

# LIBERTAD DE EXPRESIÓN, ODIO Y REDES SOCIALES: LAS OPCIONES DEL LEGISLADOR DEMOCRÁTICO ESPAÑOL

Freedom of expression, hate and social networks: the options of the Spanish democratic legislator

EVA SÁENZ ROYO

Universidad de Zaragoza

[evasaenz@unizar.es](mailto:evasaenz@unizar.es)

## *Cómo citar/Citation*

Sáenz Royo, E. (2024).

Libertad de expresión, odio y redes sociales: las opciones del legislador democrático español.

*Revista Española de Derecho Constitucional*, 132, 71-103.

doi: <https://doi.org/10.18042/cepc/redc.132.03>

## **Resumen**

La libertad de expresión y sus límites han adquirido nuevas dimensiones en la era digital. Resulta evidente que se ha ampliado el campo de acción de la libertad de expresión a través del uso de internet. Sin embargo, las ventajas de este medio van acompañadas de una serie de riesgos. Contenidos difamatorios, odiosos o violentos pueden difundirse como nunca antes por todo el mundo en cuestión de segundos y hacerse bajo la impunidad del anonimato. Esta nueva realidad ha tenido un doble efecto.

Por una parte, las instituciones públicas intentan poner límites a las expresiones de odio en las redes. A este respecto se plantea si una ley puede exigir responsabilidades a las plataformas de internet por el discurso de odio vertido por terceros o cuándo puede exigirse a esas plataformas que retiren el contenido y exigirles responsabilidad en caso de no hacerlo.

Por otra parte, las grandes plataformas de internet (Twitter —ahora X—, Facebook, Instagram o Google) se han erigido en las controladoras de esas expresiones de odio creando sus propias normas para controlar el contenido que circula en ellas. A este respecto las cuestiones que se plantean es si son las plataformas digitales libres a

la hora de moderar sus contenidos y la eficacia mediata o inmediata que tiene el derecho a la libertad de expresión de los usuarios respecto a estas plataformas.

Se trata de dos cuestiones complejas que exigen respuestas coherentes para ambas. Sin existir todavía en España un marco jurídico definido para estas cuestiones, en este trabajo abordaremos las posibilidades que tiene el legislador democrático español a partir de un análisis de la normativa y la jurisprudencia que sobre estas dos cuestiones se han desarrollado tanto en Estados Unidos como en la Unión Europea

### ***Palabras clave***

Libertad de expresión; odio; redes sociales.

### ***Abstract***

Freedom of expression and its limits have taken on new dimensions in the digital age. It is evident that the scope of freedom of expression has expanded through the use of the Internet. However, the advantages of this medium are accompanied by a series of risks. Defamatory, hateful or violent content can be spread like never before around the world in a matter of seconds and done so with impunity and anonymity. This new reality has had a double effect.

On the one hand, public institutions try to put limits on expressions of hate on the networks. In this regard, the question arises whether a law can hold Internet platforms accountable for hate speech expressed by third people or when these platforms can be required to remove the content and held accountable if they do not do so.

On the other hand, the large Internet platforms have become the controllers of these expressions of hate, creating their own rules to control the content that circulates on them. In this regard, the question that arises is whether digital platforms are free when it comes to moderating their content and the mediate or immediate effectiveness of the right to freedom of expression of users with respect to these platforms.

These are two complex questions that require coherent answers for both. Without a defined legal framework for these issues still existing in Spain, in this work we will address the possibilities that the Spanish democratic legislator has based on an analysis of the regulations and jurisprudence that have been developed on these two issues both in the United States and in the European Union

### ***Keywords***

Freedom of expression; hate; social networks.

## SUMARIO

---

I. INTRODUCCIÓN. II. LA INMUNIDAD DE LAS GRANDES PLATAFORMAS RESPECTO A LAS EXPRESIONES DE ODIO DE TERCEROS Y LA LIBRE MODERACIÓN DE CONTENIDOS EN EE. UU.: 1. La inmunidad en EE. UU. 2. El poder de moderación de las grandes plataformas en EE. UU. III. EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS RESPECTO A LAS EXPRESIONES DE ODIO DE TERCEROS Y LA MODERACIÓN DE CONTENIDOS EN EUROPA: 1. El odio en las redes sociales y el «deber de diligencia» de las grandes plataformas en Europa. 2. El poder de moderación de las grandes plataformas en Europa. IV. EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS Y SU PODER DE MODERACIÓN EN ESPAÑA: 1. El odio en las redes y el «deber de diligencia» de las grandes plataformas en España. 2. El poder de moderación de las grandes plataformas en España. V. OPCIONES DE LEGE FERENDA DEL LEGISLADOR DEMOCRÁTICO ESPAÑOL: 1. La regulación de la responsabilidad de las plataformas digitales respecto a las expresiones de odio en las redes sociales. 2. La regulación del poder de moderación de las plataformas digitales. BIBLIOGRAFÍA.

---

## I. INTRODUCCIÓN

La libertad de expresión y sus límites han adquirido nuevas dimensiones en la era digital. Como ha señalado el Tribunal Europeo de Derechos Humanos,

[...] la posibilidad de que los individuos se expresen en internet constituye una herramienta sin precedentes para el ejercicio de la libertad de expresión. Se trata de un hecho indiscutible. Sin embargo, las ventajas de este medio van acompañadas de una serie de riesgos. Contenidos claramente ilícitos, incluido el difamatorio, odioso o violento, pueden difundirse como nunca antes por todo el mundo en cuestión de segundos, y a veces permanecer en línea durante mucho tiempo» (asunto *Delfi AS v. Estonia*, 16 de junio de 2015, § 110).

En la misma línea se ha pronunciado el TJUE al señalar: «Además, el efecto de la injerencia en dichos derechos del interesado se multiplica debido al importante papel que desempeñan Internet y los motores de búsqueda en la sociedad moderna, que confieren a la información contenida en tal lista de resultados carácter ubicuo» (caso C31/2012, *Google Spain*, § 80).

De esta forma las expresiones de odio contra una persona o grupo de personas han encontrado en el mundo digital un campo abonado para su proliferación y penetración mediante el uso de los *bots*<sup>1</sup> u otro tipo de técnicas de comunicación digital. La enorme difusión del discurso del odio en las redes sociales y el anonimato que va detrás de este han tenido un doble efecto.

Por una parte, las instituciones intentan poner límites al discurso de odio en las redes. A este respecto se plantea si una ley puede exigir responsabilidades a las plataformas de internet por el discurso de odio vertido por terceros amparados por el anonimato o cuándo puede exigirse a esas plataformas que retiren el contenido y exigirles responsabilidad en caso de no hacerlo.

Por otra parte, las grandes plataformas de internet (Twitter —ahora X—, Facebook, Instagram o Google) se han erigido en las controladoras de esas expresiones de odio creando sus propias normas para controlar el contenido que circula en ellas. A este respecto las cuestiones que se plantean es si son las plataformas digitales libres a la hora de moderar sus contenidos y la eficacia mediata o inmediata que tiene el derecho a la libertad de expresión de los usuarios respecto a estas plataformas.

Se trata de dos cuestiones complejas que exigen de respuestas coherentes para ambas. Sin existir todavía en España un marco jurídico definido para estas cuestiones, en este trabajo abordaremos las posibilidades que tiene el legislador democrático español a partir de un análisis de la normativa y la jurisprudencia que sobre estas dos cuestiones se han desarrollado tanto en EE. UU. como en la Unión Europea.

## II. LA INMUNIDAD DE LAS GRANDES PLATAFORMAS RESPECTO A LAS EXPRESIONES DE ODIOS DE TERCEROS Y LA LIBRE MODERACIÓN DE CONTENIDOS EN EE. UU.

### 1. LA INMUNIDAD EN EE. UU.

Tempranamente en EE. UU. se estableció la inmunidad de los proveedores de servicios por los contenidos publicados por terceros. Fue en 1996, a través del art. 230 de la *Communications Decency Act*. Se estableció esta

<sup>1</sup> Programas informáticos que efectúa automáticamente tareas repetitivas a través de internet. Imitan el comportamiento humano, actúan a través de perfiles falsos en redes sociales y se pueden emplear para difundir contenidos. El problema se encuentra en las llamadas granjas de *bots*, en las que una o varias personas gestionan miles de bots. *Vid.* Sánchez Muñoz, 2020: 35.

inmunidad de los proveedores de servicios por los contenidos publicados por terceros incluso si las plataformas actúan «de buena fe para restringir el acceso o disponibilidad de material que el proveedor o usuario considere ser obsceno, lascivo, sucio, excesivamente violento, acosador o de otra manera objetable». Es decir, se reconocía la inmunidad de las plataformas, aunque ejercieran un cierto control editorial sobre discursos «objetables»<sup>2</sup>.

Esta ley fue la respuesta que dio el Congreso frente al caso *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995), donde se consideró que la plataforma ejercía un cierto control editorial sobre los contenidos que se publicaban y podía ser responsable de los contenidos que se publicaban por terceros. A través de esta ley quedó clara la inmunidad de los proveedores de servicios de los contenidos publicados por terceros, salvo que las propias plataformas sean parte en la «creación o desarrollo» del contenido.

En EE. UU., por tanto, las grandes plataformas digitales no son consideradas editores, a diferencia de los medios de comunicación tradicionales, como radio, televisión y prensa, y ello se basó en que «los operadores de internet eran neutrales y no participaban en el proceso comunicativo» (Teruel Lozano, 2023: 185). En este sentido, Thompson (2016: 785-786) habla de *neutrality thesis*, según la cual las plataformas no pueden adoptar decisiones sobre la licitud de los contenidos, lo que debe estar reservado a las autoridades judiciales, para evitar la «censura colateral» y limitarse a ser esas plataformas simples implementadoras de las decisiones judiciales.

La tendencia jurisprudencial ha sido interpretar de manera muy extensiva esta inmunidad, de manera que, solo si resulta probado que la plataforma ha participado directamente en la creación del contenido cuya ilicitud también

<sup>2</sup> «47 U.S. Code § 230—Protection for private blocking and screening of offensive material

(c)Protection for “Good Samaritan” blocking and screening of offensive material

(1)Treatment of publisher or speaker

No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.

(2) Civil liability No provider or user of an interactive computer service shall be held liable on account of—

(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph».

ha sido probada, perdería su inmunidad. Según recientes pronunciamientos jurisprudenciales, esa inmunidad se mantiene, aunque se haya ayudado a la proliferación de discursos de odio y contenidos terroristas a través de las recomendaciones de los algoritmos<sup>3</sup>. En estos pronunciamientos se entiende que las recomendaciones algorítmicas son «simplemente parte de la infraestructura a través de la cual todo el contenido de sus plataformas es filtrado», que es el principal argumento utilizado por las plataformas para mantener su inmunidad. No se considera, por tanto, ejercicio de una decisión «editorial».

Hay que advertir que la inmunidad de las grandes plataformas reconocida legalmente es poco compatible con la interpretación que han hecho esas mismas grandes plataformas de su poder de moderación. Lo vemos a continuación.

## 2. EL PODER DE MODERACIÓN DE LAS GRANDES PLATAFORMAS EN EE. UU.

En EE. UU. la jurisprudencia del Tribunal Supremo, desde *Reno vs. American Civil Liberties Union* 521 U.S. 844, de 1997, parte de la idea de que el ciberespacio y las redes sociales constituyen los nuevos foros públicos donde tiene lugar el proceso de la comunicación pública y, por esta razón, están protegidos por la primera enmienda, que protege la libertad de expresión. Los poderes públicos están obligados a asegurar su libertad y accesibilidad y cualquier normativa que restrinja el acceso a estos foros constituye una injerencia en la libertad de expresión que debe someterse a un escrutinio constitucional estricto. Así, en *Packingham v. North Carolina*, 582 U.S. 15-1194, de 19 de junio de 2017, el Tribunal Supremo declaró contraria a la primera enmienda una ley de Carolina del Norte de 2008 que prohibía el uso de redes sociales a personas condenadas por delitos sexuales. En esta misma línea, se considera que los representantes públicos en las redes sociales tampoco pueden bloquear

<sup>3</sup> En el caso *Twitter v. Taamneh*, 598 U.S. (2023), resuelto el 18 de mayo de 2023, se alega ante la Corte Suprema la no aplicación del art. 230 de la *Communications Decency Acts*, sino la *Antiterrorism Act*, al entender que con las recomendaciones del algoritmo las plataformas «brindaron asistencia sustancial a sabiendas» en virtud de la ley y «ayudaron e instigaron» un acto de terrorismo internacional. Sin embargo, la Corte Suprema considera que las recomendaciones del algoritmo no cumplen las condiciones para considerarse una «asistencia sustancial a sabiendas» y considera que no es aplicable la *Antiterrorism Act*. Véase <https://tinyurl.com/mpsja8jb> (consulta: 26 de julio de 2023). En el mismo sentido se resuelve *Gonzalez v. Google*, 598 U.S. (2023). Véase <https://tinyurl.com/u4nz5dhy> (consulta: de 26 de julio 2023).

a usuarios de su perfil privado. Así, en *Knight First Amendment Inst. At Columbia Univ. Vs. Trump*, se declara que Donald Trump había vulnerado el derecho a la libertad de expresión de una serie de usuarios bloqueados desde su perfil privado de Twitter cuando era presidente. Para reconocer esta vinculación el tribunal considera que la cuenta privada tiene que ser usada por alguien que desempeña un cargo público y, además, tiene que publicar información relacionada con ese cargo. Se observa así la vinculación de la libertad de expresión respecto a los poderes públicos, que no son libres de limitar los contenidos en las redes sociales. La cuestión es si las plataformas están igualmente vinculadas por la libertad de expresión de los usuarios y si, por tanto, no son libres a la hora de limitar el contenido que en ellas se publica. O si son los tribunales o el legislador democrático los que pueden restringir esa libertad de moderación.

La jurisprudencia del Tribunal Supremo en EE. UU. ha reiterado que la primera enmienda solo es eficaz frente a restricciones de la libertad de expresión procedentes del Gobierno y no para las procedentes de compañías privadas (entre otros muchos, asunto *Williby v. Zuckerberg*, n. 3:18-cv-06295-JD, Dkt. N. 19 at 1 [ND Cal. June 18, 2019]). Solo resulta eficaz la libertad de expresión frente a empresas privadas si se demuestra que están operando bajo el efecto del Gobierno según la *state-action doctrine*. Esta doctrina señala que, en algunas situaciones, la autoridad gubernamental domina una actividad hasta tal punto que se debe considerar que los participantes actúan con la autoridad del gobierno y deben estar sujetos a las obligaciones constitucionales (*Edmonson v. Leesville Concrete Co.*, 500 U.S. 614, 620 [1991]). Por tanto, en EE. UU. los derechos constitucionales no resultan eficaces directamente frente a los particulares y solo el legislador democrático puede establecer esa vinculación.

Este modelo se ha confirmado por la Corte del Distrito Norte de California en el caso *Trump v. Twitter*, donde se ha desestimado la pretensión de Trump de que se considerara vulnerada su libertad de expresión al haber sido bloqueado por Twitter. Según este Tribunal, los derechos fundamentales como la libertad de expresión de los usuarios no pueden alegarse frente a los particulares y solo si se demostrara que Twitter ha actuado siguiendo las instrucciones del poder público, conforme a la *state-action doctrine*, podría considerarse vulnerado el derecho. No obstante, el Tribunal considera que alegar las declaraciones de algunos congresistas demócratas a favor del bloqueo de la cuenta de la Trump no es suficiente para aplicar la *state-action doctrine*, sino que sería necesaria una «concreta y específica acción gubernamental» que probara que la compañía privada actúa como poder público, algo que no se prueba en este caso concreto (Sentencia del Tribunal del Distrito Norte de California de 6 de mayo de 2022, asunto *Trump et al. v. Twitter*, c. 21-cv-09378-JD). Se niega

así en EE. UU. la eficacia directa de los derechos constitucionales frente a las empresas privadas, siendo el legislador democrático el único que puede establecer esa vinculación.

Como ya hemos señalado más arriba, en 1996, a través del art. 230 de la *Communications Decency Act*, se estableció en EE. UU. la inmunidad de los proveedores de servicios por los contenidos publicados por terceros incluso si las plataformas actúan «de buena fe para restringir el acceso o disponibilidad de material que el proveedor o usuario considere ser obsceno, lascivo, sucio, excesivamente violento, acosador o de otra manera objetable». Es decir, se reconocía la inmunidad de las plataformas, aunque ejercieran un cierto control editorial sobre discursos «objetables». Por tanto, aunque pudiera resultar paradójico, se establecía que las plataformas no deben considerarse editores a efectos de su responsabilidad por los contenidos publicados por terceros, pero se les reconoce cierto control editorial en esos contenidos, cuando tradicionalmente el reconocimiento de la libertad editorial y derecho de veto a los medios de comunicación tradicionales conlleva, lógicamente, su responsabilidad solidaria por el contenido que una persona, generalmente un periodista, transmita a través de ellos<sup>4</sup>.

Por tanto, el legislador democrático decidió reconocer expresamente un concreto control editorial o derecho de veto a las grandes plataformas. Es decir, internet es un foro público donde se ejerce la libertad de expresión, pero el derecho a la libertad de expresión de los usuarios, si bien vincula en todo caso a los poderes públicos, no es así respecto a las grandes plataformas, a las que, además, se les reconoce legalmente un derecho de veto concreto: si actúan «de buena fe para restringir el acceso o disponibilidad de material que el proveedor o usuario considere ser obsceno, lascivo, sucio, excesivamente violento, acosador o de otra manera objetable».

Estas plataformas interpretaron este derecho de veto reconocido legalmente de manera muy amplia. Por una parte, fue interpretado por las plataformas como libertad editorial que permitía eliminar libremente cualquier

---

<sup>4</sup> Vázquez Alonso, 2022: 110-111. Como el propio autor señala, «no deja de ser contradictorio. Y es que si, por un lado la idea de neutralidad de la red sirve de fundamento para eximir de responsabilidad a las empresas de intermediación de contenidos, incentivado así la ausencia de controles que puedan provocar un “efecto desaliento” en el ejercicio de la libertad de expresión, por otro lado esa misma impunidad jurídica por la censura de contenidos que estas decidan llevar a cabo parece jugar en sentido contrario. Es decir, permite asimilar esta labor de control y moderación a una labor editorial clásica, solo que, en este caso, con una prerrogativa inexistente en otros ámbitos del Derecho de la comunicación: la irresponsabilidad por los contenidos que a través de ella sean difundidos» (*ibid.*: 112).

contenido publicado por sus usuarios. Por otra parte, estas plataformas entienden que esta libertad editorial estaría amparada, no solo por la *Communications Decency Act*, sino por la primera enmienda, que protege y garantiza la libertad de expresión —en este caso de las plataformas—. La imposición legal o judicial de la obligación de mantener una determinada publicación en contra de su criterio o sus propias políticas de moderación de contenidos sería *compelled speech* y, por tanto, una violación de la primera enmienda de la Constitución de EE. UU. (Barata Mir, 2022: 104).

Con ello la libertad de moderación de contenidos ha resultado en unas regulaciones discrecionales, ya que no son aplicadas de manera estandarizada para todos los casos, y poco transparentes para los usuarios, ya que no dan explicaciones ni abundan en detalles sobre el motivo por el cual toman una decisión determinada y la forma de apelación es confusa<sup>5</sup>.

Pero, en el caso de EE. UU., donde la *Communications Decency Act* reconoce ya una concreta libertad de moderación a las plataformas, ¿pueden los Estados, en ejercicio de su poder legislativo, establecer algún tipo de límite al poder de moderación que no vulnere la libertad editorial reconocida en la legislación federal?

Ha habido algunos intentos legislativos por establecer vinculación de la libertad de expresión de los usuarios respecto a las grandes plataformas. Se trataría de establecer legalmente la obligación de las plataformas de facilitar el acceso a internet y el uso de sus servicios libremente y en igualdad de condiciones para cualquier usuario sin que se pueda discriminar o monitorizar por razón de contenidos (la llamada *Common Carriage Doctrine*) (Villaverde Menéndez, 2020: 87).

La *Executive Order on Preventing Online Censorship*, de 28 de mayo de 2020<sup>6</sup>, fue la respuesta de Trump ante la calificación por Twitter de dos de sus tuits como «potencialmente engañosos». Esta orden proponía que reguladores federales como la Comisión Federal de Comercio puedan sancionar a compañías tecnológicas sobre los criterios que usan para moderar contenidos. Así, se abría el camino a que cualquier usuario pudiera emprender medidas legales si considerara que las razones por la que su cuenta ha sido suspendida o su mensaje borrado hubieran sido discriminatorias o no se hubieran basado en la «buena fe», haciendo así una interpretación estricta de la libertad editorial reconocida en la *Communications Decency Act*. Las razones que se dan en dicha

<sup>5</sup> En Facebook se ha creado su *Oversight Board*, que fija unos estándares en la resolución de los *leading case* y a través de recomendaciones generales. Una visión crítica de este órgano, en Cetina Presuel, 2021: 499-556.

<sup>6</sup> Véase <https://tinyurl.com/6m7fhkmt> (consulta: 2 de enero de 2023).

orden ponen de manifiesto la contradicción que supone el reconocimiento de una libertad editorial a la vez que su inmunidad por los contenidos vertidos por terceros. Señala que, si estas plataformas no son estrictamente neutrales en su función de intermediación, han de ser consideradas como medios de comunicación y asumir la responsabilidad por los contenidos vertidos por terceros. En mayo de 2021 Biden revocó dicha orden.

En ese mismo mes, mayo de 2021, se aprobó en Florida una ley que permitía sancionar redes sociales que expulsaran o sometieran a filtrado algorítmico a candidatos políticos, o cancelaran las cuentas de empresas periodísticas. Esta ley fue anulada por la Corte para el Distrito Norte de Florida. Entre otros argumentos, la Corte afirma que esta ley sería contraria a la sección 230 de la *Communications Decency Act*, pero, además, se señala que las plataformas son titulares del derecho reconocido en la primera enmienda, y, por ello, no pueden ser obligadas a publicar o mantener contenidos en contra de su propio criterio (Barata Mir, 2021).

En septiembre de 2021 en Texas se aprobó una ley que prohíbe que las plataformas de redes sociales con al menos cincuenta millones de usuarios activos bloqueen, eliminen o desmoneticen contenido según las opiniones de los usuarios. En coherencia con la ley federal, la ley estatal permite que las plataformas censuren ciertos discursos que «incitan a la actividad criminal o conlleven amenazas específicas». En el caso *NetChoice v. Paxton*, la compañía de redes sociales argumenta que su libertad editorial estaría amparada por la primera enmienda. Es interesante la argumentación del tribunal de apelaciones que, el 16 de septiembre de 2022, y contrariamente a lo ocurrido en Florida, ha fallado a favor del Estado. Señala este tribunal que la primera enmienda no puede proteger el derecho a censurar reclamado por las plataformas y que, por tanto, las leyes que prohíben la censura por parte de las plataformas están protegiendo la libertad de expresión de los usuarios. Considera este tribunal que esta solución es la más coherente con la no consideración de las plataformas como medios de comunicación y la inmunidad de la que gozan por las opiniones de terceros establecida en la secc. 230 de la *Communications Decency Act*. Señala expresamente el tribunal:

La sección 230 refleja el juicio del Congreso de que las Plataformas no funcionan como editores tradicionales y no están «hablando» con el contenido enviado por los usuarios. El juicio del Congreso refuerza nuestra conclusión de que la censura de las plataformas no es un discurso bajo la primera enmienda<sup>7</sup>.

[...]

---

<sup>7</sup> Véase <https://tinyurl.com/ydsm4rsu> (consulta: 28 de julio de 2023), pp. 39-44.

La sección 230 socava los dos argumentos de las plataformas para sostener que su censura de usuarios es discurso protegido. Hay que recordar que se basan en dos argumentos clave: primero, sugieren que el contenido enviado por el usuario que alojan es su discurso; y segundo, argumentan que son editores similares a un periódico. La sección 230, sin embargo, instruye a los tribunales a no tratar las plataformas como «editores u oradores» del contenido enviado por el usuario que alojan. Identificación. § 230(c)(1). Y esas son exactamente las dos categorías que invocan las plataformas para apoyar su argumento de la primera enmienda. Entonces, si § 230(c)(1) es constitucional, ¿cómo puede un tribunal reconocer que las plataformas son editores u oradores de los contenidos que alojan sus usuarios y que están protegidas por la primera enmienda?

Y, aunque según la ley, resultan inmunes incluso si las plataformas actúan «de buena fe para restringir el acceso o disponibilidad de material que el proveedor o usuario considere ser obsceno, lascivo, sucio, excesivamente violento, acosador o de otra manera objetable», el tribunal señala que este control editorial debe interpretarse restrictivamente y que no da libertad editorial a las plataformas para censuras basadas en la ideología o en la geografía (*viewpoint-based or geography-based censorship*) y que no confiere un derecho a censurar libremente. Por tanto, el Parlamento de Texas está legitimado para evitar que las plataformas vulneren la libertad de expresión de los usuarios de Texas.

Tal y como puede observarse, frente a la libertad de moderación amparada en la primera enmienda que reclaman las grandes plataformas —y que resulta contradictoria con su irresponsabilidad respecto al contenido vertido por terceros—, se empieza a abrir una línea jurisprudencial que entiende que su libertad de moderación se basa estrictamente en lo reconocido por la legislación federal y que puede ser modulada por el legislador estatal. Veremos lo que señala la Corte Suprema.

### III. EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS RESPECTO A LAS EXPRESIONES DE ODIOS DE TERCEROS Y LA MODERACIÓN DE CONTENIDOS EN EUROPA

#### 1. EL ODIOS EN LAS REDES SOCIALES Y EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS EN EUROPA

El modelo de irresponsabilidad por los contenidos vertidos por terceros en las redes fue recibido en Europa a través de la Directiva 2000/31/CE del

Parlamento Europeo y del Consejo, de 8 de junio de 2000, sobre comercio electrónico. Esta irresponsabilidad se ha mantenido en el Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, que modificó la Directiva (Reglamento de Servicios Digitales).

El Reglamento de Servicios Digitales (art. 16), en línea con la Directiva sobre comercio electrónico, prevé que las plataformas son, en general, irresponsables por los contenidos ilícitos vertidos por terceros en la red, y solo serán responsables por esos contenidos cuando desempeñen un «papel activo» en el contenido del mensaje (art. 18) o «colabore[n] deliberadamente» (art. 20).

Por otra parte, el art. 22 establece un «deber de diligencia» que se traduce en la obligación de retirar el contenido en el momento en el que el prestador del servicio tenga «conocimiento efectivo» de su carácter ilícito<sup>8</sup>. En este último caso, nótese que no se trata propiamente de la imputación de responsabilidad por el contenido vertido por un tercero, sino más bien de un sistema de reparación de la vulneración producida por un tercero consistente en un «deber de diligencia» para la plataforma en la eliminación del contenido.

El problema con este «deber de diligencia» es que no solo se puede establecer la responsabilidad cuando se trate de no retirar contenidos ilícitos, así declarados por la autoridad competente tras la ponderación de los derechos en juego (derecho al honor o a la dignidad del ofendido frente al derecho a la libertad de expresión del usuario), sino que también pueden ser responsables por no retirar contenidos controvertidos cuando han recibido notificaciones que «sean suficientemente precisas y estén adecuadamente fundamentadas para que un operador económico diligente, de manera razonable, pueda detectar y evaluar el contenido presuntamente ilícito y, en su caso, actuar contra él» (art. 22; en el mismo sentido, art. 53). Siendo que «[l]a retirada o bloqueo del acceso debe llevarse a cabo respetando los derechos fundamentales de los destinatarios del servicio, incluido el derecho a la libertad de expresión y de información», se obliga a la plataforma a hacer una ponderación entre los derechos fundamentales en juego que se suele reservar a las autoridades judiciales (así, el art 20.5 CE).

Por último, al permitir que autoridades independientes controlen los sistemas algorítmicos de las grandes plataformas (art. 92), parece apartarse de la consideración en EE. UU. de que las recomendaciones algorítmicas son «simplemente parte de la infraestructura a través de la cual todo el contenido

---

<sup>8</sup> En el caso de difusión de contenidos terroristas se da un paso más en el Reglamento 2021/784, de 29 de abril, sobre la lucha contra la difusión de contenidos terroristas en línea, donde se contempla la obligación de las plataformas de adoptar medidas proactivas para prevenir su difusión, pudiendo ser sancionadas si no lo hacen.

de sus plataformas es filtrado». Se permitiría con ello que los Estados miembros, para garantizar el control efectivo de los sistemas algorítmicos, prevean que las autoridades competentes tengan la posibilidad de sancionar (art. 117) los casos en los que, por ejemplo, se probase que un algoritmo privilegia el discurso de odio. Esta posibilidad acercaría las recomendaciones algorítmicas a su consideración como decisión «editorial».

Al margen de lo establecido por el Reglamento de Servicios Digitales, tal y como se deduce de la jurisprudencia del TJUE y del TEDH, los Estados podrían reducir más el ámbito de inmunidad de las plataformas en dos sentidos. Por una parte, los Estados podrían ampliar el «deber de diligencia» exigida a las grandes plataformas. Según el TJUE, en demandas por responsabilidad civil por daños, bastaría que la plataforma «haya tenido conocimiento de hechos o circunstancias a partir de los cuales un operador económico diligente hubiera debido deducir ese carácter ilícito»<sup>9</sup>. Por tanto, se permite que los Estados puedan exigir responsabilidad a las plataformas por la no retirada de contenido sin necesidad de aviso previo<sup>10</sup>. En este margen se sitúan la *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (NetzDG)* alemana, de enero de 2018, y la *Loi n° 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet*, que contemplan la obligación de retirar contenidos «manifestamente ilegales» en un plazo perentorio, y, en caso contrario, incurrirían en responsabilidad por estos. La *NetzDG* alemana establece, en todo caso, obligaciones en el procedimiento, ya que señala que la plataforma *puede* proporcionar al usuario un derecho de respuesta antes de

<sup>9</sup> STJUE de 12 de julio de 2011, C-324/09, *L'Oréal SA et al.*, § 120; STJUE de 7 de agosto de 2018, C-521/17, *Coöperatieve Vereniging SNB-REACT U.A. y Deepak Mehta*.

<sup>10</sup> En esta misma línea, el TEDH permitió la condena penal por delito de odio por los jueces franceses a un político como «productor de un sitio web destinado a la comunicación al público en general» por no eliminar rápidamente, aún sin notificación previa, la publicación por parte de terceros de comentarios de odio contra la comunidad musulmana en el muro de su cuenta de Facebook, utilizada en el contexto de su campaña electoral (*Caso Sánchez v. Francia*, de 15 de mayo de 2023). En este caso, existe una amplia interpretación del «deber de diligencia» que pueden manejar los Estados, ya que puede aplicarse a las personas individuales como «productores de sitios web», sin necesidad de que exista notificación previa (basta probar que se era «consciente de los controvertidos comentarios» y, además, al condenarlo penalmente se hace corresponsable, junto con los autores de los comentarios, del contenido ilícito). En este caso el Tribunal también considera la mayor responsabilidad que deben asumir los políticos en campañas electorales respecto a la propagación de este tipo de discursos.

decidir sobre la ilegalidad del contenido en casos de ilegalidad no manifiesta (art. 3.3.<sup>a</sup>), así como *debe* informar al usuario de su decisión y la consiguiente justificación (art. 3, 2.<sup>a</sup> y 5.<sup>a</sup>). En cuanto a la ley francesa, la *Décision n° 2020-801 DC du 18 juin 2020* declaró que la ley violaba la libertad de expresión ante la dificultad de apreciar en tan breve plazo la ilicitud de los hechos, aunque se admitió la legitimidad constitucional de responsabilizar a las redes sociales por no retirar contenidos publicados por terceros, siempre que se hiciera con proporcionalidad.

Por otra parte, se permite que los Estados puedan exigir responsabilidad a las plataformas digitales si las plataformas se niegan a comunicar los datos personales del tercero. Se trataría de una responsabilidad *in vigilando* y, por tanto, los Estados podrían imponer en su legislación interna el deber de comunicar datos personales con objeto de garantizar la protección efectiva frente a contenidos ilícitos en el marco de un procedimiento civil (STJUE, Gran Sala, *Promusicae v. Telefónica*, de 29 de enero de 2008, C-275/06). De hecho, el TEDH ha dado la razón a los jueces nacionales que habían condenado a un sitio web por los comentarios ofensivos y anónimos alojados en este, a pesar de que el titular de dicho portal había retirado los comentarios ofensivos al solicitarlo la víctima de estos, ya que debieron poner filtros para evitar mensajes insultantes e identificar a los infractores (STEDH, Gran Sala, *Delfi AS v. Estonia*, de 16 de junio de 2015)<sup>11</sup>.

Por tanto, en Europa rige un marco general de exención de responsabilidad de las plataformas respecto a las expresiones de odio de los usuarios, si bien los Estados, tanto en el ámbito de la Convención Europea como de la UE, gozan de un amplio margen para fijar tanto una responsabilidad *in vigilando* de los operadores digitales por contenidos anónimos vertidos en la red como su «deber de diligencia» y de reparación, pudiendo así reducir el ámbito de inmunidad de estos. También, desde mi punto de vista, el Reglamento de Servicios Digitales permitiría a los Estados sancionar a las grandes plataformas por la utilización de recomendaciones algorítmicas que fomenten el discurso de odio.

---

<sup>11</sup> En la sentencia de la Sección Cuarta del Tribunal Europeo de Derechos Humanos de 2 de febrero de 2016 (caso *Magyar TARTALOMSZOLGÁLTATÓK. Egyesülete y INDEX.HU ZRT v. Hungría*), el TEDH, en un caso similar al de *Delfi v. Estonia*, considera que los tribunales húngaros en esta ocasión no han hecho una adecuada ponderación entre los derechos a la libertad de expresión y el derecho al honor, prevaleciendo el primero, ya que, aunque ofensivos y vulgares, los comentarios no constituían expresiones de carácter claramente difamatorio, y no suponían una incitación al odio o a la violencia.

## 2. EL PODER DE MODERACIÓN DE LAS GRANDES PLATAFORMAS EN EUROPA

En Europa, reconociéndose también que las redes sociales, aún de titularidad privada, se han convertido en un foro de información y debate público sin existencia de una alternativa real en su ámbito, la libertad de expresión de sus usuarios vincula a los poderes públicos, y de ahí que, según el art. 30 del Reglamento de Servicios Digitales, los poderes públicos no puedan imponer a las plataformas sistemas de filtrado o monitoraje activo de los contenidos difundidos en ellas.

Han sido las propias plataformas las que han terminado estableciendo sistemas de filtrado, incentivadas por las propias autoridades desde dos perspectivas. Por una parte, las instituciones públicas europeas tratan de limitar la incitación al odio en la red estableciendo códigos de buenas prácticas con las principales redes sociales<sup>12</sup>. Por otra parte, como hemos señalado más arriba, la jurisprudencia del TJUE y del TEDH permite que los Estados responsabilicen a las grandes plataformas por la no retirada de contenidos claramente ilícitos sin necesidad de previo aviso. Por tanto, aunque los Estados no pueden imponer sistemas de filtrado en las redes, las plataformas, para no incurrir en responsabilidad o seguir las recomendaciones de buenas prácticas, tienen que tener mecanismos de control y filtrado previo. Pero ¿son las plataformas libres a la hora de establecer ese poder de moderación en las redes o se hallan vinculadas de alguna manera por la libertad de expresión de los usuarios?

En la práctica, las grandes plataformas establecen criterios genéricos que van más allá de los límites que el ordenamiento jurídico establece respecto a la libertad de expresión por vulneración del derecho al honor de las personas o por discurso de odio hacia colectivos históricamente marginados. Además, su aplicación a través de procedimientos automatizados de inteligencia artificial elimina contenidos que, además de estar amparados por la libertad de expresión, difícilmente pueden considerarse contrarios a las propias reglas de moderación de la corporación. Todo ello supone «una restricción a un medio importante de difusión de contenidos en línea y constituyen de tal forma una

---

<sup>12</sup> Código de conducta para la lucha contra la incitación ilegal al odio en internet, de mayo de 2016, adoptado en el marco de la Estrategia para el mercado único digital de la Comisión Europea. Este código fue suscrito por Facebook, Microsoft, Twitter y YouTube. Instagram, Google+, Snapchat, Dailymotion y Jeuxvideo.com se adhirieron posteriormente. Véase [https://tinyurl.com/kr343ess\\_en](https://tinyurl.com/kr343ess_en) (consulta: 6 de octubre de 2023). En la misma línea, la Recomendación (UE) 2018/334 de la Comisión Europea, de 1 de marzo, sobre medidas para combatir eficazmente los contenidos ilícitos en línea. Disponible en: <https://tinyurl.com/mryu38fc> (consulta: 6 de octubre de 2023).

limitación del derecho garantizado en el artículo 11 de la Carta» (STJUE de 26 de abril de 2022, C-401/19, *Polonia v. Parlamento y Consejo*, § 55). El Tribunal europeo reconoce una «libertad editorial» protegida por la libertad de empresa si las plataformas no se consideran medio de comunicación o por la libertad de expresión e información si se consideran medios de comunicación, aunque matizada por su vinculación con la libertad de expresión de los usuarios. La cuestión es determinar las características que han de tener esos sistemas de filtrado y control para que la restricción a la libertad de expresión de los usuarios sea proporcionada.

El TJUE ha señalado al respecto que, si un sistema de filtrado implicara un riesgo de que no se distingan claramente contenidos lícitos e ilícitos, «de modo que su establecimiento podría dar lugar al bloqueo de comunicaciones de contenido lícito, sería incompatible con el derecho de libertad de expresión y de información» (STJUE de 26 de abril de 2022, C-401/19, *Polonia v. Parlamento y Consejo*, § 86; siguiendo la jurisprudencia sentada en STJUE de 16 de febrero de 2012, C-360/10, SABAM). Pero señala:

La normativa que conlleve la injerencia en derechos fundamentales debe establecer reglas claras y precisas que regulen el alcance y la aplicación de la medida en cuestión e impongan unas exigencias mínimas, de modo que las personas que vean limitado el ejercicio de estos derechos dispongan de garantías suficientes que permitan protegerlas de manera eficaz contra los riesgos de abuso. En particular, dicha normativa deberá indicar en qué circunstancias y con arreglo a qué requisitos puede adoptarse tal medida, garantizando así que la injerencia se limite a lo estrictamente necesario. La necesidad de disponer de tales garantías reviste especial importancia cuando la injerencia es resultado de un tratamiento automatizado (§ 67).

Concretamente, el TJUE ha subrayado la necesidad de reconocimiento de recurso frente a un bloqueo (§ 94).

En esta línea, la normativa europea establece garantías de proceso, como deberes de transparencia, motivación de decisiones y recursos extrajudiciales (Reglamento UE 2021/784, de 29 de abril de 2021 [art. 5.2.b], sobre la lucha contra la difusión de contenidos terroristas en línea; Directiva [UE] 2018/1808, de 14 de noviembre de 2018 [art. 28 ter 3.d], sobre servicios de comunicación audiovisual; Directiva 2019/790, de 17 de abril de 2019, sobre los derechos de autor [art. 17.9]).

El art. 26 del Reglamento de Servicios Digitales, en la línea establecida en la secc. 230 de la *Communications Decency Act*, reconoce la inmunidad de las plataformas, aunque ejerzan el poder de moderación. No obstante, se establecen límites a ese poder de moderación, que debe siempre ejercerse «de buena

fe y con diligencia», es decir, «de manera objetiva, no discriminatoria y proporcionada, teniendo debidamente en cuenta los derechos e intereses legítimos de todas las partes implicadas, y proporcionar las garantías necesarias contra la retirada injustificada de contenidos lícitos».

Estas garantías se traducen en obligaciones de transparencia (el deber de que cualquier plataforma tenga que informar sobre sus políticas y procedimientos de moderación de contenidos y de reclamación), de debido proceso al moderar los contenidos, estando obligadas a poner a disposición de los usuarios puntos de contacto, mecanismos de notificación y acción (art. 50), así como a motivar sus decisiones de restricción (art. 54). A las plataformas que son grandes empresas se les impone la obligación de tener sistemas internos de tramitación de reclamaciones sobre moderación de contenidos (art. 58), así como mecanismos de resolución extrajudicial de litigios (art. 59), además de la obligación de realizar un informe anual sobre la moderación de contenidos que lleven a cabo (art. 49).

Por tanto, en el ámbito de la UE, si bien se reconoce un poder de moderación de las grandes plataformas, este poder de moderación debe quedar modulado por el derecho a la libertad de expresión de los usuarios, lo que se traduce en obligaciones para las grandes plataformas de transparencia de proceso debido, incluidos mecanismos de resolución extrajudicial de litigios.

#### IV. EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS Y SU PODER DE MODERACIÓN EN ESPAÑA

##### 1. EL ODIOS EN LAS REDES Y EL «DEBER DE DILIGENCIA» DE LAS GRANDES PLATAFORMAS EN ESPAÑA

En España, dentro del marco fijado por la Directiva europea sobre el comercio electrónico y que mantiene el Reglamento de Servicios Digitales, el art. 16 de la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico (en adelante, LSSI), prevé la responsabilidad de las grandes plataformas cuando hubieran tenido «conocimiento efectivo» del carácter ilícito y no lo hubieran retirado con diligencia. Ese «conocimiento efectivo» podría ser a través de la notificación por parte de la autoridad administrativa o judicial correspondiente o «de otros medios de conocimiento efectivo que pudieran establecerse»<sup>13</sup>.

<sup>13</sup> Señala el art. 16 LSSI:

«Responsabilidad de los prestadores de servicios de alojamiento o almacenamiento de datos.

Por tanto, en los casos de expresiones de odio en la red por parte de sus usuarios, tras la ponderación entre el derecho al honor (art. 18.1 CE) y la libertad de expresión (art. 20.1 CE), la responsabilidad de las plataformas depende de las dos condiciones legalmente exigibles: el conocimiento efectivo y la diligencia en la eliminación del contenido difamatorio. Por tanto, la responsabilidad de las plataformas según la normativa no es por el contenido vertido por terceros, sino por la falta de diligencia en su eliminación. Se establece, por tanto, un «deber de diligencia» a las plataformas para retirar el contenido considerado ilícito.

Para ponderar el cumplimiento de la diligencia debida que se exige en el art. 16 LSSI habrá de tenerse en cuenta el contexto de los comentarios, las personas implicadas, las medidas aplicadas por la plataforma para impedir o eliminar los comentarios difamatorios y, en su caso, la responsabilidad de los autores reales de los comentarios (STEDH de 16 de junio de 2015, asunto *Delfi AS v. Estonia*, §§ 42-43).

Respecto a la aplicación de la exigencia del «conocimiento efectivo», existen dos tendencias jurisprudenciales. Por una parte, muchos tribunales han exigido para aplicar el «conocimiento efectivo» la comunicación de una resolución judicial que declarase difamatorio el contenido controvertido. Así, la STS (Sala de lo Contencioso-Administrativo, Sección 4.ª), 1231/2022, de 3 de octubre de 2022, señala que cuando esté en juego la libertad de expresión solo podrán decidir el bloqueo de una página web un juez o tribunal, pero no una autoridad administrativa<sup>14</sup>, con base en el art. 20.5 CE:

---

1. Los prestadores de un servicio de intermediación consistente en albergar datos proporcionados por el destinatario de este servicio no serán responsables por la información almacenada a petición del destinatario, siempre que:

- a) No tengan conocimiento efectivo de que la actividad o la información almacenada es ilícita o de que lesiona bienes o derechos de un tercero susceptibles de indemnización, o
- b) Si lo tienen, actúen con diligencia para retirar los datos o hacer imposible el acceso a ellos.

Se entenderá que el prestador de servicios tiene el conocimiento efectivo a que se refiere el párrafo a) cuando un órgano competente haya declarado la ilicitud de los datos, ordenado su retirada o que se imposibilite el acceso a los mismos, o se hubiera declarado la existencia de la lesión, y el prestador conociera la correspondiente resolución, sin perjuicio de los procedimientos de detección y retirada de contenidos que los prestadores apliquen en virtud de acuerdos voluntarios y de otros medios de conocimiento efectivo que pudieran establecerse».

<sup>14</sup> En el caso concreto, se plantea si una entidad administrativa como es la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) podía solicitar la

Que el art. 20.5 de la Constitución, seguramente como reacción a prácticas arbitrarias y abusivas del pasado, tiene por finalidad prohibir el secuestro administrativo o gubernativo de publicaciones es innegable. Así, lo constitucionalmente prohibido no es aprehender publicaciones que incurran en alguna ilegalidad ni impedir su difusión —esto tiene que ver con los límites de las libertades de información y expresión, de que se ocupan los anteriores apartados del propio art. 20 de la Constitución—, sino que lo constitucionalmente prohibido es que el secuestro lo decida por sí sola la Administración. Al igual que en otras reservas de jurisdicción previstas en el texto constitucional, el constituyente consideró preferible que ciertas decisiones especialmente delicadas para la efectividad de algunos derechos fundamentales sean tomadas por un órgano judicial. Con ello no se trata sólo de poner coto a posibles tentaciones administrativas de arbitrariedad, sino sobre todo de encomendar la valoración de los hechos y la ponderación de los intereses a una autoridad imparcial, independiente y sometida únicamente a razones jurídicas. Debe tenerse en cuenta que decidir si una publicación merece ser secuestrada —al igual que ocurre con la interceptación de comunicaciones, con el registro domiciliario o con la disolución de asociaciones— a menudo exige un razonamiento jurídicamente complejo e intelectualmente templado.

Pues bien, a la vista de todo ello, esta Sala entiende que los sitios web —aun no siendo «publicaciones» o «grabaciones» en sentido propio— entran dentro de la categoría de «otros medios de información». A través de Internet circulan públicamente noticias, datos y juicios de hecho (información), así como opiniones, posicionamientos y juicios de valor (expresión); y, en este sentido, los sitios web cumplen una función equiparable a la de los soportes tradicionales de la información y la expresión. De aquí se sigue que, en principio, el art. 20.5 de la Constitución es aplicable a la interrupción del acceso a los sitios web.

[...] A juicio de esta Sala, dicho precepto constitucional entra en juego cuando las publicaciones, las grabaciones o los otros medios de información son canales para la emisión y circulación de ideas, tanto si versan sobre hechos como si versan sobre valores (FJ 9).

Esta interpretación del art. 20.5 CE sería acorde con los arts. 8 y 11 LSSI, que señalan, respectivamente:

Art. 8.1. [...] En todos los casos en los que la Constitución y las leyes reguladoras de los respectivos derechos y libertades así lo prevean de forma excluyente,

---

interrupción de una página web sin necesidad de una previa autorización judicial. El Tribunal considera que cuando está en juego la libertad de expresión solo una autoridad judicial puede solicitar la interrupción de la página.

sólo la autoridad judicial competente podrá adoptar las medidas previstas en este artículo, en tanto garante del derecho a la libertad de expresión, del derecho de producción y creación literaria, artística, científica y técnica, la libertad de cátedra y el derecho de información.

Art. 11.3. [...] En particular, la autorización del secuestro de páginas de Internet o de su restricción cuando ésta afecte a los derechos y libertades de expresión e información y demás amparados en los términos establecidos en el artículo 20 de la Constitución solo podrá ser decidida por los órganos jurisdiccionales competentes.

En esta línea, García Morales (2013: 258-260) subraya la reserva de jurisdicción que el art. 20.5 CE impone, de manera que toda expulsión de contenidos del proceso de comunicación pública debe ser acordada judicialmente. Cotino Hueso (2015: 209) defiende la necesidad «constitucional» de una interpretación restrictiva y no expansiva de los medios de «conocimiento efectivo», de manera que solo se admita como tal si hay resoluciones administrativas o judiciales declarativas de la ilicitud del contenido. Por su parte, Villaverde Menéndez (2020: 134-135) amplía la posibilidad de hacer valer como «conocimiento efectivo» la presentación por el ofendido a la plataforma de «otras resoluciones judiciales sobre contenidos similares [...], por ejemplo, con contenidos negacionistas o xenófobos que la jurisprudencia internacional o constitucional han considerado objetivamente lesivos de la dignidad y reputación».

Por otra parte, está aquella línea jurisprudencial que adopta el sistema *notice and take down*, es decir, que basta una notificación del afectado a la plataforma sobre el posible contenido ilícito para que esta plataforma tenga la obligación de retirarlo sin necesidad de acudir a los tribunales. Es esta la línea seguida por la STS 235/2020, de 2 de junio, y la sentencia de la Audiencia Provincial 82/2018, de 5 de febrero. Estas sentencias hacen una interpretación amplia en lo referente al «conocimiento efectivo» sin que sea necesaria «en modo alguno» para deducir ese conocimiento una «previa resolución de la autoridad competente». Estas sentencias declaran la responsabilidad de una entidad prestadora de un servicio de intermediación por no haber retirado de su sitio web, pese a ser requerida dos veces para ello, un comentario de un usuario anónimo en el que se calificaba como «hijo de puta» a un concejal. En estas sentencias, tras resolver el conflicto entre el derecho al honor del concejal y la libertad de expresión del usuario anónimo en favor del primero, consideran que la plataforma debía haber eliminado el contenido tras el requerimiento del concejal por correo electrónico, por tanto, sin necesidad de sentencia previa. Se equipara el «conocimiento efectivo» con una petición de un particular formulada por *e-mail* y por burofax. La STC 83/2023, de 4 de julio, se limita

a resolver el conflicto entre el derecho al honor del concejal y la libertad de expresión del usuario anónimo en favor del primero, «[d]ado el carácter puramente vejatorio, con publicidad y anonimato, de la expresión utilizada» (FJ 6), aunque avala implícitamente la decisión adoptada en las sentencias anteriores de responsabilizar a la plataforma por no retirar el contenido a partir del requerimiento por el concejal<sup>15</sup>.

## 2. EL PODER DE MODERACIÓN DE LAS GRANDES PLATAFORMAS EN ESPAÑA

En España no hay una normativa interna que regule el poder de moderación de las plataformas de internet. Sin embargo, de la jurisprudencia del Tribunal Constitucional podemos deducir dos consideraciones: que ese poder de moderación no entraría en el concepto de «censura previa» prohibida constitucionalmente, y que el poder de moderación de las plataformas tiene como límite los derechos fundamentales de los usuarios.

Efectivamente, el poder de moderación de las grandes plataformas no puede considerarse «censura previa» y, por tanto, no entraría en la prohibición de esta establecida en el art. 20.2 CE<sup>16</sup>. En la línea marcada por la STC 187/1999, de 25 de octubre, la «censura» prohibida constitucionalmente es

[...] la intervención preventiva de los poderes públicos para prohibir o modular la publicación o emisión de mensajes escritos o audio visuales. La presión de ciudadanos o grupos de ellos para impedir esa difusión, aunque consiga obtener el mismo resultado, puede llegar a ser una intromisión en un derecho ajeno, con

---

<sup>15</sup> En esta línea se situaría Vázquez de Castro al mostrarse partidario de que en casos de clara ilicitud el prestador de servicio, en función de su capacidad de control, debería retirar el contenido, aunque no exista previo aviso (2017: 24-27).

<sup>16</sup> En contra, el voto particular que formuló el vocal don Lázaro al Acuerdo 146/2021, de 25 de febrero, de la JEC. Este voto particular entendió que la acción de Twitter de suspender durante ocho días la cuenta de Vox durante la campaña electoral en Cataluña priva de acceso al conjunto de la actividad de esa cuenta. Añade que la sanción se impone inaudita parte y sin explicitar las razones que llevan a considerar que el tuit incita expresamente a la discriminación, la xenofobia o el racismo. De haber excesos en la libertad de expresión, lo procedente es ir a un procedimiento judicial con garantía de contradicción y defensa. Considera que lo hecho por Twitter es un caso de censura previa proscrita por el art. 20 de la Constitución, «un verdadero secuestro sin resolución judicial y una decisión desproporcionada». En definitiva, que consolida el poder de censura unilateral por las redes sociales y el recurso ante la Administración electoral no soluciona esa carencia de regulación.

relevancia penal en más de un caso y desde más de un aspecto, pero no «censura» en el sentido que le da la Constitución.

Tampoco encaja en este concepto la que a veces ha dado en llamarse «autocensura», utilizada en algunos sectores —la cinematografía o la prensa—, en algunos países o en algunas épocas para regular la propia actividad y establecer corporativamente ciertos límites. Más lejos aún del concepto constitucionalmente proscrito está la carga, con su cara y reverso de derecho-deber, que permite e impone a los editores y directores un examen o análisis de texto y contenidos, antes de su difusión, para comprobar si traspasan, o no, los límites de las libertades que ejercen, con especial atención a los penales. Se trata de algo que, en mayor o menor grado, precede siempre a la conducta humana, reflexiva y consciente de que el respeto al derecho ajeno es la pieza clave de la convivencia pacífica. En tal sentido hemos dicho ya que la «verdadera censura previa» consiste en «cualesquiera medidas limitativas de la elaboración o difusión de una obra del espíritu, especialmente al hacerlas depender del previo examen oficial de su contenido» (STC 52/1983, fundamento jurídico 5.<sup>o</sup>). Por ello el derecho de veto que al director concede el art. 37 de la Ley de Prensa e Imprenta de 18 de marzo de 1966 no puede ser identificado con el concepto de censura previa (SSTC 171/1990 y 172/1990). Tampoco lo es la autodisciplina del editor, cuya función consiste en elegir el texto que se propone publicar, asumiendo así los efectos positivos o negativos, favorables o desfavorables de esa opción, como puedan ser el riesgo económico y la responsabilidad jurídica (STC 176/1995) (FJ 5).

Además, ese poder de moderación debe verse limitado por los derechos fundamentales de los usuarios, aún sin legislación específica que así lo establezca. Tal y como hemos visto, a diferencia de lo que ocurre en EE. UU., en España se reconoce la eficacia horizontal (sin necesidad de una legislación específica) de los derechos fundamentales, y, entre ellos, la libertad de expresión de los usuarios frente a las plataformas digitales.

Los derechos fundamentales son «derechos públicos subjetivos», es decir, derechos frente al Estado. De hecho, el art. 53.1 CE señala que los derechos del capítulo II del título primero «vinculan a todos los poderes públicos». Pero, al igual que se planteó en el constitucionalismo alemán en los años cincuenta y principios de los sesenta, y también en EE. UU. bajo la etiqueta de *state action doctrine*, tempranamente el Tribunal Constitucional español tuvo que dar respuesta a la vinculación que tienen los derechos fundamentales en las relaciones privadas.

Si se consideraba que la Constitución es un «ordenamiento-marco» que tiene como objeto organizar la vida político-estatal y regular la relación

Estado-ciudadano, los derechos fundamentales se conciben como derechos subjetivos de libertad dirigidos a la defensa frente al Estado y «la competencia para la concretización creadora de derecho no sería una competencia concurrente de Legislador y Poder Judicial, sino que recaería (de nuevo) en el legislador» (Böckenförde, 1993: 40-41). En línea con lo interpretado por el Tribunal Supremo de EE. UU., el TC no podría intervenir en la determinación de las posiciones jurídicas singulares, esta tarea correspondería al legislador democrático. Los derechos fundamentales desplegarían eficacia frente a particulares cuando y en la medida en que el legislador ordinario democrático así lo hubiese previsto. Entonces la eficacia de los derechos fundamentales sería mediata, ya que la inhibición del legislador, el encargado de concretar el alcance de los derechos fundamentales en las relaciones jurídicas entre particulares, no es fiscalizable en sede jurisdiccional.

Frente a esta posición, si se entiende que la Constitución «no se limita a la ordenación de la vida estatal», como señala Hesse (2011: 44), y es «la norma fundamental y fundamentadora de todo el orden jurídico» (STC 9/1981, de 31 de marzo), se excluye que el derecho constitucional y privado puedan concebirse como compartimentos estanco. En la Constitución se descubre la opción por un determinado modelo de las relaciones entre los particulares. Todas las normas de derecho privado deben, pues, interpretarse conforme a la Constitución. Los derechos fundamentales se configurarían como normas objetivas de principio que actúan en todos los ámbitos del derecho<sup>17</sup>. En este caso, el TC podría intervenir en la determinación de las posiciones jurídicas singulares.

La jurisprudencia constitucional española argumentó tempranamente a favor de la tesis de la eficacia directa de los derechos fundamentales frente a los particulares. Como señala el propio TC, «[c]iertamente, el art. 53.1 del Texto constitucional tan sólo establece de manera expresa que los derechos fundamentales vinculan a los poderes públicos, pero ello no implica una exclusión absoluta de otros posibles destinatarios» (STC 177/1988, de 10 de octubre, FJ 4). Es el art. 9.1 CE, que reconoce expresamente la sujeción de los ciudadanos a la Constitución, el artículo que daría sustento a esa eficacia directa, aunque la sujeción establecida en el art. 9.1 «se traduce en un deber de distinto signo para los ciudadanos y los poderes públicos; mientras los primeros tienen

---

<sup>17</sup> Señala Hesse que la parte de la Ley Fundamental «dedicada a los derechos fundamentales se alza un orden objetivo de valores y con ello se produce un reforzamiento principal de la eficacia de los derechos fundamentales [...] Este sistema de valores, que encuentra su punto central en el valor del ser humano y su desarrollo libre en la comunidad social, ha de tomarse como al decisión constitucional fundamental para todos los ámbitos del derecho» (2011: 157).

un deber general negativo de abstenerse de cualquier actuación que vulnere la Constitución... los titulares de los poderes públicos tienen, además, un deber general positivo de realizar sus funciones de acuerdo con la Constitución» (STC 101/1983, de 18 de noviembre, FJ 4). Además, señala el TC, en apoyo a la eficacia directa de los derechos fundamentales, la previsión legal de protección jurisdiccional civil y penal de los derechos fundamentales, ya que esta previsión resulta difícilmente compatible con una concepción de los derechos fundamentales en que estos son solo derechos frente al poder. La Sentencia 18/1984, de 7 de febrero, refiriéndose al art. 41.2 de la LOTC, declara:

[...] esta concretización de la Ley Suprema no debe interpretarse en el sentido de que sólo se sea titular de los derechos fundamentales y libertades públicas en relación con los poderes públicos, dado que en un Estado social de derecho como el que consagra el art. 1 de la CE no puede sostenerse con carácter general que el titular de tales derechos no lo sea en la vida social, tal y como evidencia la Ley 62/1978 [...], la cual prevé la vía penal (aplicable cualquiera que sea el autor de la vulneración [...]), la contenciosa-administrativa [...], y la civil, no limitada por razón del sujeto autor de la lesión.

En línea con esta argumentación, existe en nuestra jurisprudencia constitucional una tendencia al reconocimiento de la eficacia directa en ausencia de previsión legal en casos concretos<sup>18</sup>. No obstante, esta vinculación, tal y como se ha señalado, no es de la misma intensidad, y se limita a un deber negativo de respeto al derecho fundamental<sup>19</sup>, un deber que todavía queda más

---

<sup>18</sup> Así, tempranamente, en un caso de suspensión de la relación contractual del personal femenino por el hecho de contraer matrimonio, la STC 7/1983, de 14 de febrero, señaló que «la discriminación no nace de su enfrentamiento con el artículo 17 del Estatuto y sólo desde su promulgación, sino que la situación [...] es discriminatoria *ex Constitutione*, esto es, por oposición al art. 14 de la Constitución [...] de ésta y no del Estatuto de los Trabajadores nace con plenitud el derecho fundamental de las recurrentes a no ser discriminadas por razón de sexo».

<sup>19</sup> En los casos en donde se ha alegado el derecho a la libertad de religiosa en el ámbito de la empresa para cambiar los turnos de trabajo y permitir así disfrutar del descanso semanal el día prescrito por su religión, el Tribunal afirma que la trabajadora no puede pretender que «se le dispense del cumplimiento de las obligaciones que libremente aceptó y que considera ajustadas a derecho [...] llevando así [...] el principio de sujeción de todos a la Constitución (art. 9.1) a extremos inaceptables» (STC 19/1985, de 13 de febrero, FJ 1). La empresa ha cumplido, según la sentencia, el deber negativo de respeto que impone el art. 9.1 CE a los particulares. A la demandante no se le ha impedido la exteriorización de sus creencias religiosas. El no haber

modulado en las llamadas «empresas de tendencia», donde queda restringida la dimensión externa del derecho, es decir, la referida a poder expresar las ideas, sin sufrir por ello sanción o demérito<sup>20</sup>.

Esta línea jurisprudencial permitiría a los tribunales, aún sin legislación específica, aplicar el derecho fundamental de libre expresión como límite también en las relaciones privadas frente a las plataformas digitales. Los tribunales civiles podrían reconocer la eficacia del derecho a la libertad de expresión frente a las grandes plataformas a través de las cláusulas generales del derecho privado y, con ello, controlar la razonabilidad y proporcionalidad del poder de moderación<sup>21</sup>. En el ámbito de una campaña electoral, los tribunales de lo contencioso-administrativo también podrían limitar el poder de moderación de las plataformas en aplicación de los principios de pluralismo político, transparencia, objetividad e igualdad, que presiden todo proceso electoral (art. 8.1 LOREG).

---

dispensado del régimen laboral que tiene establecido con carácter general no equivale a «una actuación coercitiva impeditiva de la práctica religiosa», contraria al principio de neutralidad que debe presidir la conducta del empresario en este terreno (FJ 2).

<sup>20</sup> En la STC 47/1985, de 27 de marzo, la demandante era una profesora de EGB que prestaba sus servicios en un colegio privado, no concertado, hasta el momento de ser despedida por no ajustarse en su actividad profesional al ideario del centro. Dos eran los motivos que, según la empresa, justificaban la decisión de rescindirle el contrato: «a) Por su disconformidad con las normas de la dirección del centro, creando con ello fricciones que deterioran los criterios que presiden la enseñanza en esta institución; b) Por desarrollar su actividad profesional en forma que no se ajusta al ideario que rige nuestro centro». La profesora, que expresa su convicción de que fue despedida por haber declarado en privado a la dirección del centro no ser católica, pide que se declare la nulidad radical del despido y se reconozca su derecho a no sufrir discriminación por sus ideas o creencias íntimas (art. 16 en conexión con el art. 14 CE).

Siguiendo la STC 5/1981, la STC 47/1985 declara que una actividad docente contraria al ideario de un centro privado puede ser causa legítima de despido del profesor, siempre que resulte probado por quien alega esta causa de despido. Pero la simple disconformidad de un profesor respecto al ideario del centro no puede ser causa de despido, si no se ha exteriorizado o puesto de manifiesto en alguna de las actividades educativas del centro (FJ 3).

<sup>21</sup> Las cláusulas generales de derecho privado que permitirían ese control del poder de moderación serían la prohibición de cláusulas contrarias a la ley o abusivas que implican la renuncia de derechos del art. 6.2 del Código Civil y el art. 82 del Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios. En esta línea, Herrera de las Heras (2021: 157).

Sin embargo, sin una regulación específica sobre la materia, las decisiones judiciales no siempre han sido coherentes con la tesis de la eficacia directa de los derechos fundamentales y, en todo caso, han visto muy limitado su margen de control.

La SAP de Mallorca 124/2020, de 26 de marzo, considera que es un contrato entre privados sobre el que no se puede alegar vulneración de la libertad de expresión, en línea con la jurisprudencia en EE. UU. El Tribunal balear da la razón a Twitter al respaldar la suspensión de una cuenta en la que un usuario vertió «comentarios homófobos» de forma reiterada. Para ello se basó en que ello «en ningún caso supone que el actor se vea privado de su libertad de expresión, sino de la posibilidad de utilizar un medio concreto para ejercerlo, en este caso, una red social de titularidad privada, pero que de ninguna forma impide que pueda seguir ejercitándolo a través del mismo medio utilizado, internet, si bien por otros canales o redes sociales distintas, o de otros medios de comunicación». Señala que no hay abuso en la cláusula que permite a Twitter eliminar una cuenta por incumplir las reglas incluidas en sus condiciones generales ya que «en ningún caso crea ningún desequilibrio frente a la otra parte, ya que en el Acuerdo de Usuario se contempla igualmente la posibilidad de que el usuario pueda cancelar su acuerdo en cualquier momento cancelando sus cuentas, sin necesidad de esgrimir motivo alguno».

En cambio, la Junta Electoral Central (Sesión JEC: 25-2-2021, Acuerdo 146/2021, de 25 de febrero, Expediente 293/1215) y, posteriormente, la STS 246/2022, de 28 de febrero, se sitúan en la tesis de la eficacia directa de los derechos fundamentales. En este caso ambos órganos declararon que la decisión de Twitter de suspender durante ocho días la cuenta de Vox durante la campaña electoral en Cataluña<sup>22</sup> había sido razonable y no discriminatoria,

---

<sup>22</sup> La plataforma Twitter suspendió la cuenta de Vox durante ocho días durante la campaña de las elecciones catalanas de 2021, al considerar que la información que transmitía, que vinculaba la inmigración procedente del norte de África a la delincuencia, incitaba al odio hacia ese colectivo. Efectivamente, Vox publicó días antes de la campaña electoral el siguiente tuit referido a la población de origen magrebí que vive en la comunidad catalana: «Suponen aproximadamente un 0,2 % y son responsables del 93 % de las denuncias. La mayoría son procedentes del Magreb. Es la Cataluña que están dejando la unánime indolencia y complicidad con la delincuencia importada. ¡Solo queda Vox!#StopIslamización». No es la primera vez en la que la publicación de un tuit por Vox o un cartel electoral de su autoría es investigada por la Fiscalía por ese motivo. Los casos más conocidos son los siguientes: 1) el de la publicación de un tuit en el que señalaba que los violadores de una chica en Cullera eran todos magrebíes (cuando era sabido por los medios que todos eran españoles) y añadía: «Imprescindible puntualizar el origen extranjero de la mayoría de los violadores para

aunque tanto la Junta Electoral<sup>23</sup> como el Tribunal Supremo<sup>24</sup> alertan de la necesidad de una regulación del poder de moderación de estas plataformas.

Tal y como señala la Junta Electoral:

«La relación de TWITTER con los usuarios es de carácter privado, pero puede tener relevancia electoral, lo que justifica la intervención de la Administración electoral para velar por el respeto de los principios de pluralismo político, transparencia, objetividad e igualdad, que presiden todo proceso electoral, principios que trascienden a esa relación privada».

La Junta considera que, si bien Twitter se ha convertido en una herramienta de interacción y comunicación social con una enorme difusión y una indiscutible relevancia, su naturaleza de «canal de comunicación que permite a sus usuarios el envío e intercambio de mensajes de texto breves a grupos reducidos de seguidores o, de forma abierta, a todo el público» hace imposible su encaje en el concepto de «medios de comunicación privados» al que hace referencia el art. 66.2 de la LOREG. Dicho precepto restringe las decisiones editoriales de este tipo de medios de comunicación en período electoral en aras del necesario respecto de «los principios de proporcionalidad y neutralidad

---

que los españoles tomen conciencia del tipo de delincuentes a los que estáis abriendo las puertas y subsidiando con el dinero de todos»; 2) el del cartel contra los menores inmigrantes donde aparecía la imagen de una señora de edad avanzada frente a un joven encapuchado con un pañuelo que tapaba la cara excepto los ojos, al igual que los jóvenes que participan de forma violenta en manifestaciones causando disturbios y no quieren ser reconocidos; el cartel iba acompañado de la frase en grande «UN MENA 4700 euros al mes, TU ABUELA, 426 euros de pensión/mes. Vox protege Madrid, vota seguro». Ambas investigaciones iniciadas por la Fiscalía fueron archivadas.

<sup>23</sup> Señala la Junta Electoral que «en esta materia —actuación de las redes sociales en período electoral— hay un vacío normativo y que es consciente de los peligros y riesgos de estas decisiones de las redes sociales, que pueden limitar seriamente la campaña electoral, sin tiempo material para obtener la tutela judicial a lo que se añade que las decisiones de los responsables de esas redes se adoptan de plano, sin oír al usuario. En tanto el legislador salva este déficit normativo, la Administración electoral sólo puede actuar en casos extremos si hay vulneración grave de los principios de pluralismo político, transparencia, objetividad o igualdad entre las candidaturas electorales, lo que no es el caso».

<sup>24</sup> Señala el TS «lo insatisfactorio que, tanto en lo procedimental como en lo material, es el escaso tratamiento legal de ese poder censor que se reservan contractualmente las redes sociales, erigidas hoy día en medios de difusión masiva de todo tipo de contenidos, también políticos, con un poder e influencia manifiestos en tiempos electorales».

informativa en los debates y entrevistas electorales, así como en la información relativa a la campaña electoral».

El Tribunal Supremo avala la no aplicación del art. 66.2 de la LOREG:

Primero porque al margen de que TWITTER no sea un medio de comunicación en el sentido tradicional del término lo cierto es que, aun cuando lo fuese, es una plataforma de publicación de conversaciones en abierto que no organiza debates ni entrevistas como tampoco emite información propia, luego falla el presupuesto del citado artículo 66.2. Y la segunda razón, porque los principios llamados a ser observados por los medios de comunicación y fiscalizados por la Administración electoral conforme al artículo 66.2, cabe también deducirlos del artículo 8.1 de la LOREG.

Por tanto, los principios de pluralismo político, transparencia, objetividad e igualdad, que presiden todo proceso electoral, son principios que trascienden esa relación privada y pueden limitar el poder de moderación de las plataformas. Entran, pues, en la proporcionalidad de la limitación a la libertad de expresión, del derecho de participación política o la igualdad, y dan por hecho la vinculación directa de estos derechos sobre las plataformas digitales, aún sin regulación legal específica.

En todo caso, en la actualidad, la aprobación del Reglamentos de Servicios Digitales impone límites de transparencia y proceso debido al poder de moderación de las grandes plataformas y que deberán aplicar los tribunales. En este sentido, las plataformas deben informar al usuario con antelación al bloqueo, explicando los motivos de esta decisión y, sobre todo, dando al usuario la oportunidad de presentar una defensa que debería ser examinada por la plataforma en una nueva decisión. En caso contrario, podría declararse vulnerado el derecho a la libertad de expresión del usuario y obligar por sentencia a la plataforma a levantar el bloqueo.

## V. OPCIONES DE LEGE FERENDA DEL LEGISLADOR DEMOCRÁTICO ESPAÑOL

### 1. LA REGULACIÓN DE LA RESPONSABILIDAD DE LAS PLATAFORMAS DIGITALES RESPECTO A LAS EXPRESIONES DE ODIOS EN LAS REDES SOCIALES

La libertad de expresión de los usuarios en la red tiene los mismos límites que en cualquier otro entorno. Como el resto de derechos fundamentales, la libertad de expresión no es un derecho absoluto y encuentra sus límites en los derechos reconocidos a las demás personas, especialmente el derecho al honor o el derecho a la intimidad (art. 18 CE). La libertad de expresión no puede incluir

el derecho al insulto ni puede amparar intromisiones en la intimidad de las personas. Esos límites están recogidos en nuestra Constitución (art. 20.4) y están amparados por los tribunales, debiéndose analizar caso por caso, siguiendo los criterios fijados para su resolución por la jurisprudencia constitucional y el Tribunal Europeo de Derechos Humanos. Por tanto, las expresiones de odio en la red podrán acarrear, en virtud de la normativa vigente y su interpretación jurisprudencial, responsabilidad civil, responsabilidad administrativa (arts. 46 ss de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación) o incluso penal de los autores de estas (delito de calumnia, 206 CP; delito de injuria, 208 CP; delito de odio, 510 CP).

Tanto en EE. UU. como en Europa las grandes plataformas digitales no son consideradas editores, al entender que son «neutrales» en el proceso comunicativo, y se las hace inmunes respecto a las expresiones de odio que puedan hacer terceros en sus redes. No obstante, en la Unión Europea, de acuerdo con el Reglamento de Servicios Digitales, que sigue la senda de la Directiva sobre el comercio electrónico, las plataformas digitales tienen un «deber de diligencia» que se traduce en la obligación de retirar las expresiones de odio en la red cuando tengan «conocimiento efectivo» de su carácter ilícito. Este «deber de diligencia» está recogido en España en el art. 16 LSSI.

Tal y como hemos comprobado a lo largo del trabajo, son dos las cuestiones en materia de responsabilidad por expresiones de odio en las redes sociales que requerirían una precisión legislativa.

En primer lugar, sería imprescindible para concretar la responsabilidad de las plataformas digitales señalar si el «conocimiento efectivo» del contenido ilícito exige o no de una resolución judicial que declare difamatorio el contenido controvertido. Desde mi punto de vista, existe en virtud del art. 20.5 CE una reserva de jurisdicción para decidir sobre la retirada de contenidos o bloqueos de páginas web, sin que sea suficiente para responsabilizar a la plataforma con una resolución administrativa o una notificación de un particular pidiendo esa retirada o bloqueo. En caso contrario, las plataformas digitales se convierten en garantes de la corrección del discurso en las redes sociales. Como señala Barata Mir, «las plataformas son forzadas a convertirse en un instrumento de aplicación de normas jurídicas», de forma que «no estarán ya actuando como meros sujetos privados, sino como intérpretes y ejecutores, *ex lege*, de la delimitación nada más y nada menos que del legítimo ejercicio de un derecho fundamental» (2022: 101). Se trataría de una *censura vicarial* (Vázquez Alonso, 2022: 118-121).

Obligar a una empresa privada a identificar un contenido ilícito y a eliminarlo supone abrir una caja de pandora sobre lo políticamente correcto difícilmente compatible con el pluralismo político. Porque una cosa es que, en ejercicio de su libertad de empresa, las grandes plataformas puedan aprobar

unas normas de moderación que vayan más allá de lo penalmente castigado en el ordenamiento jurídico (y en estos casos habría que establecer con qué límites), y otra cosa es obligarlas a identificar y eliminar un contenido ilícito al margen de la autoridad judicial, cuando, ya de por sí, tras más de veinticinco años de aplicación penal del odio como límite a la libertad de expresión en España, las respuestas judiciales están lejos de ser previsibles<sup>25</sup>.

En segundo lugar, existe un espacio de impunidad en el caso de expresiones de odio en las redes realizadas bajo anonimato, ya que ni se puede responsabilizar al autor de esas expresiones ni las plataformas digitales se hacen responsables por el contenido de dichas expresiones, más allá del «deber de diligencia» de retirarlas. En estos casos de anonimato, para evitar ese espacio de impunidad, debería responsabilizarse a las grandes plataformas por el contenido de odio vertido por terceros si esa plataforma no comunica los datos del infractor. Se trataría de regular una «responsabilidad *in vigilando*» similar a la prevista en las cartas al director de la prensa escrita<sup>26</sup> y legítima en el marco de la normativa y jurisprudencia europea.

Como ya hemos señalado, si bien el derecho de la UE no obliga a los Estados miembros a imponer en su legislación interna el deber de comunicar datos personales con objeto de garantizar la protección efectiva frente a contenidos ilícitos en el marco de un procedimiento civil, admite que puede ser una opción legítima (STJUE, Gran Sala, *Promusicae v. Telefónica*, de 29 de enero de 2008, C-275/06). De hecho, el TEDH ha dado la razón a los jueces nacionales que habían condenado a un sitio web por los comentarios ofensivos y anónimos alojados en la página, ya que debieron poner filtros para evitar mensajes insultantes e identificar a los infractores. Al no hacerlo, incurrieron en responsabilidad *in vigilando* (STEDH, Gran Sala, *Delfi AS v. Estonia*, de 16 de junio de 2015) (*vid.* Barrero Ortega, 2021: 122).

La regulación de esta «responsabilidad *in vigilando*» exigiría, además, distinguir entre los operadores que son meros intermediarios<sup>27</sup> y aquellas

---

<sup>25</sup> Recientemente, sobre la imprecisión legal de los delitos de odio y los criterios jurisprudenciales sobre su aplicación, *vid.* Lucía Morlans, 2024.

<sup>26</sup> Esta responsabilidad de la prensa escrita por el contenido difamatorio vertido en una carta al director por no comunicar el autor de esta fue establecida en la STC 336/1993, de 15 de noviembre. En favor del reconocimiento de esta responsabilidad para los operadores de internet que utilizan algoritmos, se muestran los votos particulares de los magistrados doña María Luisa Balaguer Callejón, don Ramón Sáez Valcárcel y doña Laura Díez Bueso a la STC 83/2023, de 4 de julio.

<sup>27</sup> Entre ellos podríamos incluir: 1) los operadores de redes que proveen la infraestructura para el acceso a la transmisión de internet entre un punto y otro; 2) los

plataformas que sí seleccionan, a través de aplicaciones algorítmicas, el contenido publicado<sup>28</sup>, siendo aplicable esta responsabilidad solo a las segundas.

## 2. LA REGULACIÓN DEL PODER DE MODERACIÓN DE LAS PLATAFORMAS DIGITALES

Las redes sociales, convertidas en medios de difusión masiva y en un campo abonado para la proliferación de expresiones de odio hacia personas o grupos de personas, se han reservado contractualmente un poder de moderación del discurso que puede afectar al ejercicio del derecho a la libertad de expresión de los internautas. En EE. UU., esta libertad de moderación, entendida bajo el amparo de la primera enmienda, ha sido absoluta, aunque se empieza a abrir una línea jurisprudencial que entiende que su libertad de moderación se basa estrictamente en lo reconocido por la legislación federal y no en la primera enmienda y que podría ser modulada por el legislador estatal. En el ámbito de la UE, tanto normativa como jurisprudencialmente se reconoce que el poder de moderación de las grandes plataformas debe quedar modulado por el derecho a la libertad de expresión de los usuarios, lo que se traduce en obligaciones para las grandes plataformas de transparencia y de proceso debido, incluidos mecanismos de resolución extrajudicial de litigios.

De cara a futuro, al margen de los límites de transparencia y procedimentales exigidos desde la UE, sería conveniente reconocer legislativamente límites sustantivos a la libertad de moderación de las grandes plataformas. En línea con la normativa europea, más estricta con las plataformas y motores de búsqueda en línea de gran tamaño —con al menos cuarenta y cinco millones de usuarios—, el legislador español podría distinguir entre las redes sociales de las que, por su diseño tecnológico y reglas de acceso, sea fácil deducir que se trata de un foro de naturaleza restringido, y empresas como Facebook o Twitter. Las redes sociales de naturaleza restringida podrían tener interés

---

proveedores de acceso a internet, incluidas compañías como Une, T-Mobile, Orange, etc., y 3) los proveedores de servicios de alojamiento de la información o *hosting* que permiten que los usuarios almacenen o guarden la información que deseen en sus servidores. En esta categoría, podemos encontrar los servicios en la nube, como YouTube, DropBox, SkyDrive, etc. *Vid.* Barrero Ortega, 2021: 109.

<sup>28</sup> Aquí se incluirían: 1) los portales de internet con enlaces de noticias que seleccionan publicaciones y las clasifican; 2) los proveedores de acceso logístico o motores de búsqueda como Google, Yahoo!, etc., y 3) los proveedores de servicios de redes sociales como Facebook, Twitter, LinkedIn, etc. *Vid.* Barrero Ortega, 2021: 109-110.

legítimo en excluir un discurso protegido por la libertad de expresión por consideraciones éticas o por consideraciones económicas y el legislador les permitiría excluir discursos protegidos por la libertad de expresión yendo más allá de lo prohibido penalmente respecto al honor de otras personas y a la dignidad de colectivos. Sin embargo, las empresas como Facebook o Twitter deberían someterse a los límites del derecho penal vigente «dada la naturaleza cuasi pública de los foros que han sido creados» por estas empresas y «de su propia posición en el mercado»<sup>29</sup>. Como argumenta Vázquez Alonso, no establecer límites sustanciales al poder de moderación de las grandes plataformas convierte a «dichas corporaciones sean usadas por el Estado como entes vicariales para que censuren en sus foros aquello que, de forma general, no podría ser vetado por ningún poder público», y denuncia que «en el ámbito de la UE las grandes corporaciones de Internet ya actúan en ciertos ámbitos como censores vicariales del Estado, al asumir dentro de sus criterios de autorregulación ciertas recomendaciones europeas» (2022: 117-119).

En definitiva, resulta imprescindible en este campo la intervención del legislador democrático precisando los límites del poder de moderación de las grandes plataformas porque las redes sociales, aún de titularidad privada, son foros públicos de discusión donde están en juego principios básicos de nuestro sistema democrático.

### Bibliografía

- Barata Mir, J. (2021). Un debate ponderado sobre leyes, redes sociales y moderación de contenidos. *Agenda Pública*, 10-10-2021. Disponible en: <https://tinyurl.com/u5xu253v>.
- Barata Mir, J. (2022). Libertad de expresión, regulación y moderación privada de contenidos. *Teoría y Derecho*, 32, 89-107.
- Barrero Ortega, A. (2021). Responsabilidad de los intermediarios de Internet en el Derecho de la UE. *Revista Española de Derecho Constitucional*, 123, 107-132. Disponible en: <https://doi.org/10.18042/cepc/redc.123.04>.
- Böckenförde, E. W. (1993). *Escritos sobre derechos fundamentales*. Baden-Baden: Nomos.
- Cetina Presuel, R. (2021). Un estira y afloja: la definición de las reglas para la libre expresión en las plataformas de redes sociales. *Jurídicas CUC*, 17 (1), 499-556. Disponible en: <https://doi.org/10.17981/juridcuc.17.1.2021.18>.
- Cotino Hueso, L. (2015). Libertades informativas y responsabilidad de los prestadores de servicios en Internet. En R. Bustos, M. Fernández y E. Fossas (dirs.). *La protección*

<sup>29</sup> En esta línea Vázquez Alonso, 2022: 117-119. En cambio, Teruel Lozano solo admite una limitación de su libertad editorial desde el punto de vista procedimental, exigiendo procesos transparentes, ya que los considera «foros privados» (2023: 212).

- jurisdiccional de los derechos. Actas del XI Congreso de la Asociación de Constitucionalistas de España* (pp. 171-230). Valencia: Tirant lo Blanch.
- García Morales, M. J. (2013). La prohibición de la censura en la era digital. *Teoría y Realidad Constitucional*, 31, 237-276. Disponible en: <https://doi.org/10.5944/trc.31.2013.10308>.
- Herrera de las Heras, R. (2021). La suspensión de cuentas de políticos por Twitter y los bloqueos de usuarios por parte de las Administraciones y cargos públicos. *Revista de Derecho, Empresa y Sociedad (REDS)*, 18, 151-167.
- Hesse, K. (2011). *Escritos de derecho constitucional*. Madrid: Fundación Coloquio Jurídico Europeo y Centro de Estudios Políticos y Constitucionales. Disponible en: <https://doi.org/10.69592/978-84-1194-435-9>.
- Lucía Morlans, J. (2024). *Líneas jurisprudenciales e instrumentos internacionales claves para el análisis del odio como límite a la libertad de expresión*. Madrid: Colex.
- Sánchez Muñoz, O. (2020). *La regulación de las campañas electorales en la era digital*. Madrid: Centro de Estudios Políticos y Constitucionales.
- Teruel Lozano, G. M. (2023). Libertad de expresión, censura y pluralismo en las redes sociales: algoritmos y el nuevo paradigma regulatorio europeo. En F. Balaguer y L. Cotino (dirs.). *Derecho público de la inteligencia artificial* (pp. 181-222). Zaragoza: Fundación Manuel Giménez Abad.
- Thompson, M. (2016). Beyond Gatekeeping: The Normative Responsibility of Internet Intermediaries. *The Vanderbilt Journal of Entertainment & Technology*, 4, 783-848.
- Vázquez Alonso, V. J. (2022). La censura «privada» de las grandes corporaciones digitales y el nuevo sistema de la libertad de expresión. *Teoría y Derecho*, 32, 109-129.
- Vázquez de Castro, E. (2017). Derecho al honor en internet, anonimato y responsabilidad de los prestadores de servicios: ¿evolución e involución? *Revista Aranzadi de Derecho y Nuevas Tecnologías*, 44, 1-44.
- Villaverde Menéndez, I. (2020). *Los poderes salvajes. Ciberespacio y responsabilidad por contenidos difamatorios*. Madrid: Marcial Pons.