




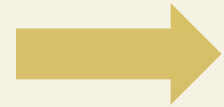


Retsstatslige og demokratiske udfordringer ved forslag til lov om regulering af sociale medier

Jacob Mchangama
Direktør Justitia
27.04.2022

 justitia-int.org
 facebook.com/justitia.dk
 linkedin.com/company/justitia

Lovforslaget udfordrer retssikkerheden



Ytringsfrihed er helt central i en moderne demokratisk retsstat, og bør derfor *altid* som udgangspunkt beskyttes og begrænses mindst muligt



Lovgivning bør baseres på fakta. Der er ingen evidens for, at ”ulovligt indhold flyder frit på de sociale medier” - tværtimod



Loven skaber incitament til overimplementering og overregulering og dermed fjernelse af meget lovligt indhold



Loven giver - stik imod hensigten – mere magt til de i forvejen magtfulde

L 146 Lov om regulering af sociale medier



§ 1. Loven har til formål at fremme beskyttelsen mod ulovligt indhold og øge gennemsigtigheden om indholdsmoderation på kommunikationsplatforme og gruppebeskedtjenester.

Bemærkningerne

1. Indledning

Dette lovforslag tager afsæt i regeringens udspil om sociale mediers ansvar, der blev lanceret 15. august 2021. Lovforslaget er en udmøntning af udspillet 1. initiativ om ”Skærpet fokus på sociale mediers ansvar i forhold til ulovligt indhold”, som bl.a. indeholder et initiativ om forslag til en ny lov, der regulerer sociale medier med skærpede krav om nedtagning af ulovligt indhold og indførelse af transparenskrav.

Ministeren (Simon Kollerup S)

Det lovforslag, vi har til behandling i dag, skaber så **bedre muligheder for at kunne regulere området**, og det vil bl.a. betyde, at ulovligt indhold – det kunne være børnepornografi – skal fjernes eller blokeres inden for 24 timer, og tiden gør en forskel. Hvis mediet ikke opfylder lovkravet, vil man blive straffet med en bøde på op til 6 pct. af virksomhedens koncernomsætning, og det er en stor bøde. Den her bøde vil blive justeret efter antallet af danske brugere, men bøden må og skal være så stor, at den kan mærkes. Lovforslaget skaber også gennemsigtighed, og det giver også brugerne bedre muligheder for at anmelde ulovligt indhold. På den måde laver vi med lovforslaget her en bedre beskyttelse – en **bedre beskyttelse mod ulovligt indhold på nettet**. Og også selv om forslaget her, og det anerkender jeg også, ikke løser alle udfordringer, er det et vigtigt **skridt i den rigtige retning**.

Meningen med det her lovforslag er, at når der er ulovligt indhold, skal det væk, og det skal gå hurtigere, end det sker i dag. Det er utilfredsstillende, og det er en **dårlig oplevelse, når ulovligt indhold ikke fjernes**.

L 146 Lov om regulering af sociale medier

Ordførertale – Malte Larsen (S)



“Desværre ser vi også, at der nu lægges masser af ulovligt indhold op på de sociale medier – indhold, der kan være sårende eller krænkende, indhold, der udleverer den enkelte eller **krænker samfundets almindelige spilleregler for respektfuldhed.**”

“...for mig vil det være en styrkelse, at man mere trygt kan begynde at ytre sig på internettet, når man ikke har den samme risiko for at få, lad os kalde det en **skidtspand hældt i hovedet**, hvis man kommer til at udtrykke sig på de sociale medier.”

Lovgivning bør baseres på fakta - Ingen evidens for at:



”ulovligt indhold flyder frit på de sociale medier”

- ”Mediernes sociale ansvar”. Regeringsudspil aug. 2022

- Justitia analyse fra 2020 (5 udvalgte nyhedsmediers profiler) viste, at kun 1,1 procent af slettet materiale viste sig at være ulovligt.
- Justitia analyse fra 2022 (et udvalg af politikeres og mediers profiler) bekræfter tendensen og finder, at kun 0,0066 procent af kommentarer er ulovlige. Det svarer til én ud af 15.000 kommentarer.
- Algoritmerne, som modererer indholdet, kan ikke kontekstvurdere – borgere kommer sandsynligvis til at gå glip af væsentlig information, som fjernes ”for en sikkerheds skyld”.

Tryk



5,2% Sproglige angreb*

Lidt mere end hvert **20.** indlæg, der står tilbage på mediers og politikeres Facebooksider er et sprogligt angreb, dvs. en stigmatiserende, nedsættende, krænkende, stereotypiserende, ekskluderende, chikanerende eller truende ytring. Det svarer til **3,3 mio.** ud af 63 mio. kommentarer siden 2019.



3,8% anstødelige 1,4 % hadefulde

Ud af de 3,3 mio. sproglige angreb er **902.029** hadefulde, svarende til 1,4 pct. af den samlede debat. De resterende **2.386.535** angrebskommentarer klassificerer algoritmen som anstødelige. Det svarer til 3,8 pct. af den samlede debat.



7,8% angreb hos politikere

Tonen i debatten er hårdere på politikersiderne end på mediesiderne. Topscoreren hos politikere har **17%** angreb i kommentarsporet på sin side. Der er forskel på, om politikere fortrinsvist er værter for angreb mod andre eller selv bliver udsat for angreb på deres side.



4,8% angreb hos medierne

Tonen i debatten på mediesiderne er mindre hård end på politikersiderne, men enkelte medier har en særligt høj andel angreb i deres kommentarspor. Topscoreren hos medierne har således **24%** angreb i kommentarsporet på siden.

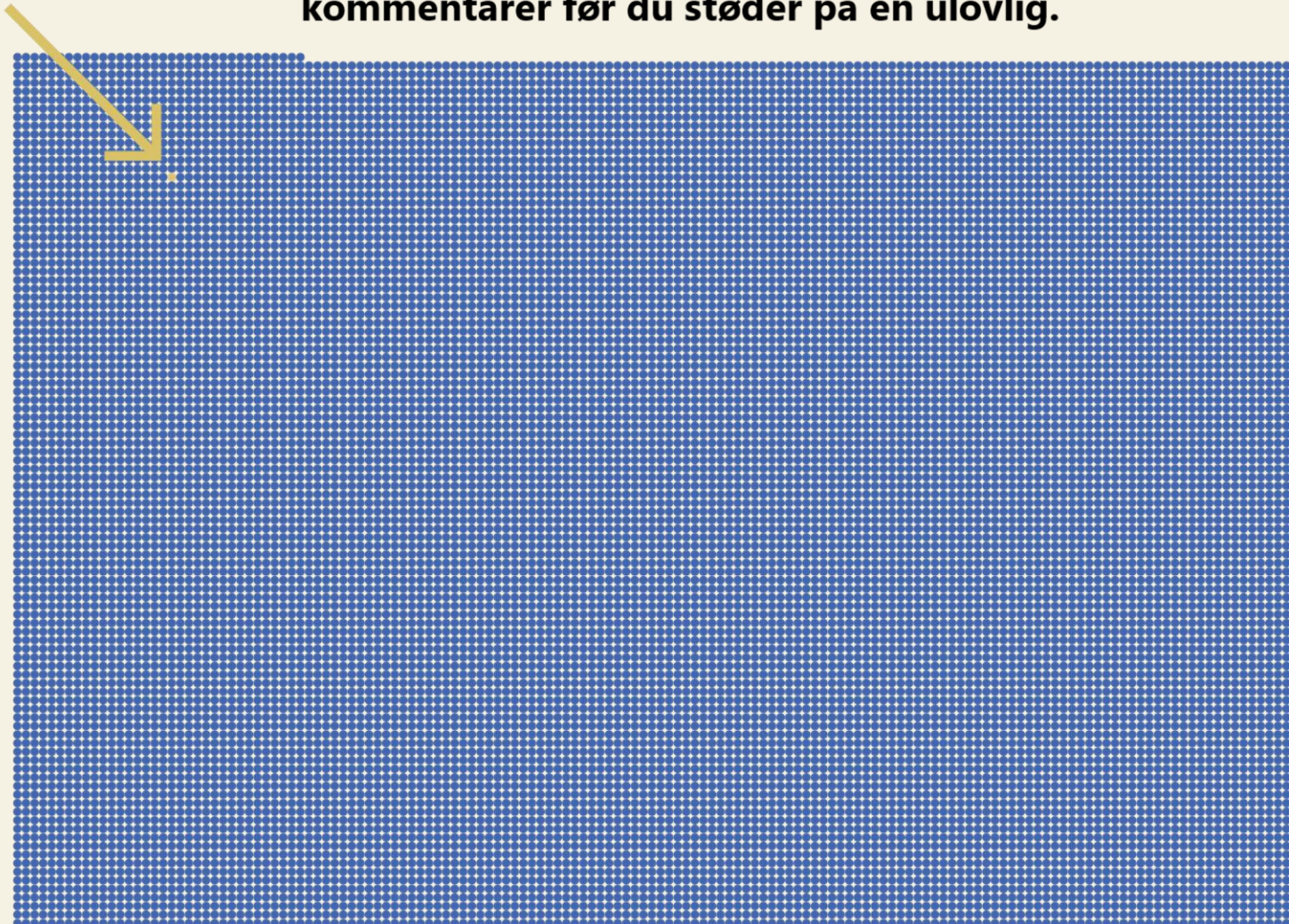


*Rapportens kernebegreber angreb, anstødelighed og hadtale er nærmere defineret i kapitel 1, side 16-20

0,0066 %

af de i alt 63 mio. kommentarer anslås til at udgøre strafbart hadefule angreb.

**Du skal altså læse over 15.000
kommentarer før du støder på én ulovlig.**



Justitia rapport fra maj 2020

Rapport:

DIGITAL YTRINGSFRIHED OG SOCIALE MEDIER - en undersøgelse af hvilke kommentarer der bliver slettet fra fem danske nyhedsmediers facebooksider

Rapporten vurderer, at

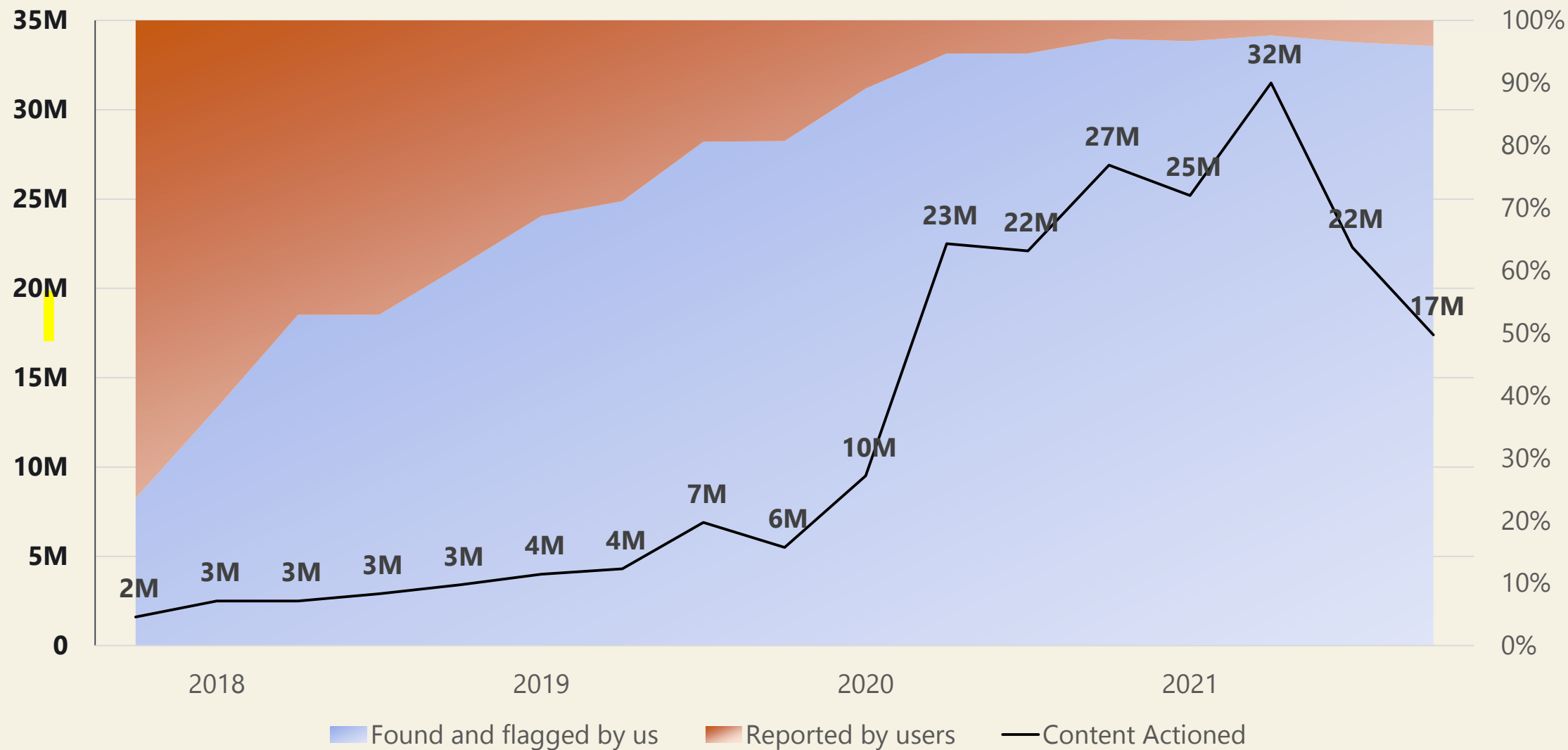
- 1,1% af de kommentarer, som bliver fjernet, er strafbare efter dansk lov. Det svarer til, at der årligt slettes 5.400 strafbare kommentarer fra de fem facebooksider i undersøgelsen.
- 6,0% af de kommentarer, som bliver slettet, kan karakteriseres som racistiske eller truende, men uden at der er tale om strafbare ytringer.
- 24,9% af de kommentarer, som bliver slettet, kan karakteriseres som nedladende eller groft nedladende.
- 39,7% af de kommentarer, som bliver slettet, kan karakteriseres som generelle holdningsytringer, der hverken er strafbare eller er vurderet til at kunne beskrives som racistiske, truende eller nedladende.

Det betyder, at for hver gang der slettes én strafbar kommentar, slettes der cirka 36 generelle holdningsytringer. Grundet det meget store antal kommentarer på Facebook generelt er det rapportens vurdering, at der årligt bliver slettet 366.479 lovlige kommentarer på de fem sider, heraf 213.191 generelle holdningsytringer om eks. politik og samfundsforhold.



Facebook reagerer på regulerende tiltag og fjerner markant mere indhold

(Facebook Transparency Report)



Hadefuld retorik

Beskyttede karakteristika	Facebook	YouTube	Straffelovens § 266 b
Race/hudfarve	✓	✓	✓
National oprindelse	✓	✓	✓
Etnisk oprindelse	✓	✓	✓
Religion	✓	✓	✓
Handicap	✓	✓	✓
Seksuel orientering	✓	✓	✓
Køn, kønsidentitet, -udtryk, -karakteristika.	✓	✓	✓
Alder (hvis refereret til med anden beskyttet karakteristika)	✓	✓	–
Kaste	✓	✓	–
Immigranter/immigrant status	✓	✓	–
Flygtninge	✓	–	–
Migranter	✓	–	–
Asylansøgere	✓	–	–
Alvorlig sygdom	✓	–	–
Beskæftigelse (hvis refereret til med anden beskyttet karakteristika)	✓	–	–
Ofre for en stor voldelig begivenhed + pårørende	–	✓	–
Veteran status	–	✓	–

Eks. På hadefuld retorik fra Facebook og YouTube's fællesskabsregler

Udtryk om at være bedre/dårligere end andre, herunder, men ikke begrænset til:
"Jeg mener, at mænd er kvinder overlegne."

Ord/sætninger, der opfordrer til seksuel aktivitet eller til kontakt med genitalier, afføring etc., herunder, men ikke begrænset til, "kys mig i røven", "æd lort".

Udtalelser om adskillelse, eller udsagn, der fortaler for eller støtter adskillelse.
Udtalelser om udelukkelse, eller udsagn, der fortaler for eller støtter udelukkelse.

Nægte at en veldokumenteret voldelig begivenhed har fundet sted.

Blasfemiske udtryk/ord, som bruges med henblik på at fornærme, herunder, men ikke begrænset til, "fuck, bitch, motherfucker".

Omtalen "den" i forbindelse med transkønnede eller ikke-binære personer.

Udtryk om at afvige fra normen, herunder, men ikke begrænset til, "freak, original".

Indhold om karaktertræk, der kulturelt opfattes som negative, herunder, men ikke begrænset til, "kujon, løgner, arrogant, ignorant".

Facebook sletter konspiratoriske og falske opslag om corona: »Det er censur«

Sociale medier som Facebook spiller en aktiv rolle i kampen mod misinformation om corona. 49-årige Per Brændgaard er en af de brugere, som har oplevet at blive blokeret og få slettet indhold. Den aktive moderation risikerer at føre til radikaliserings og mere splittelse, advarer eksperter

Historiker lagde billede op fra Istedgade:
Så blev han udelukket fra Facebook

Folketingskandidat skrev oikofobi og så fik han 30
dages karantæne på Facebook

**Torben Sangild er blevet smidt af Facebook
uden nogen forklaring**

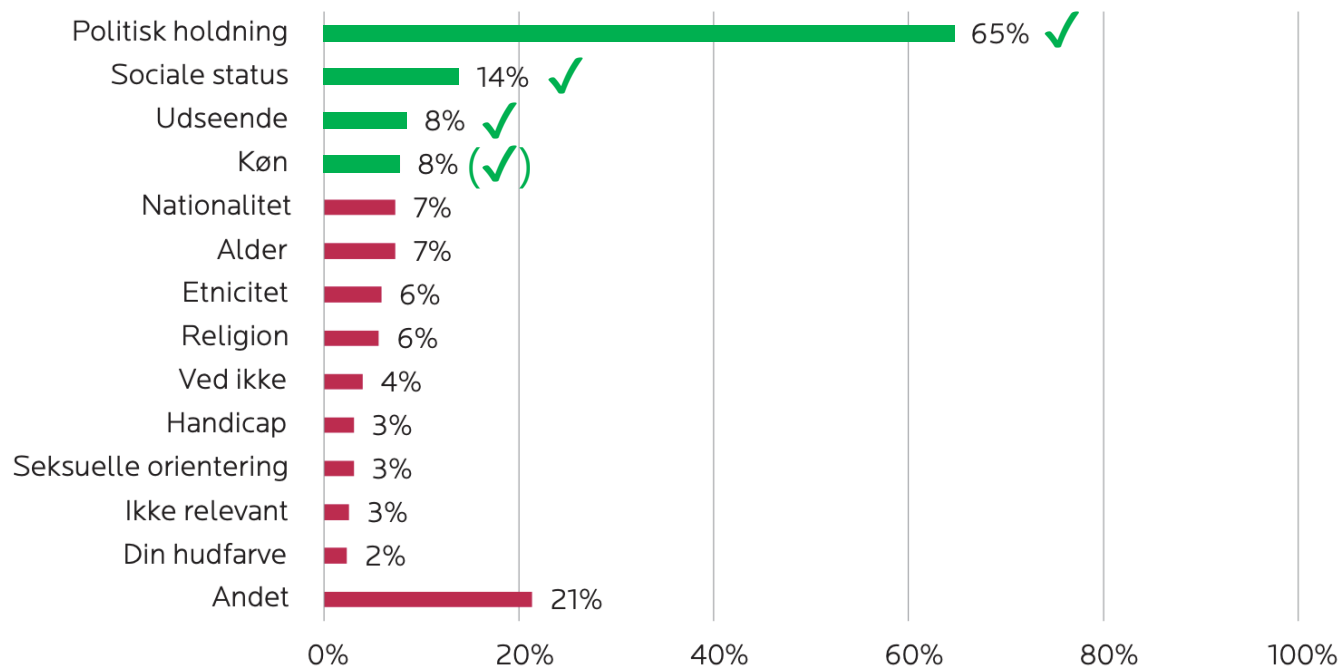
**Hvis du kun skriver »Tommy
Robinson» på Facebook, bliver dit
opslag fjernet**

**Instagram sletter
folketingsmedlems
regeringskritiske billede**

Flest oplever at blive krænket af *lovligt* (IMR 2019)

FIGUR 40: 426 respondenter. Spm. 9: "Hvis du tænker på sidste gang, du oplevede en nedsættende eller krænkende kommentar rettet mod dig på Facebook, hvad gik kommentaren så på? Du kan markere flere svar.". Andelen af alle respondenter, der deltager aktivt i samfundsdebatter på Facebook, der har oplevet at modtage nedsættende eller krænkende kommentarer rettet mod sig på Facebook og hvad de var rettet imod.

Hvad var den nedsættende eller krænkende kommentar rettet mod?



Loven skaber incitament til overregulering, og kan – stik imod hensigten – give mere magt til de i forvejen magtfulde



- Korte reaktionstider og store bøder giver de kommercielle techvirksomheder incitament til pr. automatik at overregulere og fjerne langt mere indhold, end loven kræver. Det giver techvirksomhederne stor magt over den offentlige debat i Danmark.
- Det forhold har magtfulde personer og organisationer omvendt mulighed for at udnytte til at klage over ethvert kritisk indlæg på deres profiler – indlæg som i mange tilfælde vil blive fjernet på grund af korte behandlingsfrister og risiko for store bøder.
- En stadigt større del af den politiske debat i Danmark er rykket over på de sociale medier . Med mere regulering følger derfor stor risiko for en kortslutning af den demokratiske debat og et retssikkerhedstab for borgerne.

Gennemsnitlige sagsbehandlingstider i racismesager – techgiganterne får, med 24 timer behandlingsfrist, incitament til overregulering

The nature of the available data does not allow direct and exact comparisons between the different jurisdictions studied in the report. Still, even when allowing for this shortcoming, all the surveyed domestic legal authorities took significantly longer than the time mandated for social media platforms to answer the question of whether the relevant content was lawful or not. As compared to the short time frames granted to platforms – ranging from hours to a week, the average figures for authorities are as follows:

				
Austria	Denmark	France	Germany	The United Kingdom
1273.5 days (ECtHR)	601 days (National authorities)	420.91 days (ECtHR)	678.8 days (ECtHR)	35.01 days (National authorities)
	1341 days (ECtHR)			393 days (ECtHR)

Overall, data extracted from all European Court of Human Rights' hate speech cases reveals that domestic legal authorities took 778.47 days, on average, from the date of the alleged offending speech until the conclusion of the trial at first instance.

Frankrigs Forfatningsråd

The court said this created an incentive for risk-averse platforms to indiscriminately remove flagged content, whether or not it was clearly hate speech.

The law's provisions "therefore infringe upon the exercise of freedom of expression and communication in a way that is not necessary, suitable, and proportionate," the court said in a [statement](#).

Ytringsfrihedskommissionen

Kommissionen finder ikke, at der – ligesom det har været tilfældet i f.eks. Tyskland – bør gennemføres lovgivning, der med trussel om høje bødestraffe pålægger de (største) sociale medier en pligt til hurtigt – og inden for faste tidsfrister – at fjerne eller blokere ulovligt indhold, efter at have modtaget en anmeldelse herom. I stedet bør *status quo* opretholdes, så de sociale mediers ansvar for indhold alene kan indtræde efter en konkret vurdering af den udviste passivitet i det enkelte tilfælde.

FNs Menneskerettighedskomite

Freedom of expression

46. Whilst appreciating the commitment of the State party to addressing online hate speech and abuse, the Committee is concerned by the broad powers introduced by the Network Enforcement Act in 2017 to remove content deemed illegal or abusive. It also notes with concern that responsibility for the removal of such content is assigned to social media companies and not subject to judicial oversight, thereby limiting access to redress in cases where the nature of content is disputed. The Committee is concerned that these provisions and their application could have a chilling effect on online expression (art.19).

Algoritmernes manglende evne til at kontekstvurdere indhold: Vil Facebook i fremtiden fjerne ubehageligt men væsentligt indhold?

Ville retfærdigheden være sket fyldest for George Floyd, hvis den voldsomt ubehagelige borgervideo ikke var gået viralt?



Ville vi forstå alvoren og konsekvenserne af flygtningestrømmene uden det forfærdelige billede af den lille dreng i vandkanten?

Eksempler

Kvinde anholdt for at bruge foto på Facebook af brændende statsministerdukke

Østjyllands Politi har sigtet 29-årig kvinde for at benytte foto af brændende dukke med statsminister Mette Frederiksens (S) billede i opslag på Facebook

▶ AUTOMATISK OPLÆSNING



Justitia foreslår

- ✓ I bedste fald tilbagerulle forslaget – alternativt begrænse omfanget til særligt alvorlige forbrydelser i straffeloven og at tidsfristen gøres mere fleksibel end de absolutte frister på 24 timer eller 7 dage
- ✓ Tilføre politi- og anklagemyndighed ressourcer til at efterforske og retsforfølge de der står bag det grove og mest alvorlige ulovlige indhold
- ✓ Tilskynde teknologiske løsninger som “Distributed Content Moderation”, hvor brugerne i højere grad kan filtrere indhold.