



Juridisk Publikation

STOCKHOLM - UPPSALA - LUND - GÖTEBORG - UMEÅ

Fredrik Thorslund

Nya medier, nya remedier

– innehållsmoderering och yttrandefrihet på
socialamedieplattformar i EU och USA

Särtryck ur häfte 1/2023

Nummer 1/2023

NYA MEDIER, NYA REMEDIER

– INNEHÅLLSMODERERING OCH YTTRANDEFRIHET PÅ SOCIALA MEDIEPLATTFORMAR I EU OCH USA

Av Fredrik Thorslund¹

Har de sociala medieföretagen ett ansvar att upprätthålla användarnas yttrande- och informationsfrihet? Frågeställningen fick fäste i den allmänna debatten förhållandevis sent. Det var först kring decennieskiftet 2020 – främst i samband med det kontroversiella beslutet att stänga av USA:s president Donald Trump från Twitter – som frågan blev att betrakta som verkligt politiskt allmängods. Sedan dess har både det juridiska landskapet och de sociala medieföretagens förhållningssätt till yttrandefrihetsrättslig normgivning genomgått drastiska förändringar. Men vissa fundamentala problem kvarstår än.

Den här artikeln behandlar huvudlinjerna i debatten om yttrandefriheten och de sociala medieföretagens innehållsmoderering, ur ett amerikanskt och europeiskt perspektiv. I artikeln diskuteras även den rättsutveckling som skett på området under senare år – och de utmaningar som återstår att hantera i takt med en ökad statlig reglering av innehållsmodereringen.

Under 2010-talet genomgick den sociala medieindustrin en expansionsfas som saknar motstycke i modern historia. Från att ha varit en virtuell bildmatrikel för Harvardstudenter kom Facebook att värva fler aktiva medlemmar än världens största stat har invånare.² Mot slutet av decenniet skickades mer än 5 000 tweets per sekund, och Youtube-användare i 91 länder runt om i världen strömmade mer än en miljard timmar video varje dag. Fyra av fem svenskar använde sociala medier: tre av fem dagligen.³ Inför riksdagsvalet 2018 uppgav knappt var tredje förstagångsväljare att sociala mediekanalet som Facebook, Twitter och Youtube var viktiga eller mycket viktiga informationskällor.⁴

¹ Biträdande jurist på TIME DANOWSKY Advokatbyrå, skribent och redaktionsledamot för idétidskriften Liberal Debatt.

² Kina, som åtminstone då utgjorde världens största stat, hade omkring 1,4 miljarder invånare 2021. Antalet aktiva användare på Facebook vid samma tidpunkt var dubbelt så stort, omkring 2,8 miljarder.

³ Internetstiftelsen, *Svenskarna och internet 2019, 2019–10–15* (2023-04-06), <https://svenskar-naochinternet.se/rapporter/svenskarna-och-internet-2019/>.

⁴ Internetstiftelsen, *Svenskarna och internet: Valspecial 2018, 2018-06-25* (2023-04-23), <https://internetstiftelsen.se/kunskap/rapporter-och-guider/svenskarna-och-internet-valspecial-2018/>. Inför valet 2022 var den siffran drygt sex av tio, se Internetstiftelsen, *Svenskarna och*

Samtidigt var det förhållandevis få som diskuterade de sociala medieföretagens ansvar för yttrande- och informationsfriheten på deras respektive plattformar. I akademiska kretsar fanns förvisso de som frågade sig om inte de sociala medieplattformarna borde liknas vid nutidens *agora* – de grekiska torg som en gång blev demokratins vagg – och om deras förvaltare följaktligen borde bära ett särskilt ansvar för att upprätthålla den fria åsiktsbildningen i samhället.⁵ Den bredare debatten befann sig däremot fortfarande i ett slags limbo mellan 90-talets optimistiska *Cyberpunk*-mentalitet⁶ och den *backlash* som skulle följa med 2020-talet.⁷

Det var först när USA:s president Donald Trump stängdes av från Twitter den 8 januari 2021 som den allmänna debatten tog fart på allvar. Två dagar efter stormningen av det amerikanska kongresshuset avskärmades en av världens mäktigaste befattningshavare från sin favoritkanal på grund av upprepade brott mot plattformens förbud mot ”glorifiering av våld”.⁸ Samtidigt rasade en diskussion om de sociala medieföretagens ansvar för desinformation kring covid-19 och den pågående vaccinationskampanjen.⁹ Twitters vd Jack Dorsey och Facebooks vd Mark Zuckerberg kallades båda upp inför den amerikanska kongressen för att ställas till svars för hur användargenererat innehåll sprids på deras plattformar.¹⁰ I december 2022 blossade debatten upp

internet: *Valspecial* 2022, 2023-04-05 (2023-04-06), <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-valspecial-2022/>.

⁵ Se till exempel Everett, C, *Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media*, *Kansas Journal of Law & Public Policy* 2018 (28:1) s 113–145 och Langvardt, K, *Regulating Online Content Moderation*, *Georgetown Law Journal* 2018 (106:5) s 1353–1358.

⁶ *Cyberpunk* var en rörelse som förespråkade frihet, personlig integritet och frånvaro av statlig inblandning på det då unga internet. Rörelsens etos sammanfattades av lyrikern och aktivisten John Perry Barlow i en skrift publicerad under namnet *A Declaration of Independence of Cyberspace* 1996: ”Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.”

⁷ Denna *backlash* är i dag så vedertagen att den fått ett eget uttryck: *techlash*. IDG beskriver *techlash* som en ”reaktion mot de stora amerikanska it-företagens dominans”, i synnerhet vad gäller ”företagens styrning av nyhetsflödet, deras kontroll av användarnas information, företagskulturens avarter och bristen på alternativ”. 2018 upptogs uttrycket i Språkrådets nyordslista. Se IDG, *Techlash*. I IDG:s ordlista, u å (2023-03-12), <https://it-ord.idg.se/ord/techlash/>.

⁸ Se Twitter, *Permanent suspension of @realDonaldTrump*, Twitter Blog, 2021-01-08 (2023-04-06), https://blog.twitter.com/en_us/topics/company/2020/suspension.

⁹ Se Niedzwiadek, N, *Psaki excoriates criticism of Biden administration work on vaccine misinformation*, *Politico*, 2021-07-16 (2023-04-06), <https://www.politico.com/news/2021/07/16/psaki-vaccine-misinformation-499850> och Coleman, A, *Covid-19: Facebook to take down false vaccine claims*, *BBC News*, 2020-12-03 (2023-04-06), <https://www.bbc.com/news/technology-55175222>.

¹⁰ Se De Vynck, G (m fl), *Big tech CEOs face lawmakers in House hearing on social media's role in extremism, misinformation*, *The Washington Post*, 2021-04-09 (2023-04-06), <https://www.washingtonpost.com/technology/2021/03/25/facebook-google-twitter-house-hearing-live-updates/>.

igen, när den nyttillträdde Twitter-ägaren Elon Musk plötsligt valde att stänga av ett antal användarkonton tillhörandes journalister som påstods ha delat privat information om Musks förehavanden.¹¹ Åtminstone i USA har uppfattningarna kring plattformarnas modereringsansvar tenderat att gå isär ungefär längs med den partipolitiska mittfåran. Demokratiska kongressledamöter har vänt sig emot plattformarnas påstådda slapphänthet mot desinformation och hatiskt innehåll. Republikanska ledamöter har i stället klandrat plattformarna för politiskt betingad övermoderering.¹²

Från ett juridiskt perspektiv finns två dominerande huvudlinjer i den efterföljande debatten. Den ena argumentationslinjen utgår från att de sociala medieföretagen har – eller åtminstone borde ha – en skyldighet att tillgodose allmänhetens rätt till yttrande- och informationsfrihet, eftersom de fungerar som ”grindvakter” i det moderna informationssamhället. Den andra argumentationslinjen fäster i stället större vikt vid den traditionella synen på fri- och rättigheter som någonting som primärt kan åberopas gentemot staten, inte mot privata företag som Twitter eller Youtube. Båda argumentationslinjerna har uppenbara styrkor, men också – som framgår nedan – mer eller mindre uppenbara svagheter.

I. GRINDVAKTSARGUMENTET: MED MAKT FÖLJER ANSVAR

Sätt, o HERRE, en vakt för min mun, bevaka mina läppars dörr.

Psaltaren 141:3

Begreppet ”grindvakt” härstammar inte ursprungligen från rättsvetenskapen eller rättspolitikerna, utan från den tidiga socialpsykologin. Uttrycket myntades av den amerikanske psykologen Kurt Lewin i en forskningsartikel från 1943, för att illustrera hur amerikanska hemmafruar utövade det bestämmande inflytandet över befolkningens matvanor.

Mat rör sig nämligen inte av egen kraft. Som Lewin påpekade slussas den i stället genom olika ”kanaler” på sin väg från producent till slutmottagare: från fabriken till affären, från affären till skafferiet, och från skafferiet till middagsbordet. Den som råder över de olika kanalerna, menade Lewin, har också kontroll över vad som når fram till konsumenten. I Lewins fall var det hemmafrun. Hon handlade

¹¹ Se Dang, S (m fl), *Twitter suspends several journalists, Musk cites ‘doxxing’ of his jet*, Reuters, 2022-12-17 (2023-04-08), <https://www.reuters.com/technology/twitter-suspends-accounts-several-journalists-2022-12-16/>.

¹² Jfr De Vyck, G (m fl) (2021).

maten, förvarade den, lagade den och serverade den. Hon besatt en särskild förmedlingsmakt: hon var en grindvakt i konsumtionskedjan.¹³

Senare spreds begreppet till andra vetenskapliga discipliner – bland annat till informationsvetenskapen, där ”grindvakt” blev en beteckning för de aktörer som kontrollerar flödet i en viss informationskedja. Tidigare handlade det främst om redaktörer och utgivare. I dag handlar det oftast om en alltmer komplex serie av infrastrukturella mellanhänder som överför, lagrar och indexerar digital information, såsom internetleverantörer, DNS-register, hosting-tjänster, sökmotorer och plattformslieferantörer.¹⁴

På senare tid har grindvaktbegreppet också anammats i lagstiftningssammanhang. I november 2022 trädde Europaparlamentets och rådets förordning (EU) 2022/1925 om öppna och rättvisa marknader inom den digitala sektorn (*Digital Markets Act* eller *DMA*) i kraft. Förordningen innehåller nya konkurrensrättsligt präglade regler som syftar till att förhindra otillbörligt agerande av plattformslieferantörer med en betydande inverkan på den inre marknaden. I förordningen benämns sådana plattformslieferantörer som just ”grindvakter”.¹⁵ De som pekas ut som grindvakter enligt förordningen kan bland annat bli skyldiga att ge företagsanvändare tillgång till data som genererats vid användning av grindvaktens plattformstjänst, och att inte diskriminera företagsanvändares produkter och tjänster till förmån för sina egna (genom att till exempel rangordna de egna produkterna och tjänsterna högre i sökresultatslistor).¹⁶ Plattformslieferantörer med stor förmedlingsmakt åläggs därigenom vissa särskilda skyldigheter gentemot andra privaträttsliga subjekt. I det här fallet handlar det om skyldigheter som syftar till att upprätthålla en rättvis konkurrens på marknaden, främja innovationskraft och minska risken för marknadsfragmentering.¹⁷

En del har argumenterat för att de sociala medieföretagens ställning som grindvakter i informationssamhället på motsvarande sätt borde föranleda ett särskilt ansvar att tillvarata användarnas fri- och rättigheter – särskilt i fråga om

¹³ Se Lewin, K, *Forces Behind Food Habits and Methods of Change. I: The Problem of Changing Food Habits: Report of the Committee on Food Habits 1941–1943*, National Research Council 1943, s 35–65.

¹⁴ Se Laidlaw, E, *A framework for identifying Internet information gatekeepers*, *International Review of Law, Computers & Technology* 2010 (24:3) s 263–276 och Zittrain, J, *A History of Online Gatekeeping*, *Harvard Journal of Law & Technology* 2006 (19:2) s 253–298.

¹⁵ För fullständig definition, se artikel 3 i förordningen.

¹⁶ Se artikel 5–7 i förordningen.

¹⁷ Se skäl 2 ff i förordningen.

användarnas yttrandefrihet.¹⁸ Till stöd för argumentet har förts fram att plattformarna i realiteten fyller samma funktion som en offentlig plats,¹⁹ med den enda skillnaden att plattformen förvaltas av ett privaträttsligt organ snarare än ett statligt. Det har också konstaterats att det finns gott om exempel på andra situationer där privaträttsliga subjekt ålagts ett särskilt ansvar att upprätthålla allmänhetens yttrande- och informationsfrihet.

I *common law*-länder som USA och Storbritannien har tillhandahållare av allmänna kommunikationsnät och infrastruktur ofta klassificerats som *common carriers*.²⁰ Tidigare har beteckningen avsett bland annat järnvägsbolag och telegrafoperatörer, men på senare tid har också internetleverantörer kommit att omfattas av klassificeringen.²¹ Karaktäristiskt för *common carriers* är att de lyder under en särskild skyldighet att opartiskt och utan diskriminering leverera varor eller meddelanden som lagligen anförtröts dem för transport.²² Internetleverantörer har därigenom ansetts lyda under en ”neutralitetsprincip” – en skyldighet att inte intervensera i informationsflödet som de kontrollerar.²³ Ett bärande skäl bakom denna neutralitetsprincip har varit att främja ett öppet internet, och därigenom medborgarnas yttrande- och informationsfrihet.²⁴

¹⁸ Se ovan not 4.

¹⁹ Ordvalet kan diskuteras. Med ”offentlig plats” avses inte bara en plats som är öppen och tillgänglig för envar, utan en *offentlighet* – en bärare av den allmänna opinionen och ett medium för fri debatt, åsiktsbildning och demokratiskt självbestämmande. Jämför det engelskspråkiga konceptet *public sphere*: ”the intermediary realm that supports the communication of opinions, the discovery of problems that need to be dealt with collectively, the channeling of these problems through the filter of the media and political institutions, and the realization of the collective will of the people in the act of democratic self-government”. Se Trezn, H-J, *Public Sphere*. I: Gianpietro Mazzoleni (red), *International Encyclopedia of Political Communication*, 1 uppl, Wiley 2015.

²⁰ Se Speta, J B, *A Common Carrier Approach to Internet Interconnection*, *Federal Communications Law Journal* 2002 (54:2) s 225–279.

²¹ Under senare år har flera omklassificeringar skett avseende internetleverantörers ställning som *common carriers*. Diskussionerna har varit turbulenta och präglade av starka politiska motsättningar. Sedan 2017 anses internetleverantörer emellertid inte längre skyldiga att iakta särskilt *common carrier*-ansvar. Se vidare *Verizon Communications Inc. v Federal Communications Commission*, 740 F.3d 623 (D.C. Cir. 2014), *Federal Communications Committee, In the Matter of Protecting and Promoting the Open Internet – REPORT AND ORDER ON REMAND, DECLARATORY RULING, AND ORDER*, GN Docket No 14–28, 2015-02-26 (p 39 och 59) samt *Federal Communications Committee, In the Matter of Restoring Internet Freedom - DECLARATORY RULING, REPORT AND ORDER, AND ORDER*, WC Docket No 17–108, 2017-12-14 (p 1–4).

²² Se Marsden, C, *Network neutrality: from policy to law to regulation*, Manchester University Press 2017, s 17.

²³ För ytterligare diskussion kring konceptet ”nätneutralitet”, se Wu, T, *Network Neutrality, Broadband Discrimination*, *Journal of Telecommunications and High Technology Law* 2003 (2) s 141–176.

²⁴ *Federal Communications Committee, In the Matter of Preserving the Open Internet Broadband Industry Practices – REPORT AND ORDER*, GN Docket No 09–191, 2010-12-21 (p 4, 23, 50, 53,

En motsvarande neutralitetsprincip finns också inom EU-rätten. Likt sin amerikanska förlaga innebär den att internetleverantörer ”ska behandla all trafik likvärdigt, utan diskriminering, begränsningar eller störanden [sic], och oberoende av sändare och mottagare, det innehåll användarna tar del av eller distribuerar [...]”.²⁵ Också här har yttrande- och informationsfriheten varit ett tungt vägande skäl bakom regleringen.²⁶

Även i svensk nationell rätt har liknande principer etablerats. Vid stiftandet av 1949 års tryckfrihetsförordning infördes en bestämmelse i grundlagen som innebar att tillståndspliktiga postbefordringsföretag och andra allmänna kommunikationstjänster, såväl statliga som privata, inte fick förhindra spridningen av en viss skrift på grund av skriftens innehåll. En motsvarande ”distributionsplikt” infördes senare i yttrandefrihetsgrundlagen för tekniska upptagningar. Distributionsplikten infördes för att förhindra att även icke-statliga aktörer, på den mycket konkurrenssvaga post- och telemarknaden, skulle få för stora möjligheter att stävja informationsdistributionen i samhället.²⁷

Den nu aktuella diskussionen om plattformsföretagens förmedlingsmakt kan ses som en förlängning av samma grundprincip: med makt följer ansvar. Flera framstående amerikanska jurister – däribland justitierådet Clarence Thomas – har föreslagit att åtminstone de större sociala medieföretagen borde behandlas som en ny sorts *common carriers*, med skyldighet att iakttä en icke-diskrimineringsplikt gentemot sina användare och de yttranden de ger uttryck för på plattformen.²⁸ Andra har gått längre än så. FN:s särskilde rapportör David Kaye har bland annat föreslagit att plattformsföretagen – i sin egenskap av grundvakter, och med sin stora inverkan på ”den offentliga arenan” – bör ansvara för att inkorporera FN:s standarder för mänskliga rättigheter direkt i sin verksamhet. Plattformsföretagen, skriver Kaye, borde ”erkänna att den auktoritativa globala standarden för att säkerställa yttrandefrihet på deras plattformar återfinns i människorättslagstiftningen” och ”omvärdera sina innehållsstandarder i enlighet därmed”.²⁹

75–78 och 94–95) samt Federal Communications Committee (2015) (p 1, 77, 545 och 554–555).

²⁵ Se Europaparlamentets och rådets förordning (EU) 2015/2120 om åtgärder rörande en öppen internetanslutning (TSM-förordningen), artikel 3(3).

²⁶ Se skäl 1, 3, 6 och 33 i TSM-förordningen.

²⁷ Tryckfrihetsförordningen 6 kap 4 §, yttrandefrihetsgrundlagen 3 kap 15 § och SOU 2012:55 s 339–353. Distributionsplikten omfattar dock inte internetleverantörer.

²⁸ Se *Biden v Knight First Amendment Institute at Columbia University*, 141 S. Ct. 1220 (2021) (Thomas, J. concurring).

²⁹ Se FN:s generalförsamling, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/HRC/38/35, 2018-04-06, p 70–72.

2. MOTARGUMENTET: SOCIALA MEDIER ÄR INTE MYNDIGHETER

I feel fundamentally uncomfortable sitting here in California at an office, making content policy decisions for people around the world. I have to, because [I lead Facebook], but I'd rather not.

Mark Zuckerberg, grundare av Facebook³⁰

Den andra huvudlinjen i diskussionen har motsatt utgångspunkt. Som tidigare nämnt utgår resonemanget från en mer traditionell syn på de individuella fri- och rättigheterna som rättsregler med företrädesvis vertikal verkan. Yttrandefriheten, lyder resonemanget, är en negativ rättighet som i princip inte innebär mer än ett skydd mot statliga inskränkningar i den fria åsiktsbildningen. Plattformleverantörer som Twitter och Youtube är privata företag med egna rättigheter. De har rätt att utforma och moderera sina plattformstjänster bäst de vill. Om en plattformstjänst sköter sin uppgift illa, diskriminerar användare och undanhåller intressanta perspektiv och yttranden, så kommer användarna sannolikt att överge tjänsten och ansluta sig till en annan plattform. En acceptabel avvägning mellan yttrande- och informationsfriheten och det ofta motstående intresset av en välfungerande tjänst – som inte exponerar användarna för våldsuppvigling, hämndpornografi eller annat obehagligt innehåll – kommer att uppstå i fri konkurrens mellan marknadens aktörer.³¹

Resonemanget har uppenbara styrkor. Det förhåller sig närmare till ekonomiska och juridiska realiteter, och bygger i mindre utsträckning på rättsliga analogier. Det är dessutom ett korrekt och relevant påpekande att även plattformslieferantörer är bärare av rättigheter. Att tvinga plattformslieferantörerna att göra avkall på sina rättigheter för att bättre tillgodose någon annan är – utöver de

³⁰ Se Swisher, K & Wagner, K, *Here's the transcript of Recode's interview with Facebook CEO Mark Zuckerberg about the Cambridge Analytica controversy and more*, Vox, 2018-03-22 (2023-04-24), <https://www.vox.com/2018/3/22/17150814/transcript-interview-facebook-mark-zuckerberg-cambridge-analytica-controversy>.

³¹ Filosofen och debattören Sam Harris, som får räknas som en av förgrundsgestalterna för denna argumentationslinje, har formulerat argumentet som följande: "The fact is that no one has a constitutional right to be on Twitter. [...] Twitter simply isn't the public square. It is a private platform. [...] These platforms need effective moderation and standards of civility to function. They are businesses, started by entrepreneurs, supported by investors who want to make money. They have employees with mortgages. They have to survive on ad revenue, or subscriptions, or some combination of the two." Harris, S, *Making Sense with Sam Harris, #304 – Why I left Twitter, Making Sense Podcast, 2022-11-28 (2023-04-10)*, <https://www.samharris.org/podcasts> (ca 00:09:26–00:11:04).

praktiska problemen det hade inneburit för incitamenten att verka på marknaden – principiellt tvivelaktigt.³²

Det finns två huvudsakliga problem med resonemanget, som båda handlar om giltigheten i resonemangets premisser. Det ena är frågan om yttrandefriheten faktiskt utgör en uteslutande negativ rättighet. Det andra är om det verkligen stämmer att plattformslieferantörerna själva råder över innehållsmodereringen på sina plattformar.

2.1 ÄR YTTRANDEFRIHETEN EN UTESLUTANDE NEGATIV RÄTTIGHET?

Både Europeiska konventionen om skydd för de mänskliga rättigheterna och de grundläggande friheterna (Europakonventionen) och rättighetskatalogen i regeringsformens andra kapitel har traditionellt sett tolkats som rättsinstrument med normativ verkan främst gentemot det allmänna. Yttrandefriheten, så som den kommer till uttryck i rättighetskatalogerna, gäller i princip i förhållande till staten. Frågan om yttrandefriheten också skulle kunna få verkningar direkt mellan enskilda parter (så kallad *Drittwirkung*) har varit desto mer omdiskuterad. Åtminstone i svensk praxis tycks det hitintills ha funnits en tydlig motvilja mot att utsträcka rättighetskatalogernas verkan på ett sådant sätt.³³

Däremot får det numera anses stå klart att yttrandefriheten under vissa förutsättningar kan få indirekta verkningar i relationer mellan enskilda. Att så är fallet framgår inte direkt av vare sig Europakonventionen eller regeringsformen. Det är en princip som huvudsakligen utvecklats genom Europadomstolens praxis under senare decennier.

I målet *Appleby and Others v the United Kingdom*³⁴ från 2003 hade klagandeparterna nekats tillstånd att demonstrera och dela ut politiska flygblad utanför stadens stora köpcentrum. Enligt klagandena hade den brittiska staten brustit i sin skyldighet att försvara deras yttrandefrihet enligt artikel 10 i Europakonventionen, genom att inte ålägga köpcentrumets ägare att låta dem demonstrera där. Europadomstolen konstaterade å sin sida att köpcentrumet var privatägt och att staten inte kunde tillskrivas något direkt ansvar för köpcentrumets beslut. Efter visst övervägande bedömde domstolen att det också funnits andra sätt för klagandena att sprida sina budskap till stadens invånare, och att klagandena således inte varit förhindrade att utöva sin yttrandefrihet.

³² Se Barak, A, *Constitutional Human Rights and Private Law*, *Review of Constitutional Studies* 1996 (3:2) s 220–280 (särskilt s 230 f).

³³ Se NJA 2020 s 293 ("Mobilfilmen"), särskilt p 24 och däri angiven praxis.

³⁴ *Appleby and Others v the United Kingdom*, no 44306/98, dom av den 6 maj 2003.

Europadomstolen fann att det inte skett någon överträdelse av klagandenas rätt enligt artikel 10.

Samtidigt konstaterade domstolen att statligt ansvar inte utan vidare kan uteslutas, för det fall att en enskilds egendomsskydd får till följd att en annan enskild helt förhindras från att utöva sin yttrandefrihet. Under sådana omständigheter kan staten likväl anses skyldig att ingripa i egendomsskyddet för att skydda utövandet av den konventionsskyddade yttrandefriheten. Som exempel nämnde domstolen det fall att en hel kommun kontrolleras av ett privaträttsligt subjekt. Exemplet var, som snart ska framgå, inte helt gripet ur luften.

Fem år senare fälldes Sverige i Europadomstolen för att ha brutit i sina förpliktelser att skydda och främja yttrandefriheten enligt artikel 10 Europakonventionen. I målet *Khurshid Mustafa and Tarzibachi v Sweden*³⁵ hade två boende i ett hyresgästhus i Rinkeby nyttjat en fasadmonterad parabolantenn för att ta del av utländska kanaler som sände program på deras modersmål. Hyresvärderna ansåg att parabolantennen stred mot hyresavtalet och beslutade att hyresgästerna skulle vråkas. Efter överklagande fastställdes hyresvärdens beslut av Svea hovrätt.

Europadomstolen, vartill målet överklagades, konstaterade att klagandena hade ett starkt och berättigat intresse av att få ta del av information på sitt modersmål, särskilt eftersom informationen innefattade politiska, sociala och kulturella nyheter av stor vikt för klagandenas känsla av samhörighet och delaktighet. De skäl som hyresvärderna och staten åberopat till stöd för förbudet, bland annat säkerhetshänsyn och estetiska hänsyn, ansågs inte väga tungt i förhållande till klagandenas yttrande- och informationsfrihet. Därför hade staten, genom att upprätthålla vråkningsbeslutet, överträtt artikel 10.

I en tidig passus om målets admissibility konstaterade Europadomstolen:

*”Admittedly, the Court is not in theory required to settle disputes of a purely private nature. That being said, in exercising the European supervision incumbent on it, it cannot remain passive where a national court’s interpretation of a legal act, be it a testamentary disposition, a private contract, a public document, a statutory provision or an administrative practice appears unreasonable, arbitrary, discriminatory or, more broadly, inconsistent with the principles underlying the Convention.”*³⁶

³⁵ *Khurshid Mustafa and Tarzibachi v Sweden*, no 23883/06, dom av den 16 december 2008.

³⁶ *Khurshid Mustafa and Tarzibachi v Sweden*, p 33.

Liknande principer har också tidigare vuxit fram i USA. Ett omdiskuterat fall från den amerikanska högsta domstolen, som också refereras i det ovan nämnda Europadomstolsmålet *Appleby and Others v the United Kingdom*, är fallet *Marsh v Alabama*.³⁷

Klaganden i målet, Grace Marsh, var ett Jehovas vittne som missionerade i den lilla hamnstaden Chickasaw i Alabama. Chickasaw såg ut och fungerade precis som många andra småstäder – fast med ett viktigt undantag: staden ägdes av det privata skeppsvarvet Gulf Shipbuilding Corporation. Markägaren hade låtit tapetsera staden med skyltar som angav (fritt översatt): ”Detta är privat egendom, och utan skriftligt tillstånd tillåts inga gatu- eller dörrförsäljare, affärsombud eller annan form av uppsökande verksamhet”. När Grace Marsh hade ställt sig utanför det lokala postkontoret för att dela ut flygblad blev hon ombedd att gå därifrån. När hon vägrade arresterades hon i stället för olovligt kvarstannande. Marsh försökte övertyga de lokala domstolarna om att hon hade utövat sin grundlagsskyddade yttrandefrihet, men utan framgång. Staden var privatägd, och ansågs därmed kunna behandla sina invånare ungefär som en gäst behandlas på världens nåder.

Högsta domstolen valde emellertid att bifalla Marshs överklagan. I domskälen, vartill fem av åtta justitieråd anslöt sig, angav domstolens majoritet:

*”In our view, the circumstance that the property rights to the premises where the deprivation of liberty, here involved, took place were held by others than the public is not sufficient to justify the State’s permitting a corporation to govern a community of citizens so as to restrict their fundamental liberties and the enforcement of such restraint by the application of a state statute.”*³⁸

Målet avgjordes 1946 – sju år före det att Europakonventionen trädde i kraft. Under senare år har domen emellertid fått ny aktualitet. Vissa debattörer har ansett att avgörandet skulle kunna tillämpas analogt i förhållande till de större sociala medieplattformarna, som Facebook, Twitter och Youtube, och därigenom utgöra ett stöd för att ålägga plattformarna ett mer eller mindre direkt ansvar att upprätthålla yttrandefriheten så som den kommer till uttryck i den amerikanska konstitutionens första tillägg.³⁹

³⁷ Marsh v State of Alabama, 326 U.S. 501 (1946).

³⁸ Marsh v State of Alabama (sjunde stycket i majoritetens domskäl).

³⁹ Se till exempel Neidenbach, S och Berezow, A, *Facebook Is Censoring Science and Journalism*, National Review, 2016-08-31 (2023-04-10), <https://www.nationalreview.com/2016/08/facebook-censors-science-journalism-removes-pro-gmo-page/>. Se även Everett (2018).

Även om yttrandefriheten inte kan sägas ge upphov till några direkta förpliktelser för plattformslieferantörerna vore det således ett misstag att avfärda frågeställningen som betydelslös, missriktad eller rättsligt irrelevant – vare sig i en amerikansk eller europeisk kontext. För det fall statens positiva förpliktelser aktualiseras finns det i själva verket få principiella hinder för att yttrande friheten skulle kunna ges en konkret inverkan på förhållandet mellan en plattformslieferantör och dess användare. Som framgår i det följande finns det dessutom goda argument för att statens positiva förpliktelser kan göras gällande med särskild styrka i förhållande till den innehållsmoderering som sker på flera av de större sociala medieplattformarna.

2.2 ÄR DET PLATTFORMSLEVERANTÖRERNA SJÄLVA SOM RÅDER ÖVER INNEHÅLLSMODERERINGEN?

Alla sociala medieplattformar utövar någon form av innehållsmoderering: de sällar, kurerar, arrangerar, raderar, ordnar och sammanställer användargenererat innehåll utifrån olika principer och intressen.⁴⁰ Modereringen sker delvis öppet (till exempel genom uttryckliga innehållsrestriktioner) och delvis automatiserat (exempelvis genom olika sorteringsalgoritmer som väljer och rangordnar innehållet i en given användares informationsflöde).⁴¹

Ytterst regleras modereringen av plattformens egen innehållspolicy. Innehållspolicyn utgör vanligtvis en del av det avtal som användaren ingår med plattformslieferantören i samband med att användaren ansluter sig till eller börjar nyttja tjänsten.⁴² Innehållet varierar från plattformstjänst till plattformstjänst. I stor utsträckning tillämpas emellertid samma centrala innehållsrestriktioner på samtliga större plattformstjänster (som Facebook, Instagram, Twitter och Youtube). Plattformarna tolererar typiskt sett inte beteenden som innefattar *spam*, identitetsförfalskning och bedrägeri, upphovsrättsintrång, eller publicering av integritetskränkande eller känsliga uppgifter om andra. Innehållspolicyerna innefattar också regelmässigt förbud mot trakasserier och hot; missaktande eller hatiskt tal mot vissa skyddsvärda kollektiv; uppvigling, uppmuntran, anstiftan eller erbjudande om att utföra brott; skildringar av våld och vissa brott; glorifiering, hyllande eller lovordande av terrororganisationer; och, allt som oftast, pornografi eller annat erotiskt innehåll.⁴³

⁴⁰ Se Gillespie, T, *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*, Yale University Press 2018, s 5 ff.

⁴¹ Se Gillespie, T, *The Relevance of Algorithms*. I: Gillespie, T, Boczkowski, P & Foot, K (red) *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press 2014.

⁴² Se Gillespie (2018), s 46.

⁴³ Se, för en längre jämförande redogörelse, Thorslund, F, *Twitter, Tumblr, troll och porr – Om sociala mediars ansvar för yttrande och informationsfrihet*, Uppsala universitet 2019, DiVA, <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-384469>.

Ofta motsvarar innehållspolicyerna en sorts allmänt accepterad lägstanivå för samtalsklimatet i ett offentligt rum. I vissa fall har policyernas utformning däremot varit mer diskutabla. Under 2018 kompletterade till exempel Facebook sina plattformregler med nya riktlinjer kring ”sexuella kontakter” (*sexual solicitation*). De nya riktlinjerna innebar att inte bara pornografiskt innehåll, utan också innehåll som ”förenklar, uppmuntrar eller koordinerar sexuell kontakt mellan vuxna” blev otillåtet på plattformen. Förbudet innefattade sexuell explicit språk som kan leda till sexuellt kontaktsökande, ”sexualiserad slang”, omnämnande av sexuella preferenser eller partnerpreferenser, omnämnande av ”vanligen sexualiserade kroppsdelar”, och ”vaga, suggestiva uttalanden” som kan innebära implicit sexuell kontaktsökande.⁴⁴ De nya riktlinjerna kom att kritiserats som alltför långtgående. Kritikerna uttryckte bland annat oro över att liknande förbud skulle kunna komma att användas för att utestänga och försvåra för HBTQ-organisationer och medlemmar av sexuella minoritetsgrupper.⁴⁵

Ett annat exempel är plattformslieferantörernas förbud mot lovordande av terrorism och kriminella organisationer, som ofta tenderar att vara mycket brett utformade. Meta, vars innehållspolicy gäller för såväl Facebook som Instagram, definierar till exempel en kriminell organisation som ”en sammanslutning av tre eller fler personer som är förenade under ett namn, färger, handgester eller erkända tecken och som har deltagit i, eller hotar att delta i, kriminell aktivitet [...]”.⁴⁶ Riktlinjerna har kritiserats som illa överensstämmande med sedvanliga yttrandefrihetsstandarder. Det har bland annat påpekats att kontroversiella organisationer, vars verksamhet kriminaliserats eller terroristklassificerats i en viss stat, i själva verket kan betraktas som legitima frihetsrörelser i en annan.⁴⁷

⁴⁴ Meta, *Facebooks communityregler – Sexuella kontakter (version 15 okt 2018)*, Meta Transparency Center, 2018-10-15 (2023-04-10), <https://transparency.fb.com/sv-se/policies/community-standards/sexual-solicitation/>. Policyn finns fortfarande kvar, med vissa justeringar.

⁴⁵ Se Austin, E, *We can no longer talk about sex on Facebook in Europe*, European Digital Rights, 2019-01-16 (2023-04-10), <https://edri.org/our-work/we-can-no-longer-talk-about-sex-on-facebook-in-europe/>. Harmon, E, *Facebook’s Sexual Solicitation Policy is a Honey-pot for Trolls*, Electronic Frontiers Foundation, 2018-12-07 (2023-04-10), <https://www.eff.org/deeplinks/2018/12/facebook-sexual-solicitation-policy-honey-pot-trolls>. York, J, *Blunt Policies and Secretive Enforcement Mechanisms: LGBTQ+ and Sexual Health on the Corporate Web*, Electronic Frontiers Foundation, 2018-10-24 (2023-04-10), <https://www.eff.org/sv/deeplinks/2018/10/blunt-policies-and-secretive-enforcement-mechanisms-lgbtq-and-sexual-health>.

⁴⁶ Se Meta, *Facebooks communityregler – Farliga individer och organisationer*, Meta Transparency Center, 2021-12-23 (2023-04-10), <https://transparency.fb.com/sv-se/policies/community-standards/dangerous-individuals-organizations/>.

⁴⁷ Se Article 19, *Twitter Rules: Analysis against international standards on freedom of expression*, Article 19, 2018-09-05 (2023-04-10), <https://www.article19.org/resources/twitter-rules-analysis-against-international-standards-on-freedom-of-expression/>. York, J & Greene, D, *Facebook’s*

Utformningen och tillämpningen av plattformarnas innehållspolicier – innehållsmodereringen – kan styras av flera olika överväganden. Modereringen kan ske av hänsyn till plattformsföretagets egna etiska övertygelse. Den kan ske med beaktande av tjänstens övergripande funktion och inriktning (om tjänsten är avsedd för politisk diskussion, ren underhållning eller livsstilsfrågor; om tjänsten är riktad mot barn eller vuxna; om tjänsten är avsedd för en bredare allmänhet eller en viss bransch eller intressegemenskap, och så vidare). Den kan också ske med hänsyn till det egna användarkollektivets presumerade uppfattningar – och, sannolikt, annonsörernas önskemål och vilja.⁴⁸

Plattformarnas innehållsmoderering kan också drivas av ansvarsincitament. En del av plattformarnas riktlinjer kan nämligen härledas till statlig lagstiftning som ålägger plattformarna skyldigheter att ingripa mot visst innehåll. Som exempel antas Facebooks ovan nämnda riktlinjer kring sexuella kontakter härstamma från amerikansk lagstiftning mot trafficking och prostitution: det så kallade *FOSTA/SESTA*-initiativet.⁴⁹ Men även i fråga om terror- och extremistrelaterat innehåll, trakasserier, hets mot folkgrupp och annan cyberbrottslighet har olika stater världen över ålagt plattformslieferantörer olika former av mellanhandsansvar som fått en direkt eller indirekt inverkan på plattformarnas innehållspolicier och moderering.⁵⁰ I praktiken blir plattformslieferantören på så sätt ett verktyg för att genomdriva policier och förbud som i själva verket utarbetats av statliga eller överstatliga aktörer.

I början av 2018 utfärdade EU-kommissionen en rekommendation som upp-muntrade medlemsstaterna att ta fram nya verktyg för att effektivt komma till rätta med olagligt innehåll online.⁵¹ Rekommendationen vände sig huvudsakligen mot olika värdtjänstleverantörer,⁵² med precisa rekommendationer

Secret “Dangerous Organizations and Individuals” List Creates Problems for the Company—and Its Users, ElectronicFrontiersFoundation,2021-12-21 (2023-04-10),<https://www.eff.org/sv/deeplinks/2021/12/facebook-secret-dangerous-organizations-and-individuals-list-creates-problems/>.

⁴⁸ Se Gillespie (2018), s 9 ff.

⁴⁹ Se Austin (2019) och Harmon (2018). *FOSTA/SESTA* är en akronym för de två amerikanska lagarna *Allow States and Victims to Fight Online Sex Trafficking Act* and *Stop Enabling Sex Traffickers Act* som trädde i kraft i april 2018. Lagstiftningen innehåller bland annat regler som syftar till att motverka medhjälp till, underlättande av och understödjande av människohandel. Se, för en mer ingående beskrivning, Born, E, *Too Far and Not Far Enough: Understanding the Impact of FOSTA*, New York University Law Review 2019 (94:1623) s 1623–1653.

⁵⁰ Se Gillespie (2018), s 33–40 och Laidlaw, E, *Regulating Speech in Cyberspace – Gatekeepers, Human Rights and Corporate Responsibility*, Cambridge University Press 2015, s 4–6. Se även FN:s generalförsamling, A/HCR/38/35, s 7–8.

⁵¹ Kommissionens rekommendation (EU) 2018/334 om åtgärder för att effektivt bekämpa olagligt innehåll online.

⁵² Begreppet ”värdtjänstleverantör” definierades i rekommendationen som ”en leverantör

som tog sikte på vilket ansvar sådana aktörer bör åläggas för användargenererat innehåll. Bland annat föreslogs inrättandet av särskilda flaggningssystem för misstänkt olagligt innehåll och att tjänstetillhandahållarna skulle inrätta ”automatiska verktyg” för att upptäcka och beivra brottsligt innehåll på deras plattformar.⁵³ Rekommendationen föreslog också att tjänstetillhandahållarna skulle åläggas att samarbeta med varandra, med medlemsstaternas myndigheter och med Europol, bland annat genom att rapportera misstänkta brott begångna på deras plattformar.⁵⁴ Särskilt riktade rekommendationen in sig på terrorismrelaterat innehåll. Rekommendationen föreskrev uttryckligen att plattformarna, i sina innehållsriktlinjer, skulle förbjuda förekomsten av terrorismrelaterat innehåll och inrätta automatiserade system för att upptäcka och radera sådant innehåll.⁵⁵

Rekommendationen kom senare att ligga till grund för en bindande förordning – Europaparlamentets och rådets förordning (EU) 2021/784 om åtgärder mot spridning av terrorisminnehåll online (även kallad *TERREG*) – som trädde i kraft i juni 2022.⁵⁶ Enligt förordningen är värdtjänstleverantörer⁵⁷ som tillhandahåller sina tjänster i EU skyldiga att radera ”terrorisminnehåll”⁵⁸ från sina plattformar senast inom en timme från det att de mottagit en särskild avlägsnandeorder från någon av EU:s medlemsstater. Förordningen stadgar också uttryckligen att plattformslieferantörer som exponeras för terrorisminnehåll är skyldiga att, i sina användarvillkor, ”inkludera samt tillämpa bestämmelser om åtgärder mot missbruk av dess tjänster för spridning till allmänheten av terrorisminnehåll” och i övrigt ”vidta specifika åtgärder för att skydda sina tjänster mot spridning till allmänheten av terrorisminnehåll”.⁵⁹ Vid systematisk eller fortgående

av informationssamhällets tjänster som består av lagring av information som tillhandahållits av tjänstemottagaren på dennas begäran, i den mening som avses i artikel 14 i direktiv 2000/31/EG, vars verksamhet riktas till konsumenterna som är bosatta i unionen, oavsett leverantörens etableringsort”. Se artikel 5(a) i rekommendationen.

⁵³ Se artikel 5 och 18 i rekommendationen.

⁵⁴ Se artikel 22–28 och 38–40 i rekommendationen.

⁵⁵ Se artikel 30–31 och 36–37 i rekommendationen.

⁵⁶ Jfr skäl 6 i förordningen.

⁵⁷ Begreppet definieras i förordningen på liknande sätt som i rekommendationen, jfr artikel 2(1) i förordningen. I förordningens preambel framgår att begreppet avses omfatta bland annat ”leverantörer av sociala medietjänster, video-, bild- och ljuddelningstjänster, samt fildelnings-tjänster och andra molntjänster, i den mån som dessa tjänster används för att göra den lagrade informationen tillgänglig för allmänheten på direkt begäran av innehållsleverantören”. Se skäl 14 i förordningen.

⁵⁸ Begreppet omfattar bland annat material som innefattar hot om, anstiftan till eller värvning av medgärningsmän till vissa gärningar som klassificeras som terrorbrott enligt artikel 3(1) i Europaparlamentets och rådets direktiv (EU) 2017/541 om bekämpande av terrorism.

⁵⁹ Se artikel 5 i förordningen.

underlåtenhet att fullgöra avlägsnandeskyldigheten ska plattformsleverantören bland annat kunna bli föremål för sanktionsavgifter uppgående till fyra procent av bolagets totala årsomsättning under föregående räkenskapsår.⁶⁰

Liknande ansvarslagstiftning har också antagits i olika EU-medlemsstater. År 2017 antog Tyskland en lagstiftningsakt (*Netzdurchführungsgesetz*, ofta kallad *NetzDG*) som bland annat ålägger leverantörer av sociala medieplattformar att blockera eller avlägsna vissa typer av olagligt innehåll inom 24 timmar efter mottagande av ett klagomål från en användare. Raderings- och blockeringskyldigheten omfattar flera olika sorters innehåll som kriminaliserats under den tyska strafflagen, bland annat uppvigling (inklusive förintelseförnekelse), förolämpningar, förtal, ”illvilligt skvaller”, olaga integritetsintrång och olaga hot. Vid systematiska överträdelse av kraven i *NetzDG* kan leverantören bli föremål för sanktionsavgifter uppgående till högst 50 miljoner euro.⁶¹

I Frankrike har två liknande lagstiftningsakter antagits. Den ena, *Loi no 2018–1202*, tar särskilt sikte på spridningen av falsk information i valtider, och möjliggör bland annat för nationella domstolar att utfärda särskilda förelägganden som förpliktigar plattformsleverantörer att blockera eller radera ”felaktiga eller missvisande påståenden eller anklagelser avseende fakta som sannolikt kan påverka integriteten av ett kommande val”. Den andra, kallad *Loi Avia (Loi no 2020–766)*, ålade plattformsleverantörer en straffsanktionerad skyldighet att radera vissa typer av olagligt innehåll (innefattande bland annat hets mot folkgrupp) inom 24 timmar från mottagandet av en anmälan. Lagen blev föremål för omfattande kritik och förklarades slutligen ogiltig av den franska konstitutionsdomstolen så som stridande mot konstitutionella yttrandefrihetsstandarder.⁶²

2.3 EN ”UTKONTRAKTERING” AV YTTRANDEFRIHETSINSKRÄNKNINGAR?

Den innehållsmoderering som i dag sker på sociala medieplattformar kan inte sägas ske uteslutande på plattformsleverantörens eget initiativ, som ett utflöde av plattformsleverantörens egendomsrätt eller närings- och avtalsfrihet. Inte sällan sker modereringen i stället som en sorts förlängning av diverse lagstiftningsinitiativ och myndighetsbeslut.

⁶⁰ Se artikel 18 i förordningen.

⁶¹ Jfr Zurth, P, *The German NetzDG as Role Model or Cautionary Tale? Implications for the Debate on Social Media Liability*, *Fordham Intellectual Property, Media and Entertainment Law Journal* 2021 (31:4) s 1084–1153.

⁶² Se Steiger, D, *Protecting Democratic Elections Against Online Influence via “Fake News” and Hate Speech – The French Loi Avia and Loi No 2018–1202, the German Network Enforcement Act and the EU’s Digital Services Act in Light of the Right to Freedom of Expression*. I: Kotzur, M, Schiedermaier, S, Steiger, D & Wendel, M, *Theory and Practice of the European Convention on Human Rights*, Nomos 2022.

Fenomenet är långt ifrån oproblematiskt. Kritiska röster har framhårdat att det i själva verket rör sig om ett slags ”utkontraktering” av statliga yttrandefrihetsinskränkningar, vilka ofta ligger utom räckhåll för de konstitutionella *checks and balances* som hade gällt om modereringen i stället hade utformats och verkställts av en statlig aktör. I en rapport från 2017 konstaterar FN:s särskilde rapportör för främjande och skyddande av yttrandefriheten att stater i allt högre utsträckning ”utövar censur genom den privata sektorn”.⁶³ Även Joe McNamee, verkställande direktör vid den ideella rättighetsorganisationen European Digital Rights, har uttryckt farhågor över att internetföretag i vissa fall åläggs att utöva ”frivillig” censur som aldrig kan bli föremål för demokratisk insyn och rättsliga prövningar.⁶⁴

Till de principiella invändningarna kommer dessutom ett antal praktiska problem. Om en plattformsleverantör åläggs att ingripa mot vissa innehållskategorier – hur väl definierade de än må vara – finns alltid en överhängande risk att plattformsleverantören väljer att tillämpa en viss ”säkerhetsmarginal” vid den faktiska modereringen.⁶⁵ Incitamenten lär med andra ord gå mot att hellre moderera för mycket än för lite.

Dessutom präglas innehållsmodereringen på sociala medieplattformar inte av den omsorg och noggrannhet som hade gällt vid en vanlig domstolsprövning. Plattformsleverantörerna använder i stor utsträckning automatiserade verktyg för att identifiera och ingripa mot innehåll som strider mot riktlinjerna.⁶⁶ I de fall manuell granskning sker kan granskningen ofta antas vara extremt summarisk.⁶⁷

Även om plattformsleverantörerna framhäver vikten av att beakta avsikt och kontext vid innehållsmodereringen är det förmodligen svårt att ombesörja verkligt nyanserade bedömningar i praktiken. Risken för misstag och felaktigheter lär vara hög. Skillnaden mellan exempelvis felaktig information,

⁶³ FN:s generalförsamling, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/HRC/35/22, 2017-03-30, s 20.

⁶⁴ Se European Digital Rights, *EU Commission's Recommendation: Let's put internet giants in charge of censoring Europe*, EDRI, 2018-03-01 (2023-04-10), <https://edri.org/our-work/eu-commissions-recommendation-lets-put-internet-giants-in-charge-of-censoring-europe/>.

⁶⁵ Se Riordan, J, *The Liability of Internet Intermediaries*, Oxford University Press (2019), 1.30 och 8.149–8.152 samt Born (2019), s 1647 ff.

⁶⁶ Se Gillespie (2018), s 97 ff. Se även Meta, *Så här fungerar verkställighetstekniken*, Meta Transparency Center, 2022-01-19 (2023-04-10), <https://transparency.fb.com/sv-se/enforcement/detecting-violations/how-enforcement-technology-works/>.

⁶⁷ Tidigare har exempelvis beskrivits hur Facebooks ”klickarbetare” – innehållsmoderatorer på lägsta nivå, som bedömer flaggat innehåll och dess förenlighet med användarvillkoren – arbetar sig igenom flaggat innehåll inom intervaller om tio sekunder. Langvardt, K, *Regulating Online Content Moderation*, Georgetown Law Journal 2018 (106:5) s 1353–1358.

som enligt Metas innehållsriktlinjer kan bli föremål för olika typer av restriktioner, och satir, som inte bör bli föremål för ingrepp och restriktioner, består inte sällan i avsändarens syfte och motiv.⁶⁸ Vad som ska anses utgöra satir respektive falska nyheter är naturligtvis svårt att avgöra utifrån en summarisk bedömning, utan god kontextuell och kulturell kännedom.

Det innehåll som kan tänkas bli särskilt hårt drabbat av plattformslieferantörernas innehållsmoderering är således samma typ av innehåll som ofta sägs ligga nära yttrandefrihetens kärna: kontroversiella politiska yttranden, satir, normbrytande innehåll och annat som stör, chockerar eller provocerar. Som framgått ovan är det osäkert om stater verkligen kan förhålla sig passiva till en sådan innehållsmoderering utan att åsidosätta ”de principer som ligger till grund för Europakonventionen”.⁶⁹ Det kan dessutom ifrågasättas om statliga aktörer verkligen kan anses ha varit ”passiva” i förhållande till innehållsmodereringen från första början.⁷⁰ Det finns, kort sagt, starka skäl att överväga någon sorts reglering av innehållsmodereringen.

Frågan är då hur ett bättre värnande av användarnas yttrandefrihet skulle kunna åstadkommas i praktiken. Även här har två skilda huvudlinjer kommit till uttryck i debatten. Den ena föreslår nya materiella ansvarsnormer för plattformsföretagen, medan den andra – som också verkar ha fått störst genomslag under senare år – i stället förordar olika sorters processuella rättigheter för användarna.

3. MATERIELLA ANSVARNORMER: EN SVÅR BALANSGÅNG

Att ålägga enskilda särskilda förpliktelser gentemot andra enskilda kräver alltid en varsam balansgång mellan olika intressen och hänsyn. Den enes rätt kan naturligtvis inte ge upphov till oproportionerligt tunga eller rent av omöjliga förpliktelser för den andre. I fråga om statens positiva förpliktelser har Europadomstolen också konstaterat att förpliktelserna inte får ge upphov till omöjliga eller oproportionerliga bördor för myndigheterna, och att hänsyn också måste tas till den reglerades konventionsskyddade rättigheter.⁷¹ Det låter sig nästan gå osagt att det rör sig om en mycket komplex avvägning mellan motstående viljor och intressen.

Den mest direkta lösningsmodellen vore förstås att implementera materiella ansvarsnormer som helt enkelt förpliktigar plattformslieferantörer att upprätthålla

⁶⁸ Se Meta, *Facebooks communityregler – Felaktig information*, Meta Transparency Center, u.å. (2023-04-10), <https://transparency.fb.com/sv-se/policies/community-standards/misinformation/>.

⁶⁹ Jfr ovan avsnitt 2.1 och not 33.

⁷⁰ Jfr ovan avsnitt 2.2.

⁷¹ *Appleby and Others v the United Kingdom* p 40–43.

användarnas yttrandefrihet, såsom den kommer till uttryck i till exempel Europakonventionen eller FN:s konvention om medborgerliga och politiska rättigheter (ICCPR). En sådan ansvarsmodell skulle uppenbarligen inte utgå från en direkt tillämpning av konstitutionella eller folkrättsliga yttrandefrihetsnormer⁷² – eftersom den horisontella verkan av sådana normer i bästa fall kan beskrivas som oklar.⁷³ Det skulle snarare vara fråga om att utforma ny lagstiftning som ”absorberar” de principer som kommer till uttryck i fri- och rättighetsnormerna och gör dem tillämpliga i förhållandet mellan enskilda.⁷⁴ Ett exempel på en sådan lösningsmodell, som också förts fram, vore att göra FN:s riktlinjer för företagsverksamhet och mänskliga rättigheter juridiskt bindande för plattformsföretagen.⁷⁵ Förslaget hade i princip inneburit att plattformsleverantörer som Meta, Twitter och Youtube skulle förpliktas att respektera yttrande- och informationsfriheten, så som den definieras enligt FN:s deklaration om mänskliga rättigheter, i sin verksamhet. Plattformsleverantörerna hade med andra ord ålagts att agera som om de vore en del av det allmänna.

En sådan ansvarsmodell har både fördelar och nackdelar. Fördelen är att de befintliga yttrandefrihetsnormerna – så som de definierats i till exempel Europakonventionen eller regeringsformen, och så som de kommit att utvecklas genom årtionden av domstolspraxis – speglar en eftertänksam avvägning mellan yttrandefriheten å ena sidan och de motstående samhällliga intressena å andra. De är allmänt erkända och väl förankrade i det allmänna rättsmedvetandet. I stor utsträckning ligger de dessutom redan till grund för utformningen av plattformarnas befintliga modereringsregler. Meta beskriver bland annat att målet bakom företagets innehållspolicyer är ”att skapa en plats för uttryck och ge människor möjlighet att göra sina röster hörda” och att reglerna bland annat bygger på ”rådgivning från experter inom [...] mänskliga rättigheter”.⁷⁶

Att ålägga plattformsleverantörer en direkt skyldighet att efterleva mänskliga rättighetsstandarder, så som de kommer till uttryck i till exempel Europakonventionen, FN:s rättighetskataloger eller nationella konstitutioner, skulle samtidigt medföra en del slagsidor. De konstitutionella och internationella människorättsnormerna är förhållandevis öppet utformade och anpassade för

⁷² Jfr diskussionen om *direct application model* i Barak (1996), s 224 ff.

⁷³ Se ovan avsnitt 2.1.

⁷⁴ Jfr diskussionen om *indirect application model* i Barak (1996), s 236 ff.

⁷⁵ Se *Statement on behalf of a Group of Countries at the 24rd [sic] Session of the Human Rights Council -Transnational Corporations and Human Rights*, Business & Human Rights Resource Centre 2013-09-01, <https://www.business-humanrights.org/en/latest-news/pdf-statement-on-behalf-of-a-group-of-countries-at-the-24rd-session-of-the-human-rights-council/>. Jfr även Everett (2018).

⁷⁶ Meta, *Facebooks communityregler – Introduktion*, Meta Transparency Center, u å (2023-04-11), <https://transparency.fb.com/sv-se/policies/community-standards/>.

försiktiga och eftertänksamma domstolsprocesser snarare än de ögonblicksbeslut som präglar de sociala mediernas praktiska moderering. Det framstår som osannolikt att en rimlig avvägning mellan den konstitutionella yttrandefriheten och ett starkt, motstående intresse – som rätten till privatliv eller skyddet för upphovsrätten – hade låtit sig göras av en Meta-anställd ”klickarbetare” inom ett tiosekundersintervall.⁷⁷ Lösningssmodellen hade också inneburit en stor börda i plattformsleverantörernas verksamhet. Om plattformsleverantörer förväntades bedöma fri- och rättighetskonflikter på samma sätt som en konstitutionsdomstol hade det sannolikt resulterat i oförsvarliga arbetsinsatser och kostnader för företagen. Det hade med andra ord utgjort ett ganska besvärande ingrepp i plattformsleverantörernas egna fri- och rättigheter.

En närliggande – men något mindre långtgående – ansvarsmodell vore att i stället introducera mer specificerade, konkreta och avgränsade skyldigheter: skyldigheter som vore ägnade att främja yttrandefriheten snarare än att reproducera dess normativa innehåll.⁷⁸ EU:s dataskyddsförordning⁷⁹ kan ses som ett exempel på en sådan lagstiftningsmodell. Förordningen ålägger företag vissa mer eller mindre specifika skyldigheter, som sammantagna syftar till att förbättra skyddet för fysiska personers personuppgifter och privatliv.⁸⁰ Ett annat exempel är de nätneutralitetsregler som följer av EU:s TSM-förordning, vars syfte delvis varit att skydda och främja yttrande- och informationsfriheten på internet.⁸¹ Som tidigare nämnt finns liknande normer om neutralitet, icke-diskriminering och objektivitet vid informationsförmedling också i flera andra sammanhang.⁸²

En reglering med liknande inslag finns numera etablerad i Europaparlamentets och rådets förordning (EU) 2022/2065 om en inre marknad för digitala tjänster (*Digital Services Act* eller *DSA*), som trädde i kraft den 16 november 2022. Förordningen, som reglerar olika typer av digitala förmedlingstjänster, syftar bland annat till att skapa en ”säker, förutsebar och tillförlitlig onlinemiljö” och göra det möjligt för EU-medborgare att ”utöva sina grundläggande rättigheter

⁷⁷ Jfr ovan not 28 och 66.

⁷⁸ Jfr Barak (1996), s 236. Barak noterar att flera grundläggande ansvarsprinciper inom civilrätten i själva verket kan anses återspegla en rimlig avvägning mellan motstående mänskliga rättigheter. Civilrättsliga koncept som ”god tro”, ”skälighet” och ”oaktsamhet” kan på så sätt betraktas som en inkarnation av mänskliga rättighetsstandarder i civilrättslig skepnad.

⁷⁹ Europaparlamentets och rådets förordning (EU) 2016/679 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter.

⁸⁰ Se skäl 1–13 i förordningen.

⁸¹ Se ovan avsnitt I.

⁸² Se ovan avsnitt I.

[...], särskilt yttrande- och informationsfriheten, näringsfriheten, rätten till icke-diskriminering och uppnåendet av en hög konsumentskyddsnivå”.⁸³

I förordningen anges särskilt att ”[l]everantörer av förmedlingstjänster ska agera på ett aktsamt, objektivt och proportionellt sätt vid tillämpningen och kontrollen av efterlevnaden av [förmedlingstjänstens allmänna villkor], med vederbörlig hänsyn till alla berörda parter rättigheter och legitima intressen, inbegripet tjänstemottagarnas grundläggande rättigheter, såsom yttrandefriheten, mediernas frihet och mångfald och andra grundläggande rättigheter och friheter enligt stadgan.”⁸⁴

Förordningens definition av förmedlingstjänster inrymmer bland annat online-plattformar – innefattande sociala medieplattformar som Youtube och Twitter.⁸⁵ Förordningen innebär därmed ett krav på sådana plattformslieferantörer att iaktta aktsamhet, objektivitet och proportionalitet när de tillämpar sina innehållspolicyer för att ingripa mot innehåll, eller mot användare som anses agera i strid med innehållspolicyerna. Bestämmelsen kan med andra ord sägas förhindra en godtycklig och diskriminerande modereringspraxis.⁸⁶

I bestämmelsen uppställs däremot inga uttryckliga krav på utformningen av plattformslieferantörens innehållspolicyer och villkor. De objektivitets- och proportionalitetskrav som framgår av artikel 14(4), liksom kraven på att iaktta ”vederbörlig hänsyn” till tjänstemottagarnas grundläggande rättigheter, är i princip avgränsade till ”tillämpningen och kontrollen av efterlevnaden av” plattformslieferantörens villkor.⁸⁷ Om kravet däremot skulle anses tillämpligt också på utformningen av innehållspolicyerna,⁸⁸ är det å andra sidan svårt att uppfatta vilka konkreta skyldigheter som bestämmelsen medför. Utfallet av regleringen skulle då bli dubbelt tveksam – med en rättslig osäkerhet

⁸³ Se skäl 3 i förordningen.

⁸⁴ Se artikel 14(4) i förordningen.

⁸⁵ Se artikel 3(g) och (i) i förordningen. Begreppet ”förmedlingstjänst” inrymmer underkategorierna värdtjänster, cachingtjänster och så kallade mere conduit-tjänster: tjänster för ”överföring i ett kommunikationsnät av information som tillhandahållits av en tjänstemottagare, eller tillhandahållande av tillgång till ett kommunikationsnät”. ”Onlineplattformar” definieras som ”en värdtjänst som på begäran av en tjänstemottagare lagrar och sprider information till allmänheten, såvida inte denna verksamhet enbart utgör en extrafunktion av mindre betydelse [...]”.

⁸⁶ Se skäl 26 i förordningen.

⁸⁷ Se skäl 45–47 i förordningen, där det emellertid nämns att plattformslieferantörerna bör avstå från diskriminering och godtycke samt iaktta användarnas ”rättigheter och legitima intressen” även vid ”utformning” av innehållsbegränsningar.

⁸⁸ Jfr Quintais, J P, Appelman, N & Fahy, R, *Using Terms and Conditions to Apply Fundamental Rights to Content Moderation*, SSRN, 2022-12-14 (2023-04-11), <https://ssrn.com/abstract=4286147>.

kring användarnas materiella rättigheter och en lika stor osäkerhet kring hur plattformslieferantörerna förväntas agera för att fullfölja sina skyldigheter.

4. EFFEKTIVA RÄTTSMEDEL: DET NYA SVARTA

Att stifta lagar som direkt ingriper i plattformslieferantörernas egna innehållspolicyer – en sorts ”moderering av modereringen” – placerar en tung börda på leverantörernas axlar. Sett i ljuset av den ansvarsreglering som implementerats i såväl EU som USA när det gäller bland annat radering av terrorismrelaterat innehåll,⁸⁹ men även blockering av upphovsrättskyddat material⁹⁰ och före byggande av prostitution och trafficking,⁹¹ skulle leverantörerna i många fall vara förpassade till en nästintill omöjlig juridisk balansgång, med ansvar för ”övermoderering” å ena sidan, och ansvar för ”undermoderering” å andra. I ett större perspektiv kan man dessutom fråga sig om en utveckling i den riktningen – ett slags utkontraktering av ansvaret för och kontrollen över det fria ordets gränser – ens vore önskvärd.

Som alternativ till att reglera vad som får och inte får publiceras på plattformarna i rent materiellt hänseende, har i stället framförts förslag om att på olika sätt utverka *processuella rättigheter* för användarna.⁹² Gemensamt för denna typ av förslag är att de inte tar sikte på att reglera normerna som styr innehållsmodereringen, utan syftar till att öka transparensen kring plattformslieferantörernas modereringspraxis och möjliggöra en prövning av modereringsbeslut som uppfattas som felaktiga eller opropotionerliga. Som framgår nedan är det också denna sorts lösningsförslag som tycks ha fått störst praktiskt genomslag: olika processuella remedier har under senare tid införts både av plattformarna själva, som en sorts självreglering, och i mer institutionaliserade former under EU:s DSA-förordning.

⁸⁹ Se ovan avsnitt 2.2.

⁹⁰ Se artikel 17 i Europaparlamentets och rådets direktiv (EU) 2019/790 om upphovsrätt och närstående rättigheter på den digitala inre marknaden och Quintais, J P, *The New Copyright in the Digital Single Market Directive: A Critical Look*, *European Intellectual Property Review* 2020 (42:1) s 28–41.

⁹¹ Se ovan avsnitt 2.2.

⁹² Se Windwehr, S & Schmon, C, *Our EU Policy Principles: Procedural Justice*, Electronic Frontiers Foundation, 2020-07-27 (2023-04-11), <https://www EFF.org/deeplinks/2020/07/our-eu-policy-principles-procedural-justice>. Mostert, F & Urbelis, A, *Your day in court: social media needs a system of due process*, *Financial Times*, 2021-05-17 (2023-04-11), <https://www.ft.com/content/48c49453-9a8f-4125-85d7-94220497d13c>.

4.1 SJÄLVREGLERING: METAS "TILLSYNSRÅD"

Under 2020 lanserade Meta ett särskilt Tillsynsråd (*Oversight Board*), med uppgift att överpröva modereringsbeslut på Facebook.⁹³ Rådet har ibland beskrivits som en sorts "högsta domstol" för plattformen.⁹⁴ Det är organisatoriskt fristående från Meta⁹⁵ och har, efter begäran från en användare, befogenhet att fatta beslut om att tillåta eller radera visst innehåll från plattformen.⁹⁶ Besluten är bindande för Meta.⁹⁷ Rådet kan också rekommendera ändringar i plattformarnas innehållspolicier.⁹⁸

I Tillsynsrådets stadgar betonas yttrandefriheten som ett bärande syfte bakom organisationsförändringen. I stadgarna konstateras att yttrandefriheten utgör en grundläggande mänsklig rättighet, och att rådets syfte är att skydda det fria ordet genom att fatta principfasta, självständiga innehållsbeslut.⁹⁹ Även om Tillsynsrådets beslutsbefogenheter primärt ligger i att granska innehållets förenlighet med "Metas innehållspolicier och värden"¹⁰⁰ har internationella yttrandefrihetsrättsliga normer också fått ett stort genomslag i många av de beslut som rådet hitintills fattat.¹⁰¹

När Tillsynsrådet i maj 2021 behandlade Facebooks beslut att stänga av Donald Trumps användarkonton på obestämd tid till följd av den dåvarande presidentens uttalanden i samband med stormningen av kongressbyggnaden,¹⁰² hänvisade rådet uttryckligen till såväl artikel 19 och 20 i ICCPR, som den allmänna kommentar om yttrandefriheten som utarbetats av FN:s kommitté

⁹³ Se Clegg, N, *Welcoming the Oversight Board*, Meta Newsroom, 2020-05-06 (2023-04-11), <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>.

⁹⁴ Se Gradoni, L, *Constitutional Review via Facebook's Oversight Board – How platform governance had its Marbury v Madison*, Verfassungsblog 2021-02-10 (2023-04-11), <https://verfassungsblog.de/fob-marbury-v-madison/>.

⁹⁵ Tillsynsrådet, *Vanliga Frågor - Vilket är rådets förhållande till Facebook-företaget?*, 2020-03 (2023-04-11), <https://www.oversightboard.com/faq/>, samt Tillsynsrådet, *Tillsynsrådets stadga*, 2023-02 (2023-04-11), <https://www.oversightboard.com/governance/>, artikel 5.

⁹⁶ Tillsynsrådets stadga, artikel 1, sektion 4 och artikel 3, sektion 5.

⁹⁷ Tillsynsrådets stadga, artikel 4.

⁹⁸ Tillsynsrådets stadga, artikel 1, sektion 4 och artikel 3, sektion 7.3.

⁹⁹ Tillsynsrådets stadga, "Introduction".

¹⁰⁰ Tillsynsrådets stadga, artikel 2, sektion 2.

¹⁰¹ Se Gradoni (2021).

¹⁰² Tillsynsrådet, *Avstängningen av tidigare president Trump (2021-001-FB-FBR)*, 2021-05-05 (2023-04-11), <https://www.oversightboard.com/decision/FB-691QAMHJ/>.

Den dåvarande presidentens användarkonto hade stängts av under hänvisning till Metas communityregel om "farliga organisationer och individer", som "förbjuder innehåll som hyllar, stödjer eller representerar händelser som Facebook identifierar som terroristattacker, hatbrott, massmord, försök till massmord, seriemord, hatbrott och olagliga evenemang".

för mänskliga rättigheter¹⁰³ och OHCHR:s handlingsplan för att förhindra främjande av diskriminering, aggression och våld (*Rabat Action Plan*).¹⁰⁴ Efter att ha konstaterat att bolaget förvisso inte ålagts några direkta skyldigheter under internationell eller konstitutionell rätt, men likväl åtagit sig att efterleva FN:s vägledande principer om företag och mänskliga rättigheter (innefattande ett åtagande om att inte orsaka eller bidra till men för de mänskliga rättigheterna) konstaterade rådet:

*”Political speech receives high protection under human rights law because of its importance to democratic debate. The UN Human Rights Committee provided authoritative guidance on Article 19 ICCPR in General Comment No. 34, in which it states that free communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential” (para. 20).*¹⁰⁵

Rådet hänvisade också till vedertagna normer för att etablera godtagbara *inskränkningar* i yttrandefriheten. I det hänseendet konstaterade rådet att ”internationell lag” tillåter begränsningar i yttrandefriheten, förutsatt att begränsningen baseras på regler som är tydliga och tillgängliga, utformade för ett berättigat ändamål, nödvändiga och proportionerliga i förhållande till risken för skada.¹⁰⁶ Efter att ha bedömt åtgärdens förenlighet med Facebooks regelverk, förekomsten av berättigat ändamål och proportionaliteten i åtgärden konstaterade Tillsynsrådet att det varit berättigat att begränsa Donald Trumps kontoåtkomst, men att det inte funnits stöd för att stänga av kontot ”på obestämd tid”.¹⁰⁷

I ett annat ärende behandlade rådet Facebooks borttagande av en svensk användares inlägg, under hänvisning till plattformens förbud mot ”sexuellt utnyttjande av barn, barnmisshandel och nakenbilder av barn”.¹⁰⁸ Inlägget som tagits bort bestod av ett arkivfoto på en ung, oidentifierad flicka tillsammans med en längre bildtext. I bildtexten fanns detaljerad information om verkliga övergrepp mot två barn och beskrivningar av de straff som förövarna tilldömts.

¹⁰³ FN:s kommitté för mänskliga rättigheter, *General comment No. 34 – Article 19: Freedoms of opinion and expression*, CCPR/C/GC/34.

¹⁰⁴ Kontoret för FN:s högkommissarie för mänskliga rättigheter (OHCHR), *Annual report of the United Nations High Commissioner for Human Rights – Addendum: Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred*, A/HRC/22/17/Add.4.

¹⁰⁵ Se punkt 8.3 i beslutet.

¹⁰⁶ Se punkt 8.3 i beslutet.

¹⁰⁷ Se punkt 9 i beslutet.

¹⁰⁸ Tillsynsrådet, *Swedish journalist reporting sexual violence against minors* (2021-016-FB-FBR), 2022-02-01 (2023-04-11), <https://www.oversightboard.com/decision/FB-P9PR9RSA/>.

I inlägget riktades dessutom skarp kritik mot det svenska rättsväsendets sätt att hantera sexuellt våld mot minderåriga. Tillsynsrådet beslutade att inlägget skulle återpubliceras.¹⁰⁹ Rådet konstaterade att kommentaren till artikel 19 i ICCPR lyfter fram vikten av att få uttrycka sin politiska åsikt, vilket enligt rådet innefattar ”att få diskutera ett lands kriminalvårdssystem och rapportera om dess funktion i specifika fall”.¹¹⁰ Rådet ansåg att det saknats tillräckligt stöd i plattformens innehållsregler för att ta bort inlägget, och att ett borttagande därmed utgjorde ett oproportionerligt ingrepp i användarens yttrandefrihet.¹¹¹

Tillsynsrådets beslutsmotiveringar är påfallande lika de bedömningar som sker i nationella och internationella domstolar vid en traditionell, ”vertikal” tillämpning av individuella fri- och rättighetsnormer.¹¹² Någon domstolsprövning är det emellertid inte fråga om: rådet liknar snarare ett slags självregleringsorgan. Oavsett Tillsynsrådets organisatoriska självständighet kan rådet inte heller anses vara helt oberoende av Meta. Meta finansierar till exempel den stiftelse som Tillsynsrådet är underställt och utser dess medlemmar.¹¹³ Tillsynsrådets verksamhet är på så sätt varken mer eller mindre än ett ambitiöst sätt att förverkliga Metas självpåtagna *corporate social responsibilities*.

5.2 EXTERNA PRÖVNINGSMEKANISMER: DSA-MODELLEN

Efter det att Meta instiftade sitt Tillsynsråd har liknande initiativ kommit att övervägas också av andra plattformslieferantörer. I oktober 2022 berättade Elon Musk att Twitter planerade att inrätta ett eget ”innehållsmodereringsråd”.¹¹⁴ Videoplattformen TikTok har å sin sida inrättat ett amerikanskt *Content Advisory Council* och ett europeiskt *Safety Advisory Council* som ska bistå med

¹⁰⁹ Se punkt 9 i beslutet.

¹¹⁰ Se punkt 8.3 i beslutet.

¹¹¹ Se punkt 8.3 i beslutet.

¹¹² Jfr till exempel Högsta domstolens beslut i mål nr B 2501-22 (”Anförandet i fullmäktige”), med hänvisning till både regeringsformen 2 kap 1 § och artikel 10 i Europakonventionen. I avgörandet konstaterar Högsta domstolen bland annat att skyddet för ”yttrandefriheten i politiska sammanhang är särskilt starkt” (p 15). Högsta domstolen betonar också att yttrandefriheten endast får underkastas sådana ”inskränkningar som i ett demokratiskt samhälle är nödvändiga” och att begränsningar inte får ”gå utöver vad som är nödvändigt med hänsyn till det ändamål som har föranlett den” (p 13–14). Att begränsningar av yttrandefriheten ska vara tydligt föreskrivna i lag, tjäna ett berättigat ändamål och stå i rimlig proportion till det äsyftade ändamålet gäller såväl enligt Europakonventionen som enligt regeringsformen, se till exempel NJA 2012 s 400 (”Mangamålet”) och *Handyside v the United Kingdom*, no 5493/72, dom av den 7 december 1976.

¹¹³ Tillsynsrådets stadgar, artikel 5, sektion 2–3.

¹¹⁴ Se Frenkel, S, *Elon Musk moves to form a content moderation council for Twitter*, *New York Times*, 2022-10-28 (2023-04-12), <https://www.nytimes.com/2022/10/28/technology/twitter-elon-musk-content-moderation.html>.

expertrådgivning kring plattformens ”innehållspolicyer och modereringspraxis”.¹¹⁵ Utvecklingen ger sken av ett sent försök till självreglering, i en bransch som allt oftare tilldragit sig kritik för osaklighet och bristande insyn i innehållsmodereringen.¹¹⁶ Frågan är emellertid om inte självregleringsinitiativen kommit för sent – eftersom liknande prövningsmodeller nu även kommit att bli föremål för obligatoriska, regulatoriska krav.

I den nya DSA-förordningen finns ett antal särskilda skyldigheter för ”online-plattformar”.¹¹⁷ Kraven innefattar bland annat vissa obligatoriska skyddsåtgärder som ska tillämpas om plattformsleverantören ingriper mot en användare på grund av innehåll som bedöms vara olagligt eller oförenligt med plattformens allmänna villkor. Plattformsanvändare som i sådana fall blivit föremål för beslut om innehållsrestriktioner, kontorestriktioner, tjänsterestriktioner eller så kallad ”demonetisering”¹¹⁸ ska bland annat tillhandahållas en ”tydlig och specifik motivering” till beslutet.¹¹⁹ Motiveringen ska ange de faktiska omständigheter som legat till grund för beslutet, den avtals- eller rättsgrund som legat till grund för beslutet, och om ”automatiserade metoder” använts för att identifiera innehållet eller fatta modereringsbeslutet. Plattformsleverantören ska även tillhandahålla ett ”effektivt internt system för hantering av klagomål” mot beslutet.¹²⁰ Den interna hanteringen av klagomål ska ske på ett ”snabbt, icke-diskriminerande, aktsamt och icke godtyckligt sätt”, och om det visar sig att beslutet inte var motiverat ska plattformsleverantören vara skyldig att ”utan onödigt dröjsmål ändra sitt beslut”.¹²¹

Förordningen föreskriver dessutom att det ska finnas möjlighet för plattformsanvändare att få innehållstvister prövade av ett ”externt” tvistlösningsorgan, genom att få klagomål prövade av ett ”certifierat organ för tvistlösning utanför domstol”.¹²² Det gäller också för innehållstvister som redan prövats

¹¹⁵ Se Pappas, V, *Introducing the TikTok Content Advisory Council*, TikTok, 2020-03-18 (2023-04-12), <https://newsroom.tiktok.com/en-us/introducing-the-tiktok-content-advisory-council>, och de Baillencourt, J, *Meet TikTok's European Safety Advisory Council*, TikTok, 2021-03-02 (2023-04-12), <https://newsroom.tiktok.com/en-gb/tiktok-european-safety-advisory-council>.

¹¹⁶ Se ovan i det inledande avsnittet.

¹¹⁷ I artikel 16 görs dock undantag från de särskilda skyldigheterna för onlineplattformar i förordningen för ”mikroföretag eller små företag enligt definitionen i rekommendation 2003/361/EG”.

¹¹⁸ Artikel 17 talar om ”tillfälligt upphävande, avslutande eller annan begränsning av penningutbetalningar” medan den följande artikel 20 talar om begränsningar i ”möjligheten att monetarisera information”. I båda fall lär avses det som ofta kallas ”demonetisering”, det vill säga begränsningar i möjligheten att tjäna pengar genom plattformen.

¹¹⁹ Se artikel 17 i förordningen.

¹²⁰ Se artikel 20 i förordningen.

¹²¹ Se artikel 20(4) i förordningen.

¹²² Se artikel 21 i förordningen.

genom en intern klagomålsmekanism hos plattformslieferantören. De ”externa” tvistlösningsorganen ska kunna certifieras av en nationell samordnare för digitala tjänster¹²³ om organet visat prov på oberoende och opartiskhet; besitter tillräcklig sakkunskap; tillämpar rättvisa och tydliga förfaranderegler och uppfyller vissa krav på tillgänglighet och effektivitet.¹²⁴ Medlemsstaterna ges också rätt att inrätta egna tvistlösningsorgan. Oavsett organets form framgår direkt av förordningstexten att förfarandet i princip ska vara kostnadsfritt för användaren, och att användaren ska ha rätt till kostnadsersättning om tvisten avgörs till hans eller hennes fördel.¹²⁵

Även om DSA-förordningen redan trätt i kraft kommer de ovannämnda skyldigheterna inte att börja tillämpas förrän i februari 2024.¹²⁶ Vilka aktörer som kan komma i fråga som certifierade tvistlösningsorgan är därmed ännu inte klart. Om till exempel Metas Tillsynsråd kan betraktas som ett certifierbart, externt tvistlösningsorgan är osäkert – inte minst med tanke på att förordningen uppställer specifika krav på att ett externt tvistlösningsorgan ska vara ”opartiskt och oberoende, även oberoende i finansiellt hänseende, från leverantörer av onlineplattformar”.¹²⁷

6. ÄR PRIVATA TVISTLÖSNINGSORGAN TILLRÄCKLIGT?

En ordning som möjliggör ökad transparens kring plattformarnas modereringsbeslut, tillsammans med väl definierade prövningsmöjligheter både internt hos plattformen och externt hos ett certifierat tvistlösningsorgan, är en god grogrund för en rättvisare modereringspraxis och rimligare avvägningar mellan plattformsföretagens och användarnas intressen. Om inte annat är det en tydlig förbättring jämfört med den ordning som rått tidigare. Ur ett yttrandefrihetsperspektiv finns det däremot en del frågor som inte direkt adresseras genom plattformsföretagens egna CSR-åtaganden eller DSA-förordningens obligatoriska prövningsmekanismer.

I Tillsynsrådets stadgar framgår tydligt att Meta aldrig ska vara skyldigt att följa ett beslut från Tillsynsrådet som skulle kunna innebära att Meta gerade i

¹²³ För svensk del har Post- och telestyrelsen, Konsumentverket och Myndigheten för press, radio och tv övervägts som samordnande myndighet för digitala tjänster enligt förordningen. Av de tre myndigheterna har Post- och telestyrelsen ansetts bäst lämpad att tilldelas uppdraget. Se SOU 2023:2 s 98 ff.

¹²⁴ Se artikel 21(3) i förordningen.

¹²⁵ Se artikel 21(5) i förordningen, där det dock framgår att tvistlösnings tjänsten får tillhandahållas ”mot en låg avgift” för användaren.

¹²⁶ Se artikel 93 i förordningen. Jfr dock artikel 92.

¹²⁷ Se artikel 21(3)(a) i förordningen.

strid med gällande lag.¹²⁸ Även när det gäller de externa tvistlösningsorgan som utpekats i DSA-förordningen framgår uttryckligen att ett certifierat organ inte har befogenhet att ”älägga parterna en bindande lösning av tvisten”.¹²⁹ Ett prövningsutslag som kan innebära att plattformsleverantören själv riskerar att exponera sig för rättsligt ansvar är med andra ord svårt att verkställa, vare sig vid en intern eller extern prövning.

Det är heller inte svårt att föreställa sig att det skulle kunna uppstå situationer där plattformsleverantörer anser sig lagligen förhindrade att tolerera visst innehåll – oavsett om en oberoende prövningsinstans bedömt att innehållet ligger inom ramen för de yttrandefrihetsstandarder som kan anses gälla på plattformen. Som konstaterat finns exempel i både amerikansk och europeisk lagstiftning där plattformsleverantörer åläggs ett eget ansvar att ingripa mot vissa innehållstyper. Ett exempel är skyldigheten enligt TERREG att radera ”terrorisminnehåll” inom en timme från det att plattformsleverantören mottagit en avlägsnandeorder från någon av EU:s medlemsstater. De privaträttsliga tvistlösningsorgan som inrättas under DSA har inget givet normprövningsmandat. Om tillämpningen av en sådan tvingande, offentligrättslig skyldighet hamnar i konflikt med en användares yttrandefrihet är det långt ifrån säkert att det kommer finnas några effektiva rättsmedel att tillgå – trots att yttrandefrihetsinskränkningen i själva verket kan sägas emanera från staten.

I och med DSA kommer plattformsleverantörer dessutom åläggas ett strängare ansvar för att komma till rätta med ”olagligt innehåll”. Förordningen utgår från ett slags *notice and takedown*-system, där plattformsleverantörer i princip kan hållas ansvariga för användargenererat innehåll på plattformen under förutsättning att de har fått kännedom om innehållet och inte ”utan dröjsmål” agerat för att avlägsna innehållet eller göra det oåtkomligt för andra.¹³⁰ Ansvarsregleringen kompletteras med regler om inrättandet av obligatoriska anmälningsmekanismer som ska kunna nyttjas av privatpersoner och organisationer för att ”flagga” olagligt innehåll på plattformen, och därigenom försätta plattformsleverantören i ond tro.¹³¹

Vad som utgör ”olagligt innehåll” specificeras emellertid inte i förordningen. I stället definieras ”olagligt innehåll” som all information som ”inte är förenlig med unionsrätten eller gällande rätt i en medlemsstat som står i överensstämmelse med unionsrätten, oavsett denna rätts exakta sakinnehåll eller art”.¹³²

¹²⁸ Se Tillsynsrådets stadgar, artikel 4 och 7.

¹²⁹ Se artikel 21(2) i förordningen.

¹³⁰ Se artikel 6 i förordningen.

¹³¹ Se artikel 16 och 22 i förordningen.

¹³² Se artikel 3(h) i förordningen.

Plattformleverantörer som Meta och Twitter skulle därigenom kunna finna sig skyldiga att verkställa ingrepp mot användarinnehåll enligt spansk, polsk eller ungersk rätt – samtidigt som användare hänvisas till att få ingreppet prövat hos ett privat tvistlösningsorgan utan möjlighet att meddela rättsligt bindande besked.

I förordningen framgår att reglerna kring certifierade tvistlösningsorgan inte ska påverka en användares rätt att ”inleda förfaranden för att inför domstol bestrida besluten från leverantörerna av onlineplattformar i enlighet med tillämplig rätt”.¹³³ Frågan är om den möjligheten inte borde ha utvecklats närmare i lagstiftningsarbetet.¹³⁴ Att väcka talan mot ett plattformsföretag inom ramen för den ordinarie civilrätten får antas vara förhållandevis komplicerat, såväl processuellt som materiellt.¹³⁵ Samtidigt, i takt med ett ökat statligt inflytande över plattformslieferantörernas innehållsmoderering, får behovet av en effektiv domstolsprövning anses ha ökat.

7. SAMMANFATTNING

Det fanns en tid då internet betraktades som ett sorts immateriellt vilda västern. En ny ”sinnenas civilisation”, som John Perry Barlow uttryckte det,¹³⁶ där ord och idéer kunde strömma fritt, och dit statligt inflytande aldrig riktigt skulle nå. Det tankegodset får idag anses förpassat till det historiska kuriosakabinettet. Sedan en tid tillbaka besitter ett antal multinationella plattformsföretag en närmast oöverträffad makt som grindvakter i informationssamhället. De större plattformarnas innehållsmoderering påverkar nu fler människor än vad några av världens största stater förmår – men ligger också utom räckhåll för de bindande folkrättsliga och konstitutionella yttrandefrihetsstandarder som demokratiska stater i EU och USA åtagit sig att respektera. Samtidigt har lagstiftare runtom i världen, både i USA och Europa, tagit aktiva steg för att uppnå ökad kontroll över plattformarnas innehållsmoderering. Med lagstiftningsinitiativ

¹³³ Se artikel 21 i förordningen.

¹³⁴ I skäl 39 i förordningens preambel anges att förordningen inte bör ”hindra de relevanta nationella rättsliga eller administrativa myndigheterna från att, på grundval av tillämplig unionsrätt eller nationell rätt, utfärda ett föreläggande om att återställa innehåll, om sådant innehåll överensstämde med de allmänna villkoren för leverantören av förmedlingstjänsten, men av den leverantören felaktigt har betraktats som olagligt och avlägsnats”.

¹³⁵ I ett antal avgöranden från nederländska distriktsdomstolar under senaste tiden har användare väckt talan mot plattformslieferantörer som Meta, LinkedIn och Youtube med yrkanden om återställande av innehåll som raderats på grund av oförenlighet med plattformens innehållsriktlinjer. I samtliga fall har talan grundats på en direkt eller indirekt tillämpning av artikel 10 i Europa-konventionen – men ingen har vunnit framgång. Se Quintais, Appleman & Fahy (2022), s 22.

¹³⁶ Se ovan not 5.

som FOSTA/SESTA i USA eller TERREG och DSA i EU kan statliga aktörer nu sägas ha skapat ett slags ”moderering av modereringen”.

Utvecklingen är begriplig och till stor del befogad, men väcker samtidigt en del frågor. Inte minst om vem som ska tillskrivas ansvar för plattformarnas moderering, vem som ska säkerställa att ingripanden är förenliga med användarnas yttrandefrihet, och, i förlängningen, hur det ska säkerställas att ”modereringen av modereringen” inte övergår till ”utkontraktering av censur”.

En del plattformsleverantörer har på senare tid själva agerat för att förbättra användarnas möjligheter att få till en saklig prövning av olika modereringsbeslut, med beaktande av användarnas yttrandefrihet. Metas Tillsynsråd, som givits rätt att åsidosätta beslut om att radera innehåll från Facebook, är ett exempel. Med DSA lär också fler prövningsmöjligheter följa: såväl interna hos plattformsleverantörerna som externa hos oberoende tvistlösningsorgan. Men de verkligt komplicerade och principmättade frågorna – om förhållandet mellan statligt inflytande, statens positiva förpliktelser, och användarnas tillgång till effektiva rättsmedel – framstår alltså som obesvarade.



