Forsknings-, Uddannelses- og Digitaliseringsministeriet

Ugennemsigtige rapporter peger på huller i techgiganters indholdsmoderation



Brief om rapporten 'Et glansbillede af gennemsigtighed'

Center for sociale medier, tech og demokrati

August 2026

Resumé

Techgiganternes indholdsmoderation er afgørende for den demokratiske samtale online, men de er, på trods af voksende krav fra EU, langtfra gennemsigtige med, hvordan de forvalter den magt. Det begrænsede indblik, som techgiganters gennemsigtighedsrapporter giver, viser en indsats præget af ulige prioritering af EU's sprog og store forskelle og faldgruber, hvor platformene kommer til kort.

Ved at gennemgå de otte største sociale mediers gennemsigtighedsrapporter viser en ny undersøgelse fra Center for sociale medier, tech og demokrati, at platformene (stadig) underprioriterer det danske sprog blandt deres indholdsmoderatorer, og at de på europæisk plan prioriterer de store vesteuropæiske sprog over de nord- og østeuropæiske. De fremlagte tal indikerer også, at flere techgiganter kun i lav grad tager handling på de anmeldelser om ulovligt indhold, de modtager — både fra almindelige brugere og fra de pålidelige indberettere, som er udpeget til at spotte ulovligt indhold.

Men undersøgelsen gør det også tydeligt, at der på trods af nye standardiserede krav til techgiganternes rapporter fortsat er langt til fuld og reel gennemsigtighed med, hvordan techgiganterne modererer indhold på deres platforme og dermed sætter rammerne for den offentlige samtale online. En række gennemsigtighedsproblemer, såsom manglende ensartethed og kontekst, og at store dele af techgiganternes moderationsindgreb opgøres i uklare "andet"-kategorier, spænder fortsat ben for, at rapporterne kan give et reelt indblik i platformenes maskinrum og sikre demokratisk kontrol med de usynlige hænder, der former vores offentlige samtale.

Indledning

I takt med, at den offentlige samtale i stigende grad foregår på en håndfuld sociale medier, er techgiganternes egne retningslinjer, algoritmer og indholdsmoderation blevet afgørende for, hvilke synspunkter der kan komme til orde, og hvilket indhold brugerne præsenteres for. Når private virksomheder har så stor indflydelse på den offentlige samtale, er der brug for gennemsigtighed med, hvordan de forvalter den — hvordan de sikrer en god tone, beskytter brugerne og holder platformene fri for skadeligt og ulovligt indhold.

Det er netop den demokratiske kontrol, der er tanken bag EU's krav i Digital Services Act (DSA) om, at techgiganterne to gange årligt skal udgive gennemsigtighedsrapporter, som giver offentligheden indblik i deres indholdsmoderation.

Boks 1
Sådan har vi gjort

Undersøgelsen bygger på otte store sociale mediers egne gennemsigtighedsrapporter, som de under DSA er forpligtet til at udgive to gange årligt. Den dækker afrapporteringsperioden 1. juli – 31. december 2025 og følger op på Center for sociale medier, tech og demokratis analyse fra 2024 *Er der et menneske til stede?* Da tallene er techgiganternes egne, giver undersøgelsen udelukkende indblik i deres selvrapporterede indholdsmoderation. Undersøgelsens fulde resultater og metode findes i rapporten.

Læs hele rapporten her:

Undersøgte sociale medier

Undersøgelsens hovedresultater

Tre ud af otte sociale medier har ingen danske moderatorer

Techgiganterne skal i deres gennemsigtighedsrapporter oplyse, hvor mange af EU's officielle sprog deres indholdsmoderatorer taler. Tallene afslører betydelige forskelle i, hvordan platformene prioriterer det danske sprog.

Tre af de otte undersøgte sociale medier — LinkedIn, X og Pinterest — har ingen dansktalende moderatorer til at håndtere dansksproget indhold. Blandt de øvrige er antallet lavt, når man tager højde for platformenes høje brugertal. Facebook og Instagram, der deles om moderatorerne, har i Danmark kun hvad der svarer til seks dansktalende moderatorer pr. million danske brugere, mens YouTube har tre.

Tabel 1
Udvikling i antallet af techgiganternes dansktalende indholdsmoderatorer siden sidst

	Dansktalende moderatorer		Moderatorer pr. en mio. aktive brugere i DK	
Sociale medier	Juli - december 2025	Udvikling siden 2023	Juli - december 2025	Udvikling siden 2023
Snapchat	56	+41	21	+15,2
Facebook	25	±0	6	±0
Instagram	25	±0	6	-0,3
YouTube	24	+6	3	-0,5
TikTok	11	-16	7	-12,0
Pinterest	0	-1	0	-0,8
LinkedIn	0	0	0	0
X	0	0	0	0

Techgiganternes indholdsmoderatorer spiller en central rolle i håndhævelsen af platformenes regler og fjernelse af ulovligt indhold, da de er med til at håndtere anmeldelser, kan gribe ind ved tvivlsspørgsmål og er med til at træne platformenes automatiserede moderationssystemer. Derfor er det vigtigt, at platformene har moderatorer, der forstår det sprog, indholdet er skrevet på.

Udviklingen går flere steder den forkerte vej. TikTok har mere end halveret sit samlede antal dansktalende moderatorer siden 2023, fra 27 til 11. Det er særligt bekymrende, da børn og unge under 18 år anslås at stå for 43 pct. af eksponeringstiden på platformen i Danmark. Snapchat er den eneste af de otte platforme, der har øget antallet af dansktalende moderatorer pr. bruger siden 2023.

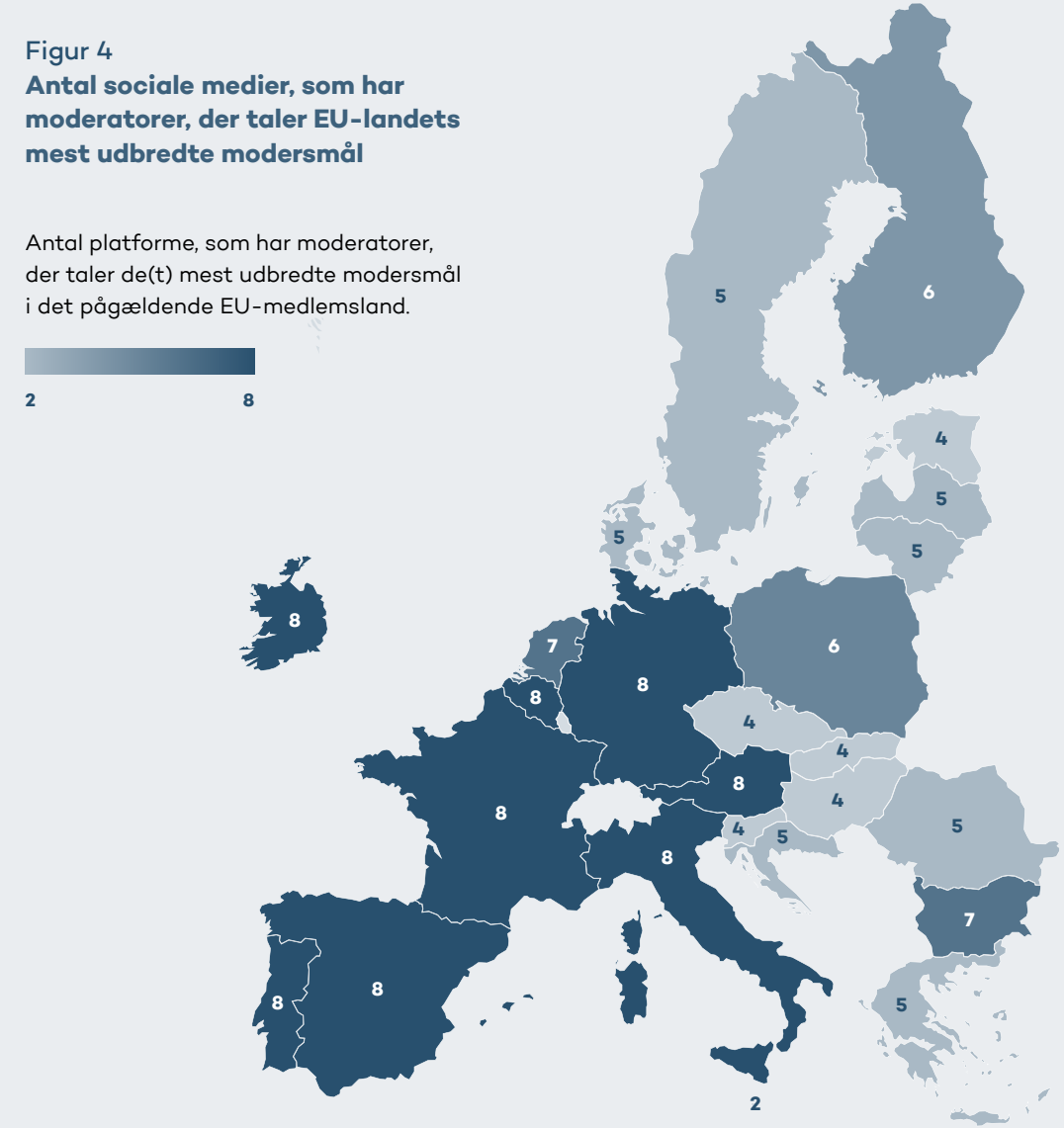
Nordiske og østeuropæiske sprog taber

Det danske sprog er ikke et særtilfælde. Ingen af de otte platforme har moderatorer, der taler alle EU's 24 officielle sprog, og de store vest- og sydeuropæiske sprog prioriteres systematisk højere end de nordiske og østeuropæiske. Mens alle otte platforme har moderatorer, der taler de store vesteuropæiske sprog som tysk, fransk, spansk og engelsk, har blot omkring fem platforme moderatorer på dansk, svensk, de baltiske og de fleste østeuropæiske sprog.

Figur 4

Antal sociale medier, som har moderatorer, der taler EU-landets mest udbredte modersmål

Antal platforme, som har moderatorer, der taler de(t) mest udbredte modersmål i det pågældende EU-medlemsland.





Store forskelle i platformenes egen moderation – og lav gennemsigtighed med, hvad de fjerner

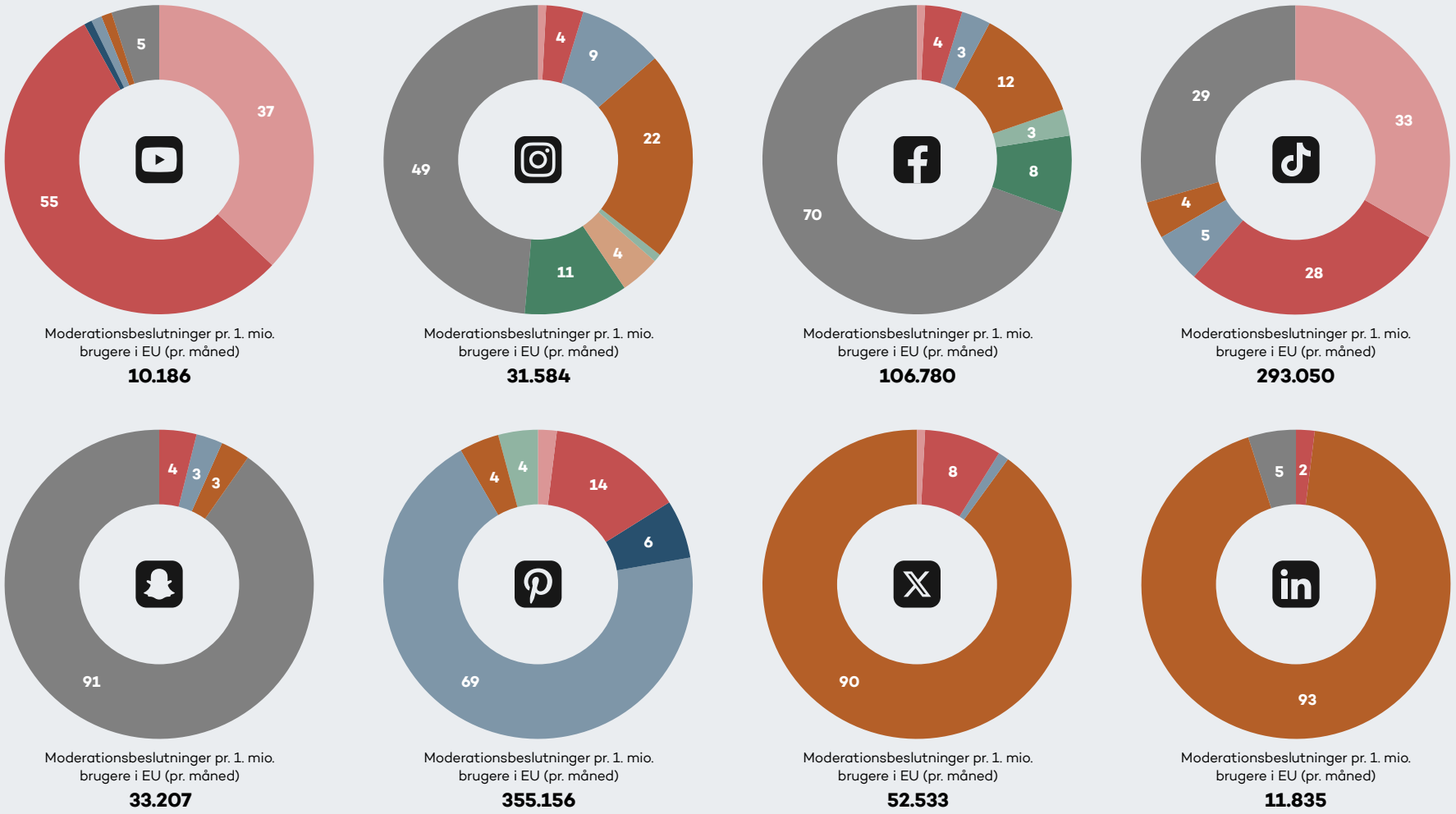
Techgiganterne skal rapportere, hvor meget og hvilket indhold de modererer på eget initiativ — altså det indhold, som platformenes egne systemer og moderatorer selv finder frem til og griber ind over for, uden det er anmeldt af brugere eller myndigheder. Her afslører rapporterne betydelige forskelle.

Set i forhold til deres størrelse modererer YouTube og LinkedIn mindst, med omkring 10.000 og 12.000 moderationsindgreb pr. million brugere i EU om måneden, mens TikTok og Pinterest ligger i den anden ende med henholdsvis 293.000 og 355.000 — altså op mod 35 gange så mange indgreb pr. bruger. Techgiganterne kan moderere ved at fjerne indhold eller brugere helt fra platformene, men kan også dække over blødere tiltag, som visningsfiltre eller at mindske algoritmisk anbefaling af indhold.

Techgiganterne opgør deres data forskelligt, hvilket gør det vanskeligt at vurdere platformene efter en fælles målestok — men et endnu større problem er, at flere af dem ikke er gennemsigtige med, hvad de modererer. Snapchat og Metas platforme kategoriserer en stor del af deres moderation i uklare "andet"-kategorier: det gælder 91 pct. af Snapchats moderationsbeslutninger, 70 pct. af Facebooks og 49 pct. af Instagrams. På tværs af de otte platforme havner hvad der svarer til 240 mio. moderationsbeslutninger i afrapporteringsperioden i sådanne uigennemsigtige kategorier. Når så store mængder indhold ikke opgøres under en specifik kategori, bliver det svært at få reelt indblik i techgiganternes indholdsmoderation.

Figur 6

Sociale mediers indholdsmoderation på eget initiativ i EU pr. måned



X modtager flest anmeldelser om ulovligt indhold fra sine brugere

Ud over platformenes egen moderation stiller DSA krav om, at brugerne skal have mulighed for at anmelde ulovligt indhold, de støder på. Disse bruger-anmeldelser giver et indblik i, hvilket indhold brugerne selv oplever som problematisk, og hvor platformenes egen moderation kommer til kort.

Her skiller X sig ud. Set i forhold til brugerantal modtager X flest anmeldelser om ulovligt indhold — omkring 1.000 pr. million brugere om måneden, omtrent dobbelt så mange som TikTok og Facebook, der i absolutte tal modtager flest brugeranmeldelser hver måned.

Anmeldelsernes indhold fordeler sig forskelligt fra platform til platform: på X og TikTok handler den største del om vold, had og ubehageligt indhold, mens anmeldelser om overtrædelser af ophavsret fylder mest hos LinkedIn og YouTube. Også her er det svært at få indblik i Facebook og Instagrams maskinrum, da omkring 40 pct. af anmeldelserne opgøres i uklare "andet" kategorier.

Meta handler sjældnest på anmeldelser om ulovligt indhold

Når brugere anmelder ulovligt indhold, er der stor forskel på, hvor ofte det fører til handling. I den lave ende ligger Metas platforme: ifølge deres egne opgørelser foretager Facebook kun ni moderationshandlinger pr. 100 anmeldte stykker indhold og Instagram 19, mod X's 64 og YouTubes 53. Samme mønster gælder anmeldelser fra de pålidelige indberettere — de uafhængige eksperter, EU har udpeget til at spotte ulovligt indhold. Her handler Facebook på blot tre ud af 100, selvom platformen modtager næsten halvdelen af alle ekspert-anmeldelser. Tallene skal fortolkes med forsigtighed, men de peger på, at anmeldelser på Metas platforme langt sjældnere fører til et indgreb end på de øvrige.

Tabel 5
Anmeldelser fra brugere om ulovligt indhold i EU pr. måned

Sociale medier	Anmeldelser om ulovligt indhold i alt	Anmeldelser pr. en mio. brugere
X	65.545	1.012
TikTok	116.498	653
Facebook	128.600	490
Instagram	57.717	200
YouTube	56.046	117
Pinterest	1.242	15
Snapchat	1.070	11
LinkedIn	285	5

Gennemsigtighed er en forudsætning for demokratisk kontrol

Indholdsmoderation sætter spillereglerne for den offentlige samtale online

Undersøgelsen tegner et billede af en indholdsmoderation, der er ujævn, flere steder mangelfuld og vanskelig at få indblik i. Men hvorfor er det et problem? Fordi techgiganternes moderation ikke blot er en teknisk opgave — den er med til at sætte spillereglerne for den offentlige samtale. Gennem beslutninger om, hvad der fjernes, nedtones eller får lov at stå, er platformene med til at afgøre, hvilke stemmer der bliver hørt, og hvilket indhold brugerne møder. Dårlig og ineffektiv moderation, der eksempelvis lader mis- og desinformation eller hadtale brede sig, kan samtidig trække den demokratiske samtale i en negativ retning.[1]

Techgiganterne skruer ned for moderationen

Techgiganternes indholdsmoderation har løbende været genstand for kritik og debat fra flere sider, hvor en række undersøgelser og interne læk fra techgiganterne selv har vist problemer med moderation af skadeligt og ulovligt indhold. I de senere år har flere af de største platforme ligeledes meldt ud, at de vil skære ned på moderationen. Efter Elon Musks opkøb annoncerede Twitter (nu X) i 2022, at de ville skære markant i deres moderationsindsats.[2] I januar 2025 fulgte Mark Zuckerberg efter og annoncerede, at Meta ville afvikle sit faktatjek-program – i første omgang i USA – og globalt skære i den automatiserede indholdsmoderation.[3] Zuckerberg erkendte samtidig, at ændringen ville betyde, at mere skadeligt indhold ville optræde på platformene.

En ny undersøgelse fra Digitalt Ansvar (juni 2026) viser, at konsekvenserne af de beslutninger nu kan aflæses i tallene.[4] På baggrund af data fra EU's gennemsigtighedsdatabase finder de eksempelvis, at Meta i løbet af 2025 reducerede sin indholdsmoderation på Instagram i EU med 74 pct.

Trods skærpede krav halter gennemsigtigheden fortsat

Gennemsigtighed er forudsætningen for, at techgiganterne kan holdes ansvarlige for, hvordan de modererer indhold på deres platforme, og sætter rammerne for den demokratiske samtale. Det er netop tanken bag DSA's krav om gennemsigtighedsrapporter. Forskere har samtidig advaret om, at rapporterne kan ende som en form for "transparency washing", der får de ellers notorisk lukkede virksomheder til at fremstå mere åbne, end de reelt er.

Undersøgelsen viser, at det trods nye krav fortsat er svært at få et reelt og sammenligneligt indblik i techgiganternes maskinrum — blandt andet fordi opgørelserne mangler ensartethed og kontekst, og store dele af moderationen havner i uklare "andet"-kategorier. Fundene understreger endnu engang, at gennemsigtighed ikke er et flueben, der kan sættes én gang for alle, men en vedvarende opgave for at sikre indblik i de techgiganter, som ved så meget om sine brugere, men som borgere, myndigheder og forskere stadig ved så lidt om.

1 Center for sociale medier, tech og demokrati, Digital democracy centre & Centre for digital citizenship (2025): Indeks for det danske digitale demokrati 2024.
2 TjekDet (2024): Moderatorerne er fyret: Nu skal brugerne kontrollere brugerne på X.
3 Meta (2025): More Speech and Fewer Mistakes.
4 Digitalt Ansvar (2026): Et kig ind i den sorte boks – en analyse af sociale mediers begrænsning af skadeligt og ulovligt indhold i 2025.