art. 5 para. 1 sentence 1 GG in the same way as the state. As a private company, it is not directly bound by fundamental rights [...] With its network, it does offer a significant communication option within the Internet, but it does not guarantee access to the Internet as such.

[80-85] It also follows from the principle of practical concordance that the defendant's right to establish rules of conduct in its terms and conditions and to take measures to enforce them is not unrestricted. Rather, the defendant must take sufficient account of the users' fundamental right to freedom of expression [...] There must be an objective reason for the removal of content and the blocking of user accounts [...] network operators such as the defendant must make reasonable efforts to clarify the facts of the case [...] it is necessary that the defendant undertakes in its terms and conditions to immediately inform the user affected by a post removal, or account blocking, of the reasons for this and to give him the opportunity to respond, followed by a new description, which is accompanied by the possibility of making the removed post accessible again.

---

<sup>2</sup> Facebook had removed a Russian news distribution organization’s page, following the 2016 United States presidential election, on the grounds that it was one of the “inauthentic” accounts that had allegedly sought to inflame social and political tensions in the United States. The Court held that Facebook had not violated the First Amendment as it is neither a public forum nor do its actions amount to state action. Further, it also argued that the company, as a provider of interactive computer services, had immunity under the Communications Decency Act.

<sup>3</sup> In this case, the claimants asked Facebook to restate their posts and accounts, which were removed due to the publication of xenophobic content. The Court declared Facebook’s Community Standards invalid, as they did not include the procedural standards necessary to reach an adequate level of protection of freedom of expression. While Facebook is entitled to remove posts violating its Community Standards, even if they do not constitute illegal hate speech, it must inform the user and grant them an opportunity to respond and appeal the decision.

## Claims against intermediaries to remove content or accounts

[Delfi As v. Estonia](#) 2015 (European Court of Human Rights)<sup>4</sup>

---

[147] Anonymity has long been a means of avoiding reprisals or unwanted attention. As such, it is capable of promoting the free flow of ideas and information in an important manner, including, notably, on the Internet. At the same time, the Court does not lose sight of the ease, scope and speed of the dissemination of information on the Internet, and the persistence of the information once disclosed, which may considerably aggravate the effects of unlawful speech on the Internet compared to traditional media.

[157] Having regard to the fact that there are ample opportunities for anyone to make his or her voice heard on the Internet, the Court considers that a large news portal's obligation to take effective measures to limit the dissemination of hate speech and speech inciting violence – the issue in the present case can by no means be equated to “private censorship”. While acknowledging the “important role” played by the Internet “in enhancing the public’s access to news and facilitating the dissemination of information in general” [...] the Court reiterates that it is also mindful of the risk of harm posed by content and communications on the Internet.

[Sanchez v. France](#) 2023 (European Court of Human Rights)<sup>5</sup>

---

[190] The Court first observes that there can be little doubt that a minimum degree of subsequent moderation or automatic filtering would be desirable in order to identify clearly unlawful comments as quickly as possible and to ensure their deletion within a reasonable time, even where there has been no notification by an injured party.

[193] It further notes that the applicant had been free to decide whether or not to make access to the “wall” of his Facebook account public [...] The Court thus finds it legitimate to make a distinction, as the domestic courts did, between limiting access to the Facebook “wall” to certain individuals and making it accessible to the general public. In the latter case, everyone, and therefore especially a politician experienced in communication to the public, must be aware of the greater risk of excessive and immoderate remarks that might appear and necessarily become visible to a wider audience.

[Glawischnig-Piesczek](#) 2019 (European Court of Justice)<sup>6</sup>

---

[34-36] Article 15(1) prohibits Member States from imposing on host providers a general obligation to monitor information which they transmit or store, or a general obligation actively to seek facts or circumstances indicating

---

<sup>4</sup> An Estonian online news outlet had been held liable by domestic courts for defamation based on comments posted by readers in the comments section of its articles. The European Court of Human Rights decided that this does not violate freedom of expression.

<sup>5</sup> In this case, the Court held that the conviction of a politician for failing to promptly delete unlawful comments published by third parties on the public wall of his Facebook account did not violate his freedom of expression despite his apparent lack of knowledge of the comments, as his conviction was based on his “lack of vigilance and responsiveness” in monitoring his page for comments that could be unlawful.

<sup>6</sup> In this case, the claimant demanded that Facebook removes a particular piece of defamatory content as well as all “equivalent content”. The Court found that monitoring for identical and equivalent content to that which was declared illegal, would fall within the allowance for monitoring in a “specific case” and thus not violate the EU Directive’s general monitoring prohibition.



# Global Freedom of Expression

## COLUMBIA UNIVERSITY

illegal activity, as is clear from recital 47 of that directive, such a prohibition does not concern the monitoring obligations ‘in a specific case’ [...] Such a specific case may, in particular, be found, as in the main proceedings, in a particular piece of information stored by the host provider concerned at the request of a certain user of its social network, the content of which was examined and assessed by a court having jurisdiction in the Member State, which, following its assessment, declared it to be illegal [...] Given that a social network facilitates the swift flow of information stored by the host provider between its different users, there is a genuine risk that information which was held to be illegal is subsequently reproduced and shared by another user of that network.

[41] To be capable of achieving those objectives effectively, that injunction must be able to extend to information, the content of which, whilst essentially conveying the same message, is worded slightly differently, because of the words used or their combination, compared with the information whose content was declared to be illegal. Otherwise [...] the effects of such an injunction could easily be circumvented by the storing of messages which are scarcely different from those which were previously declared to be illegal.

### [Twitter v. Tammeh](#) 2023 (US Supreme Court)<sup>7</sup>

---

[22-23] The only affirmative “conduct” defendants allegedly undertook was creating their platforms and setting up their algorithms to display content relevant to user inputs and user history [...] The mere creation of those platforms, however, is not culpable. To be sure, it might be that bad actors like ISIS are able to use platforms like defendants’ for illegal—and sometimes terrible—ends. But the same could be said of cell phones, email, or the internet generally. Yet, we generally do not think that internet or cell service providers incur culpability merely for providing their services to the public writ large. Nor do we think that such providers would normally be described as aiding and abetting, for example, illegal drug deals brokered over cell phones—even if the provider’s conference-call or video-call features made the sale easier.

To be sure, plaintiffs assert that defendants’ “recommendation” algorithms go beyond passive aid and constitute active, substantial assistance. We disagree. By plaintiffs’ own telling, their claim is based on defendants’ “provision of the infrastructure which provides material support to ISIS” [...] Viewed properly, defendants’ “recommendation” algorithms are merely part of that infrastructure. All the content on their platforms is filtered through these algorithms, which allegedly sort the content by information and inputs provided by users and found in the content itself. As presented here, the algorithms appear agnostic as to the nature of the content, matching any content (including ISIS’ content) with any user who is more likely to view that content. The fact that these algorithms matched some ISIS content with some users thus does not convert defendants’ passive assistance into active abetting.

---

<sup>7</sup> Here, family members of a terrorist attack victim filed a lawsuit against Facebook, Google, and Twitter for providing substantial assistance to ISIS. They alleged that these social media companies knew that ISIS used their platforms to recruit people and raise funds for the attacks, yet failed to detect and remove their accounts, posts, and videos. Further, the respondents contended that the “recommendations” algorithm of these companies matched ISIS’ content with users more likely to be interested in their posts. The Supreme Court ruled that the companies were not liable for “aiding and abetting” the terrorist attack since the mere failure to remove the content could not constitute “substantial assistance” unless an independent duty to act was identified.

## Cases against public officials and institutions

[Knight First Amendment Institute v. Donald J. Trump](#) 2019 (US Court of Appeals for the Second Circuit)<sup>8</sup>

---

[23-24] To determine whether a public forum has been created, courts look “to the policy and practice of the government” as well as “the nature of the property and its compatibility with expressive activity to discern the government’s intent” [...] Opening an instrumentality of communication “for indiscriminate use by the general public” creates a public forum [...] The Account was intentionally opened for public discussion when the President, upon assuming office, repeatedly used the Account as an official vehicle for governance and made its interactive features accessible to the public without limitation. We hold that this conduct created a public forum. If the Account is a forum—public or otherwise—viewpoint discrimination is not permitted.

## Cases on state enforcement of private content moderation

[NetChoice v. Attorney General, State of Florida](#) 2022 (US Court of Appeals for the Eleventh Circuit)<sup>9</sup>

---

[25] Social-media platforms exercise editorial judgment that is inherently expressive. When platforms choose to remove users or posts, deprioritize content in viewers’ feeds or search results, or sanction breaches of their community standards, they engage in First-Amendment-protected activity.

[41-43] Social-media platforms are not—in the nature of things, so to speak—common carriers [...] While it’s true that social-media platforms generally hold themselves open to all members of the public, they require users, as preconditions of access, to accept their terms of service and abide by their community standards. In other words, Facebook is open to every individual if, but only if, she agrees not to transmit content that violates the company’s rules. Social-media users, accordingly, are not freely able to transmit messages “of their own design and choosing” because platforms make—and have always made—“individualized” content- and viewpoint-based decisions about whether to publish particular messages or users [...] Finally, Congress has distinguished internet companies from common carriers. The Telecommunications Act of 1996 explicitly differentiates “interactive computer services”—like social-media platforms—from “common carriers or telecommunications services.”

[SERAP v. Federal Republic of Nigeria](#) 2022 (ECOWAS Court)<sup>10</sup>

---

---

<sup>8</sup> In this decision, the court affirmed a lower court ruling that said that then-President Donald J. Trump engaged in unconstitutional viewpoint discrimination after he blocked users from his Twitter account for posting comments he disliked.

<sup>9</sup> In this case, the Court granted a preliminary injunction regarding specific provisions from a Florida Senate Bill which sought to “combat the ‘biased silencing’ of ‘our freedom of speech as conservatives [...] by the “big tech” oligarchs in Silicon Valley.’”

<sup>10</sup> In this case, the court held that the Nigerian government violated the Applicant’s right to freedom of expression and access to information and the media by suspending the operation of Twitter on June 4, 2021. The Nigerian authorities claimed the action was necessary to protect its sovereignty because the platform was being used by a separatist leader to sow discord.



# Global Freedom of Expression

## COLUMBIA UNIVERSITY

[67-68] It is clear from the analysis of Articles 9 of the ACHPR and 19 of the ICCPR that they did not only guarantee freedom of speech but also provided for a derivative right to access information, which is not a stand-alone right but a complementary right to the enjoyment of the right to freedom of expression [...] Therefore, the Court will hold that access to Twitter, being one of the social media of choice to receive, disseminate and impart information, is one such derivative right that is complementary to the enjoyment of the right to freedom of expression pursuant to the provisions of Article 9 (1) & (2) of the ACHPR and Article 19 of the ICCPR.

[par. 85] Whilst the Court agrees with the Respondent that the enjoyment of the right to freedom of speech is not absolute and that the same is exercised within the ambit of the laws, the Court is of the opinion that any limit to regulate the exercise of this right or to derogate from it must be expressly and specifically provided for by legislation for that purpose and the same must not have a retrospective effect in terms of its application. The Court, however, notes that the Respondent in all its statement of defense, has failed to adduce proof or evidence referencing a particular law or Court order or otherwise justifying the suspension of Twitter.