

De globale plattformenes etterlevelse av bransjenormen mot desinformasjon

Vurdering av grunnlagsrapporter og statusrapporter fra Meta, TikTok, Google, Microsoft, Twitch og Twitter



Hovedpunkter

- Bransjenormen mot desinformasjon (The Code of Practice on Disinformation)¹ er en frivillig ordning der underskriverne forplikter seg til å begrense desinformasjon, og rapportere på dette. Bransjenormen ble lansert i 2018, og revidert i 2022 med nye indikatorer og retningslinjer.
- Medietilsynet har på oppdrag fra Kultur- og likestillingsdepartementet gjennomgått grunnlags- og statusrapporter fra Meta, Google, Microsoft, TikTok, Twitter og Twitch publisert i februar og september 2023.
- Rapportene viser at de fleste tilbyderne tar arbeidet mot desinformasjon på alvor, og etterlever bransjenormen i større grad enn før. Det må understrekes at det som er vurdert er etterlevelse av bransjenormen, ikke den reelle effekten av tiltakene mot desinformasjon.
- Utviklingen av kunstig skapt tekst, bilder, lyd, foto og video har ført til at de fleste tilbyderne har revidert retningslinjene sine og innført tiltak for å dempe risikoen for desinformasjon.
- Tilbydere av veldig store internettplattformer og søkemotorer, som Microsoft, Meta, Google, TikTok og Twitter, har plikt til å redusere desinformasjon etter forordningen for digitale tjenester (DSA). Rapporteringen viser at tilbyderne tilpasser seg DSA, og at regulering har effekt.
- Rapportene har imidlertid store mangler, etter Medietilsynets vurdering. Den kvantitative rapporteringen er lite egnet for å måle og sammenligne effekten av tilbydernes tiltak. Den kvalitative rapporteringen er svært omfattende og upresis. Statusrapportene fra september viser imidlertid en positiv utvikling for de fleste tilbyderne, sammenlignet med grunnlagsrapportene i januar.
- Det er stor variasjon i kvaliteten på tilbydernes rapportering. Medietilsynet har rangert de ulike plattformenes etterlevelse av bransjenormen slik, fra best til dårligst: TikTok, Microsoft, Google, Meta, Twitch og Twitter. Medietilsynet vurderer at Metas rapporter er svake. Grunnlagsrapportene fra Twitch og Twitter vurderes som svært svake, i den grad at Medietilsynet mener at de to tilbyderne ikke etterlever bransjenormen.
- Det er store mangler i rapporteringen fra EØS/EFTA-landene. Av de vurderte rapportene har Google, TikTok og Microsoft oppgitt tall fra Norge, Island og Liechtenstein. Meta, Twitter og Twitch har ikke inkludert norske tall.

¹¹ <https://disinfocode.eu/introduction-to-the-code/>

Medietilsynets anbefalinger

- Det bør arbeides systematisk med å forbedre kvalitet på rapportene fra tilbyderne for å sikre bedre presisjon, konsistens og sammenlignbarhet. Rapportene som er vurdert av Medietilsynet som grunnlag for denne rapporten utgjør omlag 1700 sider, der rapportene for TikTok, Meta, Google og Microsoft alene er på over 800 sider. Den kvalitative rapporteringen bør kortes ned, og det bør vurderes å redusere antallet kvantitative indikatorer.
- Vurdering av bransjenormen bør kombineres med egne, selvstendige undersøkelser av effekten av tiltakene mot desinformasjon. Bransjenormen er tilbydernes egenrapportering, som etter Medietilsynets vurdering krever kontroll og verifisering opp mot andre kilder, samt egne analyser og undersøkelser.
- Det bør undersøkes om veldig store internettplattformer og veldig store søkemotorer overholder plikten til å identifisere og redusere systemiske risikoer i DSA knyttet til desinformasjon. Eksempelvis bør forskning på uønsket påvirkning av nasjonale og lokale valg vurderes opp mot disse bestemmelsene.
- Det bør etableres en terskelverdi og tilsynsmetodikk for manglende etterlevelse av bransjenormen, og hvilke mangler som kan utgjøre et brudd på de systemiske risikoene i DSA. Dette vil bidra til klargjøring av hva nasjonale tilsynsmyndigheter bør dokumentere av mulige brudd på DSA, og hva som bør rapporteres til Europakommisjonen, som har håndhevingsansvar for de største plattformene.
- Det må arbeides for at tilbyderne inkluderer EØS/EFTA-landene i rapporteringen. Medietilsynet har meldt fra om at manglende rapportering fra Norge, Island og Liechtenstein til Europakommisjonen og European Regulators Group for Audiovisual Media Services (ERGA). Det anbefales at norske myndigheter og EFTA-sekretariatet jobber for at bransjenormen og andre rammeverk knyttet til systemiske risikoer i DSA, som rammeverket mot hatefulle ytringer, inkluderer tall for EØS/EFTA-landene.
- Det er en reell risiko for at bransjenormen og DSA fører til overmoderasjon på veldig store internettplattformer, og Medietilsynet vurderer det som problematisk at bransjenormen ikke skiller mellom feilinformasjon og desinformasjon. Norge har en høy terskel for forhåndssensur. En politisk føring fra EU om å redusere feil- og desinformasjon, eller blokkering av innhold, på plattformene, kan utfordre informasjon- og ytringsfriheten for norske brukere.
- Kombinasjonen av algoritmisk forsterkning av desinformasjon og kunstig skapt innhold på globale plattformer utfordrer mulighetene til å skille sant fra usant. Det er ikke gitt at redaksjonelle medier, faktasjekk og kritisk medieforståelse er en tilstrekkelig motvekt. Risikodempende tiltak mot algoritmisk og annonsefinansiert forsterkning av desinformasjon bør prioriteres, fremfor tiltak for å blokkere eller slette enkeltinnlegg som kategoriseres som feil- eller desinformasjon, da førstnevnte ikke er et inngrep i ytringsfriheten.

- Bransjenormen bærer preg av egenvurdering fra tilbydernes side, med fare for skjevheter og mangler i rapporteringen. Det anbefales derfor at vurderingen av bransjenormen kombineres med uavhengige analyser fra tilsynene av plattformenes risikodempende tiltak opp mot desinformasjon. Dette for å dokumentere mangler og måle effekten av tiltakene, samt gjøre nærmere undersøkelser av interessante data i bransjenormens rapporter.
- Bransjenormen bør ha en kvantitativ indikator for antall moderatorer for hvert land og hvilke språk de behersker. En krise som kan endre risikobildet, kan oppstå raskt, slik man for eksempel har sett i forbindelse med svensk og finsk NATO-medlemskap. Det bør også etableres en indikator for behandlingen av redaksjonelt innhold, som er en viktig motvekt til desinformasjon, men som det siste året er blitt systematisk nedprioritert av globale plattformer.
- Det bør legges press på plattformene for et utvidet samarbeid med anerkjente faktasjekkorganisasjoner med norsk kontekstforståelse, som faktisk.no. Medietilsynet anbefaler at plattformenes bruk av norske faktasjekkere kartlegges.
- Vurderingen av bransjenormen mot desinformasjon i Norge bør fortsette, særlig opp mot de globale plattformene. Dette er nødvendig for å dokumentere eventuelle brudd på forpliktelsene om å redusere de systemiske risikoene i DSA. Medietilsynet påtar seg gjerne ansvaret for å jevnlig gjennomføre slike kartlegginger og vurderinger. Arbeidet med denne rapporten har vist seg å være omfattende og ressurskrevende, som bør sees opp mot Medietilsynets samlede ressursbehov relatert til nye oppgaver.
- Den nordiske tenketanken for teknologi og demokrati har anbefalt flere områder for nordisk samarbeid og koordinering av reguleringer relatert til globale teknologigiganter, særlig i lys av DSA.² Det er naturlig å etablere et samarbeid om vurdering av bransjenormen, og at dette skjer mellom de nordiske mediemyndighetene.

² <https://pub.norden.org/nord2023-004/nord2023-004.pdf>

Innhold

Innledning	7
1. Om desinformasjon.....	9
2. Om bransjenormen	13
3. Sammenligning av rapportene	15
Kapittel 2: Kontroll av annonser	19
Kapittel 3: Politiske annonser	21
Kapittel 4: Tjenestens integritet	23
Kapittel 5: Styrke befolkningens motstandskraft.....	24
4. Nærmere vurdering av de ulike tilbydernes rapporter.....	28
Meta.....	28
TikTok.....	34
Google	41
Microsoft	50
Twitter	58
Twitch	59

Forkortelser:

API: Application Programming Interface, grensesnitt som gir direkte tilgang til data og funksjonalitet i et datasystem.

CoP eller bransjenormen: Strengthened Code of Practice on Disinformation.

DSA: Digital Services Act (forordningen for digitale tjenester)

EC: European Commission

EDMO: European Digital Media Observatory.

EEA: EØS (EU-landene og Island, Liechtenstein og Norge)

ERGA: European Regulators Group for Audiovisual Media Services (ansvar for monitorering av bransjenormen CoP).

GR: Grunnlagsrapport (baselinerapport) fra februar 2023

KI: Kunstig intelligens

PTF: Permanent Task-Force («arbeidsgruppen»)

QREs: Qualitative Reporting Elements – kvalitativ rapportering

SLIs: Service Level Indicator – kvantitativ rapportering

SR1: Statusrapport 1 (tilbydernes rapportering første halvår 2023, levert september 2023)

TTP: Taktikker, teknikker og prosedyrer for spredning av desinformasjon og påvirkningskampanjer.

VLOP: Very Large Online Platform, jf. DSA³

VLOSE: Very Large Online Search Engine, jf. DSA⁴

³ https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413

⁴ Ibid.

1 Innledning

Medietilsynet har på oppdrag Kultur- og likestillingsdepartementet (KUD) vurdert rapportene fra tilbydere som har signert den reviderte bransjenormen mot desinformasjon – The Strengthened Code of Practice on Disinformation fra 2022. I vurderingen inngår grunnlagsrapporter som tilbyderne leverte i februar 2023 og statusrapporter fra juli, publisert i september. Vurderingen omfatter 6 tilbydere og 11 tjenester:

- Google (YouTube, Google Search og Google Advertising)
- Meta (Facebook og Instagram)⁵
- Microsoft (LinkedIn, Bing Search og Microsoft Advertising)
- ByteDance (TikTok)
- Twitter⁶
- Twitch⁷

Det er viktig å understreke at Medietilsynet har vurdert kvaliteten på rapportene, ikke effekten av tilbyderens tiltak mot desinformasjon. En tilbyders innsats mot desinformasjon kan følgelig være bedre eller dårligere enn hva kvaliteten på rapportene skulle tilsi. Bransjenormen vil bidra til at tilbyderne rapporterer på en ensartet måte, som gjør det mulig å etterprøve og sammenligne plattformenes påståtte effekter av tiltakene, samt å utarbeide *best practices*. For å vurdere den faktiske effekten av tilbyderens innsats mot desinformasjon, bør bransjenormens rapporter etter Medietilsynets vurdering sammenstilles med øvrige analyser.

Bransjenormen inneholder 10 kapitler og 152 indikatorer. De vurderte rapportene er totalt på nærmere 1700 sider, og det store omfanget har gjort det nødvendig med prioriteringer opp mot risiko og relevans i gjennomgangen. Medietilsynet har valgt å legge vekt på kapitlene 2-5 om annonser, politiske annonser, tjenestens integritet og kritisk medieforståelse. Vurderingen av meldingstjenester som WhatsApp og Messenger er valgt bort, til fordel for sosiale medier og søkemotorer, som er de viktigste kanalene for desinformasjon i Norge.

Medietilsynet har rangert rapportene fra 1-3, der 1 er dårlig og markert med rødt, 2 er middels og markert med oransje, og 3 er god og markert med grønt. Medietilsynets vurdering er gjort parallelt med at Kantar, på oppdrag av Europakommisjonen, utarbeider felles metodikk for vurdering av bransjenormen. En slik felles metodikk vil bidra til mer enhetlig rapportering og evaluering av rapportene fremover. EDMO⁸ har publisert sin vurdering av grunnlagsrapportene. ERGA har gjort en tilsvarende vurdering, som ikke er publisert.⁹ Europakommisjonen har også gjort en selvstendig undersøkelse av plattformenes rolle i spredningen av russisk desinformasjon, og har konkludert med at innsatsen var mangelfull i forhold til de systemiske forpliktelsene i DSA, og at det i særdeleshet gjaldt X/Twitter.¹⁰

⁵ Rapporteringen omfatter også Messenger og WhatsApp, men er lite relevant. Medietilsynet har ikke vurdert denne delen av rapportene.

⁶ Twitter har trukket seg fra bransjenormen, og har kun levert grunnlagsrapport.

⁷ Twitch har ikke plikt til å levere statusrapport før januar 2024.

⁸ <https://www.politico.eu/wp-content/uploads/2023/09/05/CoP-Monitor-Report.pdf>

⁹ ERGAs utkast til rapport legges ved, under forutsetning av at den behandles konfidensielt.

¹⁰ <https://op.europa.eu/en/publication-detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-de>

Desinformasjon er et sensitivt område for regulering, og er ikke klart juridisk definert. En stor andel desinformasjon er følgelig lovlig informasjon, der statlige inngrep kan være et urettmessig inngrep i ytringsfriheten. Allikevel kan desinformasjon i ytterste konsekvens utfordre demokratiet. Selv om bransjenormen er en frivillig ordning, har de største globale plattformene etter DSA plikt til å identifisere og redusere den systemiske risikoen for spredning av desinformasjon. Medietilsynet støtter denne systemiske tilnærmingen, samtidig som det er en risiko for at bransjenormen og DSA kan føre til overmoderering på plattformene. Denne problemstillingen behandles i kapittel 1.

Kapittel 1 tar opp noen sentrale prinsipielle problemstillinger knyttet til behandlingen av desinformasjon. Kapittel 2 inneholder en kort oppsummering av bransjenormen. I kapittel 3 sammenlignes tilbydernes innsats opp mot kapitlene 2-5 i bransjenormen, mens kapittel 4 inneholder en detaljert gjennomgang av tilbydernes rapportering, der også kapitlene 2-7 er vurdert. Medietilsynet har utarbeidet grafer for kvantitative indikatorer som er relevante, og som tilbyderne har rapportert på.

2 Om desinformasjon

Desinformasjon, feilinformasjon og propaganda utfordrer befolkningens evne til å foreta informerte valg, og er «en stor utfordring» for demokratiene i Europa, jf. bransjenormens forord, bokstav b. Pandemi og krig i Europa har synliggjort de helsemessige og sikkerhetsmessige risikoene ved feil- og desinformasjon. Russlands systematiske bruk av koordinerte påvirkningsoperasjoner er godt dokumentert. Den systematiske spredningen og forsterkningen av desinformasjon kan i ytterste konsekvens utfordre tilliten til demokratiet og demokratiske institusjoner. Det er derfor behov for å innføre tiltak for å redusere disse risikoene.

Samtidig er de fleste ytringer som kategoriseres som feil- eller desinformasjon lovlige i Norge, og har et konstitusjonelt vern i Grunnlovens paragraf 100. Tiltak mot desinformasjon gir en risiko for statlige inngrep, direkte, eller indirekte, ved at det legges føringer for tilbydernes tiltak, og for at dette fører til overmoderasjon på plattformene. Begge deler kan utfordre ytringsfriheten.

Risikoen ved desinformasjon ble aktualisert i det amerikanske presidentvalget og Brexit-kampanjen i 2016, og ved stormingen av kongressen 6. januar 2021. Diskusjonen om desinformasjon og påvirkningskampanjer ble intensivert etter den russiske invasjonen i Ukraina 24. februar 2022. Det er blitt en mer militær retorikk mot desinformasjon, som at Europa står i en «informasjonskrig» med «desinformasjon som våpen», og at denne kampen skal vinnes.^{11 12} Medietilsynet deler bekymringen for at desinformasjon kan brukes som et middel for å undergrave demokratiet, men vil samtidig påpeke risikoen ved at en slik retorikk kan føre til et «statlig overslag». Det er viktig å understreke at kampen mot desinformasjon ikke må gå på bekostning av den ytringsfriheten man ønsker å beskytte.

Desinformasjon bør så langt det er mulig møtes med demokratiske idealer som ytringsfrihet, og dens begrunnelser om at meningsmangfold fører til sannhet, demokrati og autonomi. Terskelen for forhåndssensur i Grunnloven er svært høy. EUs sanksjoner mot russiske statskontrollerte medier førte til en konstitusjonell vurdering i Norge, som konkluderte med at sanksjonene ikke kunne forsvares i Grunnloven. Norge er følgelig det eneste landet i det indre markedet som ikke har gjennomført disse sanksjonene. Ytringsfriheten er ikke absolutt, men det bør være opp til enkeltland å gjøre inngrep i ytringsfriheten når det er nødvendig, og slike inngrep må kunne forsvares opp mot Grunnloven.

I vurderingen av hvordan desinformasjon bør begrenses, uten inngrep i ytringsfriheten, bør det skilles mellom tre aktører: Mottager, distributør (kanal) og avsender.

Mottager: God kritisk medieforståelse, tilgang til redaktørstyrte medier og faktasjekkere er viktige forutsetninger for å sikre tilstrekkelig motstandsdyktighet på mottagersiden. Ytringsfrihetskommisjonen konkluderte med at Norge er relativt godt rustet på mottagersiden, med et høyt utdanningsnivå, et mangfold av uavhengige redaktørstyrte medier og andre tiltak

¹¹ <https://mediacitybergen.no/home/desinformasjon-og-falske-nyheter-er-en-stor-og-alvorlig-trussel-mot-demokratiet/>

¹² <https://forskning.no/krig-og-fred-media-russland/slik-driver-russiske-myndigheter-informasjonskrig-mot-vesten-og-sitt-egget-folk/1989827>

som har ført til en høy grad av kritisk medieforståelse i befolkningen. Norge er følgelig i en god posisjon for å håndtere desinformasjon.¹³

Kanal: Globale teknologiaktører har de siste 15-20 årene fått enorm makt over distribusjonen av innhold, og eier store deler av informasjonsinfrastrukturen i Norge. Sosiale medier og søkemotorer er også de viktigste kanalene for spredning av desinformasjon. Kombinasjonen av kunstig skapt innhold, bruk av bots og effektive anbefalingssystemer har gitt et nytt risikobilde. Forretningsmodellen kan bidra til en algoritmisk forsterkning av spredning av desinformasjon, siden emosjonelt innhold skaper mer engasjement enn faktabasert informasjon. Det er usikkert om redaksjonelt innhold, faktasjekk og kritisk medieforståelse gir en tilstrekkelig motstandskraft mot KI-generert innhold og forsterket spredning. Det er derfor behov for ytterligere tiltak mot distribusjonen av desinformasjon.

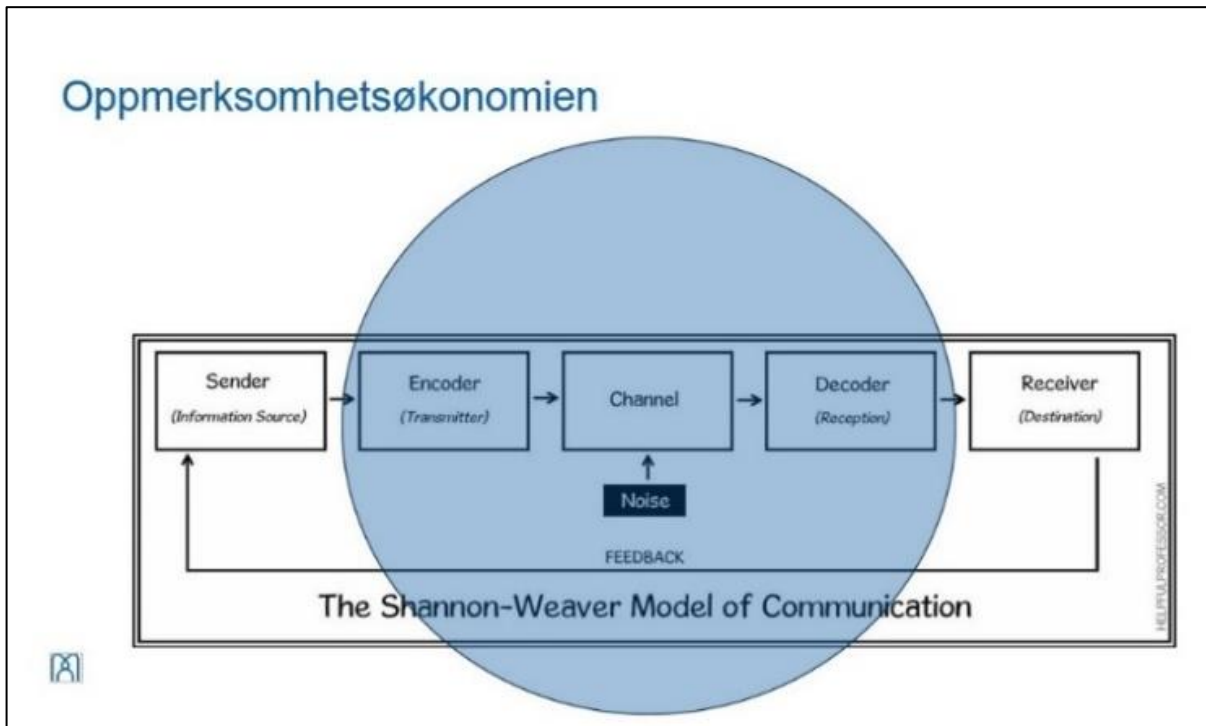
Avsender: Det må først skilles mellom en manuell og automatisert avsender. Ytringsfriheten gjelder mennesker, ikke maskiner. Begrensninger av *boter* som sprer desinformasjon, er ikke et inngrep i ytringsfriheten. Europol har i en rapport vist til at 90 prosent av innholdet på nettet i 2026 kan være kunstig skapt.¹⁴ Bruk av falske og automatiserte kontoer – *boter* – på sosiale medier er utbredt, og kan føre til en betydelig økning av volumet desinformasjon, uten at mottageren er i stand til å avdekke om avsenderen er menneske eller maskin.¹⁵ Det er behov for et strengt regelverk for å begrense innlegg fra automatiserte kontoer eller *boter*. Når det gjelder menneskelige avsendere, omfatter ytringsfriheten retten til å legge ut feilinformasjon eller desinformasjon. Staten bør være forsiktig med å legge føringer som innskrenker enkeltpersoners ytringsfrihet. Et statlig virkemiddel rettet mot avsendersiden kan være å øke forståelse av ytringsansvar, slik kritisk medieforståelse er et virkemiddel for å styrke motstandskraften på mottagersiden.

¹³ NOU 2022: 9, kapittel 7.5.2

¹⁴ Facing reality? Law enforcement and the challenge of deepfakes (Europol, 2022)

¹⁵ <https://www.aftenposten.no/verden/i/e1LWkM/sverige-valget-automatiserte-kontoer-sprer-innvandringskritikk-paa-sosiale-medier>

Figur 1: Shannon-Weavers klassiske modell definerer fem ledd i kommunikasjon, inkludert avsender, mottager og kanal. Globale plattformer er en viktig del av informasjonsinfrastrukturen, og kontrollerer distribusjonskanalene (markert i blått), som i den digitale økonomien omfatter søkemotorer, sosiale medier og applikasjonsbutikker. De globale aktørene er en grunnleggende infrastruktur for informasjon og desinformasjon. Plattformenes forretningsmodell er optimalisert for maksimal tidsbruk, for mest mulig eksponeringstid for annonser. Oppmerksomhetsøkonomien har gitt utilsiktede konsekvenser når det gjelder forsterkningen av engasjerende innhold som feil- og desinformasjon. Medietilsynet mener reguleringer mot algoritmisk og annonsefinansiert forsterkning er viktig, også fordi det er mindre inngripende enn statlige føringer for å slette enkeltinnlegg.



Medietilsynet mener det er behov for reguleringer for å hindre forsterkningen av desinformasjon, som å begrense (eller innføre forbud mot) algoritmisk og annonsefinansiert forsterkning¹⁶. Dette kan bidra til å ivareta yttringsfriheten for den enkelte, og samtidig begrense distribusjon av desinformasjon: *Freedom of Speech* og *Freedom of Reach* er to forskjellige ting. Den systemiske tilnærmingen i DSA er et riktig virkemiddel, som legger vekt på ansvarlig design av den underliggende teknologien. I DSA pålegges de største globale plattformene og søkemotorene å identifisere og redusere risikoene for «enhver faktiske eller forutsigbare negative effekter på den offentlige samtalen og valgprosesser; og offentlig sikkerhet», jf. artikkel 34, nr.1, bokstav c.¹⁷ Dette er grunnlaget for bransjenormen, som inngår som et risikodempende tiltak opp mot denne bestemmelsen.

Algoritmenes innvirkning på et innleggs synlighet er stor. I 2018 anslo YouTube at anbefalingsalgoritmene sto for 70 prosent av visningene på YouTube.¹⁸ En studie fra Massachusetts Institute for Technology (MIT) har vist at falske nyheter og desinformasjon

¹⁶ Algoritmisk forsterkning brukes her som forsterkning av en type innhold på bekostning av annet innhold for økt synlighet eller distribusjon, eksempelvis på bakgrunn av anbefalingssystemer. Annonsefinansiert forsterkning brukes om betalt forsterkning for økt synlighet eller distribusjon.

¹⁷ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065&from=EN>

¹⁸ <https://qz.com/1178125/youtubes-recommendations-drive-70-of-what-we-watch>

spres 70 prosent raskere enn redaksjonelt innhold.¹⁹ At desinformasjon begrenses, ikke forsterkes, kan derfor ha stor effekt. Medietilsynet mener en slik tilnærming er mindre inngripende for ytringsfriheten enn å motivere plattformene til å redusere antall menneskeskapte innlegg definert som feil- og desinformasjon.

Samtidig vil Medietilsynet peke på at bransjenormen har problematiske sider, særlig når det gjelder risikoen for at for mye innhold fjernes av plattformtilbydere. Desinformasjon defineres som «bevisst villedning som spres for å skade» mens feilinformasjon «feilaktig informasjon som spres uten skadehensikt».²⁰ Det finnes imidlertid ingen klar, rettslig definisjon av desinformasjon, og grensen mellom lovlig og ulovlig desinformasjon er diffus.

Bransjenormens bruk av «desinformasjon» omfatter feilinformasjon (misinformation), desinformasjon, påvirkningsoperasjoner og utenlandsk innblanding i den offentlige samtalen, jf. forordets bokstav a. Bransjenormen legger føringer som kan resultere i overmoderasjon av innhold på plattformene.

Det er allerede flere eksempler på at plattformene har blokkert innhold på sviktende grunnlag.²¹ Politikere og partier er blitt utestengt fra plattformene i forkant av en valgkamp.²² Det er følgelig viktig å finne en god balanse mellom nødvendige tiltak for å forhindre manipulering av demokratisk valg, uten å gjøre inngrep i politikeres legitime bruk av sosiale medier for å nå ut til velgerne.

¹⁹ <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>

²⁰ Definisjoner av feilinformasjon og desinformasjon er hentet fra NOU 2022:9, kap. 9.2

²¹ <https://www.theguardian.com/technology/2021/may/27/facebook-lifts-ban-on-posts-claiming-covid-19-was-man-made>

²² <https://www.nrk.no/osloogviken/en-rekke-lokale-ap-sider-utestengt-fra-facebook-1.16526120>

3 Om bransjenormen

[The Strengthened Code of Practice on Disinformation fra 2022](#) – bransjenormen mot desinformasjon – bygger videre på bransjenormen fra 2018, og har som mål å redusere spredning og finansiering av desinformasjon på digitale plattformer, sikre mer åpenhet om plattformenes tiltak, og styrke befolkningens kritiske medieforståelse. Bransjenormen består av 10 kapitler, 44 forpliktelser, 128 tiltak, med rapporteringskrav på totalt 152 kvantitative og kvalitative indikatorer. Bransjenormens kapitler og sentrale forpliktelser:

- 1) Forord.
- 2) Kontroll av annonser (forpliktelse 1–3): Forpliktelser til å stanse annonseinntekter fra desinformasjon og forhindre algoritmisk og annonsefinansiert forsterkning av spredningen av desinformasjon.
- 3) Politiske annonser (forpliktelse 4–13): Forpliktelser om åpenhet om politisk annonsering. Tilbyderne skal verifisere og tydelig merke politisk reklame, jf. definisjonen i forordningen for åpenhet og målretting av politisk reklame, artikkel 2, som er til behandling i EU.²³ Det skal etableres en søkbar database for politisk reklame.
- 4) Tjenestens integritet (forpliktelse 14–16): Inneholder forpliktelser til å redusere manipulerende teknikker for spredning av desinformasjon. Omfatter tiltak mot *bots*, falske kontoer og kunstig skapt innhold, med en kobling til forordningen for kunstig intelligens, som er til behandling i EU.²⁴
- 5) Styrke befolkningens motstandskraft (forpliktelse 17–25): Tiltak for å sette brukerne bedre i stand til å gjenkjenne, forstå og varsle om desinformasjon. Omfatter tiltak for å styrke brukernes kritiske medieforståelse, «sikker design», åpenhet om anbefalingssystemene.
- 6) Styrke forskning om desinformasjon (forpliktelse 26–29): Skal sikre forskere bedre og enklere innsyn i plattformenes data. Omfatter forpliktelser til å gi forskere automatisert tilgang til ikke-sensitive data, jf. DSAs artikkel 40 om datatilgang til «sertifiserte forskere».
- 7) Styrke faktasjekk av desinformasjon (forpliktelse 30–33): Skal legge til rette for faktasjekk-tilbud på tvers av medlemsland og språk og sikre at plattformene gjør mer konsekvent nytte av disse. Inkluderer forpliktelser til samarbeid med faktasjekkere, sikre nødvendig tilgang til data og merking av faktasjekkett innhold.
- 8) Etablering av et transparenscenter²⁵ (forpliktelse 34–36).
- 9) Etablering av en fast arbeidsgruppe (forpliktelse 37).
- 10) Monitorering av bransjenormen (forpliktelse 38–44).²⁶

²³ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0731>

²⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206>

²⁵ <https://disinfocode.eu/>

²⁶ Permanent Task Force: Europakommisjonen, EUs utenriksstjeneste (European External Action Services – EEAS), European Regulators Group for Audiovisual Media Services (ERGA) og European Digital Media Observatory (EDMO).

Etter en evaluering av den første versjonen av bransjenormen, utført i 2020, konkluderte med at det var behov for revideringer.²⁷ Den «forsterkede» bransjenormen inneholder en rekke nye forpliktelser, og har innført 152 indikatorer som tilbyderne skal rapportere på slik at det kan måles effekt over tid. Indikatorene er delt inn i kvalitative indikatorer (Qualitative Reporting Elements - QRE) og kvantitative indikatorer (Service Level Indicators - SLI). En viktig endring i den reviderte bransjenormen er koblingen til DSA, som har betydelig håndhevningskraft. Etterlevelse skal inngå som et risikodempende tiltak i DSA, jf. bransjenormens forord, bokstavene h og i. Bransjenormen refererer også til andre EU-forordninger, som forordning om politisk reklame, jf. kapittel 3, bokstav c.

Europakommisjonen har utarbeidet retningslinjer for å sikre felles forståelse av rapporteringskravene.²⁸ I tillegg ligger det en rettleider i skjemaene plattformene skal fylle ut, som ikke er offentlig tilgjengelig. 44 selskaper har ifølge Europakommisjonen signert bransjenormen,²⁹ og listen oppdateres fortløpende på bransjenormens nettsider.³⁰ Underskrivere inkluderer globale teknologiselskaper som er definert som portvoktere etter Digital Markets Act,³¹ og som tilbyr tjenester som er definert som VLOP/VLOSE etter Digital Services Act,³² der risikoene for en negativ innvirkning på det offentlige ordskiftet og demokratiske valg er størst. VLOP og VLOSE forplikter seg til å rapportere to ganger i året. Øvrige underskrivere rapporterer årlig.

Tilbyderne leverte grunnlagsrapporter i februar 2023, mens statusrapportene for VLOP/VLOSE ble publisert 27. september på transparenscenterets nettsider (disinfocode.eu).

Som det kommer frem av bransjenormens forord, bokstav t, omfatter bransjenormen EØS, og det skal også rapporteres fra de tre EØS/EFTA-landene. I rapportmalene er EØS/EFTA-landene inkludert.³³ Etter Medietilsynets vurdering er det uheldig at ikke alle tilbyderne rapporterer granulerte tall fra Norge, Island og Liechtenstein. I vurderingen har Medietilsynet lagt vekt på om retningslinjene på dette området er fulgt.

²⁷ <https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>

²⁸ <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation>

²⁹ https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_23_4645

³⁰ <https://disinfocode.eu/signatories-archive/>

³¹ https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4328

³² https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413

³³ "(t) The application of this Code is limited for each Signatory to services provided in the States that are Contracting Parties to the European Economic Area."

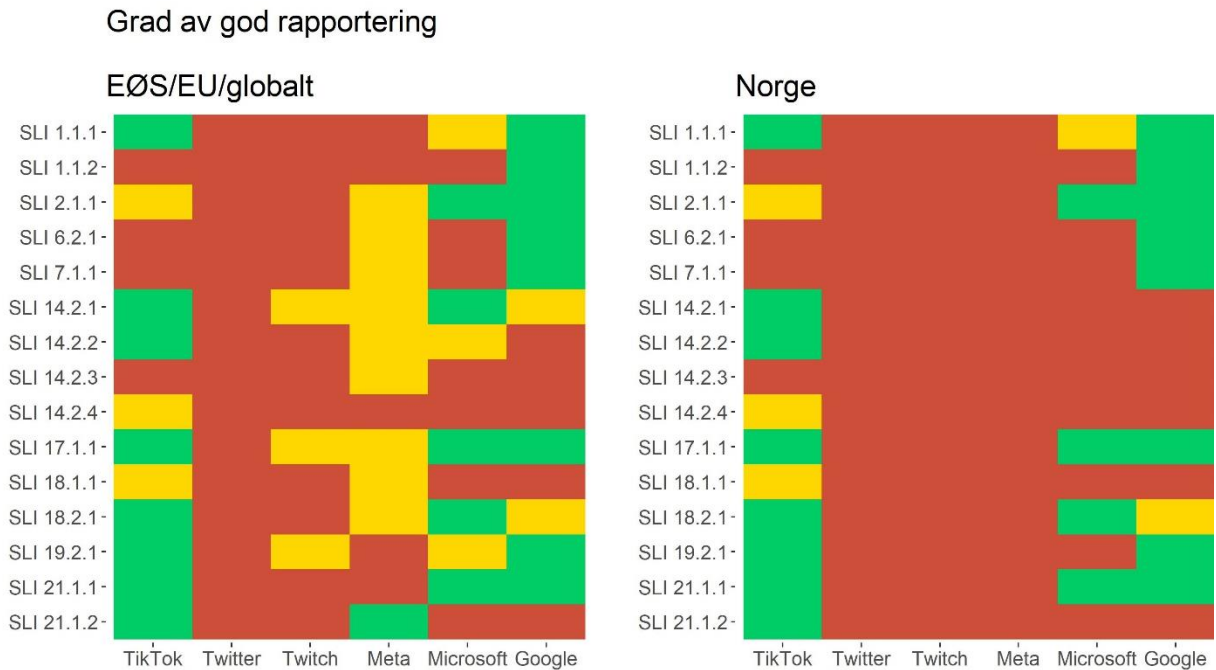
4 Sammenligning av rapportene

I denne delen sammenstilles hovedpunkter fra tilbyderens rapportering for kapitlene 2–5. Det er stor variasjon på kvaliteten i rapportene. Selv om rapporteringen fra noen av tilbyderne langt på vei følger retningslinjene, vurderer Medietilsynet at datakvaliteten er gjennomgående svak. Dette gjør sammenligningen på tvers av tilbyderne, eller en måling av utvikling fra GR til SR1, krevende. Eksempelvis skal rapporteringsperioden for GR være én måned (16.12.22–16.1.23), noe tilbyderne i liten grad har forholdt seg til. For statusrapportene skal tilbyderne rapportere for en periode på seks måneder, som gjør det enklere å sammenligne utviklingen over tid.

Rapporteringen er gjennomgående omfattende og upresis. De kvalitative rapportene er i stor grad hentet fra tilbyderens nettsider, og er i for liten grad tilpasset bransjenormens retningslinjer. Flere rapporter mangler landspesifikke data, selv om bransjenormen, retningslinjene og rapportmalene tydeliggjør at det skal rapporteres på landnivå. Enkelte tilbydere har rapportert globale tall, mens flere har rapportert tall fra EU, uten EØS/EFTA-landene, inkludert Norge. Dette gjelder Twitter, Twitch, Meta og til dels Microsoft LinkedIn (GR), mens TikTok, Google og Microsoft Bing og Microsoft Advertising har inkludert norske tall. LinkedIn har levert norske tall i statusrapporten.

Medietilsynets har rangert de ulike tilbyderens etterlevelse av bransjenormen fra best til dårligst: TikTok, Microsoft, Google og Meta. Twitch og Twitter har levert så svake rapporter at Medietilsynet vurderer at de ikke etterlever bransjenormen.

Figur 2: Kvalitet på utvalgte kvantitative rapporteringspunkter i plattformenes grunnlagsrapport: Grønt indikerer tilfredsstillende rapportering, gult indikerer rapportering med betydelige mangler, rødt indikerer manglende eller svært mangelfull rapportering. Venstre graf viser vurdering av tall rapportert på internasjonalt nivå (EØS, EU eller globalt), mens høyre graf viser vurdering av tall rapportert for Norge. Vurderingen gjenspeiler detalj- og presisjonsnivået på de rapporterte tallene, og er ikke en vurdering av om tallene er korrekte eller gode.



Medietilsynet har vurdert den kvantitative rapporteringen i GR for EØS/EU og Norge, der TikTok kommer best ut, etterfulgt av Google og Microsoft (som ikke rapporterer for Norge for LinkedIn i GR). Metas kvantitative rapportering vurderes som svak, både på norsk og europeisk nivå. Som det kommer frem av figur 2, vurderes rapportene fra Twitter og Twitch som svært mangelfulle.

Hvis bransjenormen skal fungere etter intensjonen, er det avgjørende at underskriverne setter seg inn i retningslinjene og faktisk følger dem. Medietilsynet mener at det må arbeides aktivt

Tabell 1: Utvalgte nøkkeltall for kvantitative indikatorer som måler plattformenes innsats mot desinformasjon. Medietilsynet har justert tallene fra grunnlagsrapporten fra januar 2023 (GR) til 181 dager, slik at de er sammenlignbare med tallene fra den første statusrapporten fra september 2023 (SR1). TikTok fjernet dobbelt så mange innlegg i EØS i SR1 sammenlignet med GR, og tallet for sletting av falske kontoer er tredoblet. Rapporten sier lite om årsaken. Det kan skyldes at TikTok har intensivert tiltakene mot desinformasjon, at nye brukervilkår har hatt effekt, at bruken av TikTok som en kanal for desinformasjon har blitt vanligere, eller skjevheter i datagrunnlaget. I den andre enden av skalaen er Meta, som har en markant nedgang i antall fjernede desinformasjons-innlegg i SR1 sammenlignet med GR. Dette er tall Medietilsynet mener det er viktig å gå dypere inn i.

Nøkkeltall, plattformenes egenrapportering for arbeid mot desinformasjon

	TikTok	Twitch	Meta	Microsoft	Google
Innlegg fjernet som desinformasjon	EØS: 71 961 (GR), 142 711 (SR1) Norge: 116 (GR), 2 055 (SR1)	N/A	Facebook: EU: > 208 543 (GR), > 140 000 (SR1) Instagram: EU: > 11 608 (GR), > 6 900 (SR1)	LinkedIn: EU: 20 103 (GR), 24 205 (SR1) Norge: N/A (GR), 113 (SR1)	YouTube: EØS: > 21 641 (GR), 16 558 (SR1) Norge: > 157 (GR), 124 (SR1) Google Advertising: EØS: 43.5 mill. (GR), 10.0 mill. (SR1) Norge: 21 061 (GR), 12 730 (SR1)
Falske kontoer fjernet	EØS: 1.7 mill. (GR), 6.0 mill. (SR1) Norge: 13 813 (GR), 27 052 (SR1)	Globalt: 1.6 mill. (GR), N/A (SR1)	Facebook: Globalt: 3.0 mrd. (GR), 1.1 mrd. (SR1)	LinkedIn: EU: 5.0 mill. (GR), 6.7 mill. (SR1) Norge: N/A (GR), 88 195 (SR1)	YouTube ³⁴ : Globalt: > 20 000 (GR), 7.4 mill. (SR1)
Følgere av falske kontoer	EØS: 36 mill. (GR), 49.1 mill. (SR1) Norge: 414 319 (GR), 1.3 mill. (SR1)	N/A	Facebook: Globalt: > 50 000 (GR), > 570 000 (SR1) Instagram: Globalt: > 70 000 (GR), > 1000 (SR1)	N/A	N/A
Stoppe finansiering av desinformasjon: antall annonser fjernet eller blokkert	EØS: 11 415 (GR), N/A (SR1) Norge: 128 (GR), N/A (SR1)	N/A	Facebook og Instagram: EU: > 17 313 (GR), 24 000 (SR1)	LinkedIn: EU: 111 (GR), 7 (SR1) Norge: N/A (GR), 0 (SR1) Microsoft Advertising: EØS: 46 108 (GR), 2.9 mill. (SR1) Norge: 216 (GR), 102 069 (SR1)	Google Advertising ³⁵ : EØS: 5.8 mill. (GR), 20.2 mill. (SR1) Norge: 2 353 (GR), 71 323 (SR1) Samlet verdi: EØS: 25.2 mill. € (GR), 30.2 mill. € (SR1) Norge: 3 472 € (GR), 768 250 € (SR1)

³⁴ YouTube-kanaler brukt for spredning av desinformasjon.

³⁵ Antall AdSense-sider gjort tiltak mot på grunn av desinformasjon.

(Tabell 1 fortsetter)

	TikTok	Twitch	Meta	Microsoft	Google
Fakta-sjekking	<p>Innlegg merket med ikke-verifisert innhold:</p> <p>EØS: 53 285 (GR), 77 542 (SR1) Norge: 1 149 (GR), 1 225 (SR1)</p> <p>Andel faktasjekkmerket innhold som ikke ble videreformidlet:</p> <p>EØS: 28 % (GR), 30 % (SR1), Norge: 23 % (GR), 22 % (SR1)</p>	<p>Antall brukere av hjelpesenter knyttet til kritisk medieforståelse:</p> <p>Globalt: 16 348 (GR), N/A (SR1)</p>	<p>Antall artikler publisert av faktasjekkere:</p> <p>Facebook: Globalt: > 295 000 (GR), > 190 000 (SR1)</p> <p>Instagram: Globalt: > 75 000 (GR), > 52 000 (SR1)</p> <p>Innlegg merket med ikke-verifisert innhold:</p> <p>Facebook: EU: > 55 mill (GR), > 40 mill (SR1)</p> <p>Instagram: EU: > 3.3 mill (GR), > 1.1 mill (SR1)</p> <p>Andel faktasjekkmerket innhold som ikke ble videreformidlet:</p> <p>Facebook: EU: 25 % (GR), 37 % (SR1)</p> <p>Instagram: EU: 38 % (GR), 38 % (SR1)</p>	<p>Antall faktasjekk-URLer på førstesiden av et søk:</p> <p>Bing Search: EØS: 6 364 (GR), 4 184 (SR1) Norge: 257 (GR), 186 (SR1)</p>	<p>Faktasjekk-artikler tilgjengelig:</p> <p>Globalt: 185 728 (GR), 211 154 (SR1) Norske: 447 (GR), 481 (SR1)</p>

Kapittel 2: Kontroll av annonser

Forsterkning av spredning av desinformasjon skjer primært på to nivåer, gjennom annonsefinansiert og algoritmisk forsterkning. Kapitlet inneholder tre forpliktelser; hindre annonseinntekter for spredere av desinformasjon, håndtere annonser kategorisert som desinformasjon, og samarbeid mot desinformasjon.

Meta har egne retningslinjer for annonsører. Innholdsleverandører kan ikke tjene penger på innlegg som bryter med Metas retningslinjer eller innlegg som er faktasjekket og merket som feilinformasjon. Annonsører kan velge å ikke plassere annonser i nærheten av et sett uønskede innholdskategorier. Meta har etablert en søkbar annonsedatabase for reklamefinansiert distribusjon av politisk innhold.

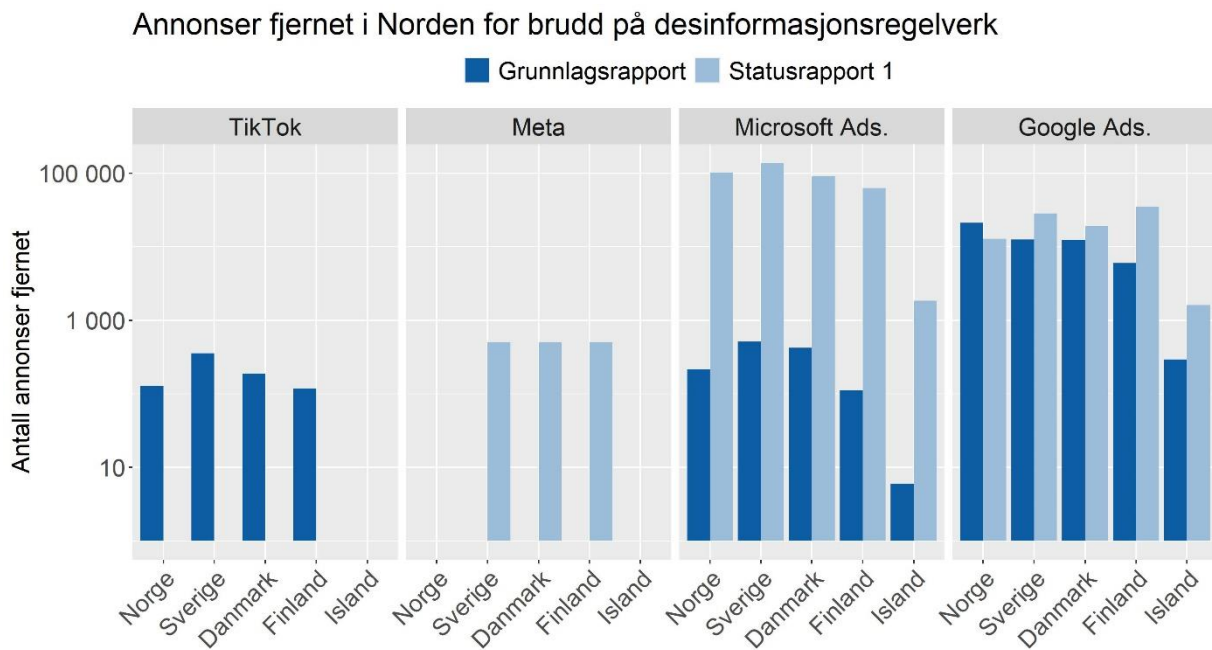
TikTok oppgir at de ikke deler annonseinntekter med annonsører eller brukere, og at de derfor ikke rapporterer på denne forpliktelsen. TikTok har definert flere annonsekategorier som ikke tillates, som ble revidert våren 2023. Hvis en annonse bryter med retningslinjene, blir den stanset, og annonsøren har rett til å klage på beslutningen.

Google forbyr politisk reklame med skjult eller villedende annonsør. Forretningsmodellen legger opp til annonseinntektsdeling, altså at de som publiserer innhold på en side, får en prosentandel av annonsene Google selger på denne siden. Restriksjoner skal hindre finansiering av desinformasjon via dette systemet, blant annet at man ikke tillater manipulerende innhold relatert til politikk. Rapportene viser at betydelige mengder innhold blir sanksjonert for brudd på disse reglene. Google har også innført restriksjoner mot visse typer annonser knyttet til klimaendringer, Covid og krigen i Ukraina. Annonsører kan selv velge vekk visse typer innhold de ikke ønsker å plassere sine annonser ved siden av.

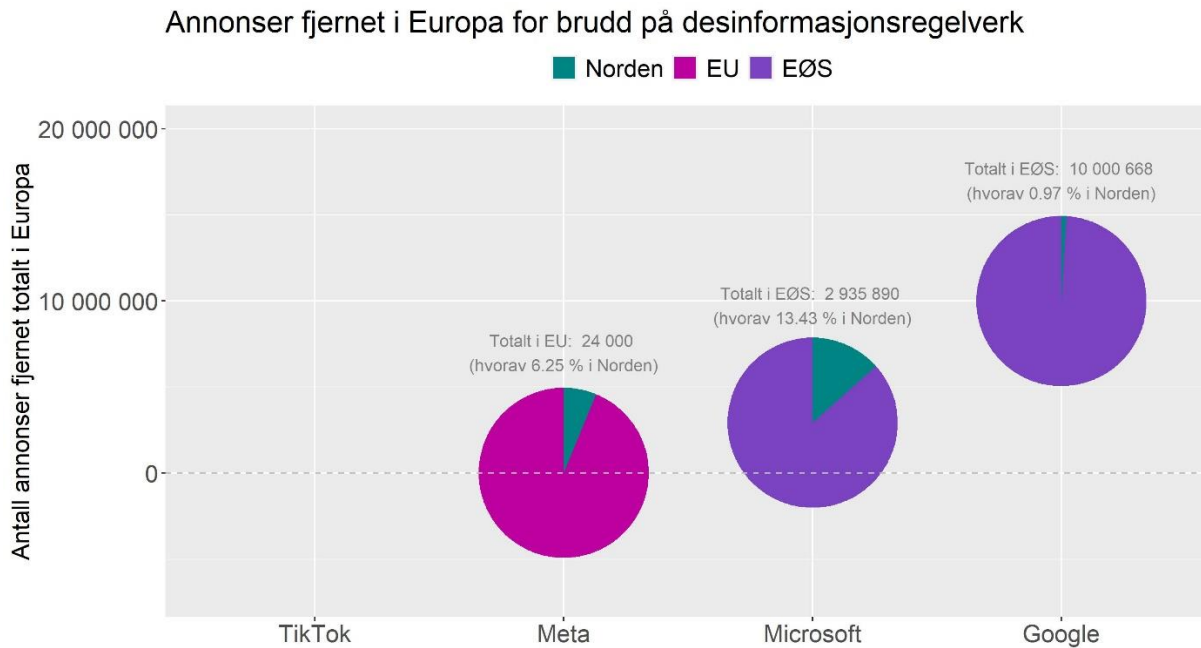
Microsoft Advertising har innført nye regler³⁶ for å stoppe publisering og spredning av skadelig desinformasjon, samt hindre plassering av annonser knyttet til desinformasjonsinnhold. LinkedIn mener problemstillingen er lite relevant for deres tjeneste, da det ikke er mulig å kjøre annonser knyttet til eget innhold. Altså har de ingen modell for annonseinntektsdeling som muliggjør inntjening fra feil- og desinformasjonskampanjer. Microsoft forbyr feil- og desinformasjon og annet villedende innhold på sin plattform. Selskapet leverer ingen rapportering om LinkedIn for Norge i GR, men dette er på plass i SR1.

³⁶ <https://about.ads.microsoft.com/en-us/policies/disallowed-content/information-integrity-and-misleading-content#information-integrity>

Figur 3: Antall annonser fjernet i de nordiske landene for brudd på desinformasjonsregelverk. For sammenligningens skyld er tallene justert for å dekke en periode på 181 dager. Legg merke til logaritmisk y-akse (y-verdien tidobles for hvert trinn). TikTok rapporterte disse tallene kun i grunnlagsrapporten (annonser fjernet på grunn av misledende eller falskt innhold). Meta rapporterte kun i statusrapporten (for Sverige, Danmark og Finland er tallene oppgitt som «mindre enn 500»; uten tall for Norge eller Island). Microsoft Advertising har hatt en markant vekst, nær en tidobling, i antall annonser som er fjernet i de nordiske landene. Rapporten gir ingen klar forklaring på årsaken til dette, men det kan ha sammenheng med at Google Advertising har innført nye begrensninger mot skadelig desinformasjon. Merk ellers at Microsoft ikke har rapportert for LinkedIn (GR). SLI 2.1.1.



Figur 4: Antall annonser fjernet i Europa for brudd på desinformasjonsregelverk i H1 2023 (tall er oppgitt på EU-nivå for Meta, og på EØS-nivå for Microsoft og Google). TikTok rapporterte ikke disse tallene i statusrapporten. Andel annonser fjernet i Norden (Norge, Sverige, Danmark, Finland og Island) er farget grønt. Figuren illustrerer store forskjeller i antall annonser som fjernes av de ulike tilbyderne i Norden, EU og EØS, 0 på TikTok, 24 000 for Meta (EU) og 10 millioner for Google (EØS). SLI 2.1.1.



Kapittel 3: Politiske annonser

Kapitlet tar for seg retningslinjer for politiske annonser og hvilke tiltak som er gjort for å unngå uønsket politisk påvirkning, og inneholder sju forpliktelser. Blant annet skal politisk reklame tydelig defineres, merkes og verifiseres, med krav til åpenhet og innsyn, og det skal etableres en database tilgjengelig for forskere. Tiltakene er koblet til DSA, og forordningen for politisk reklame,³⁷ jf. Kapittel 3, bokstav d.

Meta rapporterer at annonsører for sosiale forhold, valg og politikk (Social Issues, Elections or Politics – SIEP) må gjennomgå en verifiseringsprosess. Denne typen annonser merkes med informasjon om den som betaler for annonsen, og annonsene er også tilgjengelig i et søkbart annonsebibliotek,³⁸ med en tilgang til API for forskere.³⁹ Meta oppgir å ha en særskilt beredskap i forbindelse med demokratiske valg.⁴⁰

TikTok oppgir at tjenesten ikke formidler politiske annonser, og har rapportert på få av forpliktelsene. Dette til tross for at TikTok tilbyr «cause-based» annonser, og at bransjenormen omfatter «issue-based» annonsering. TikToks definisjon av politisk reklame fremstår etter Medietilsynets vurdering som snever, og bør etterprøves av arbeidsgruppen for

³⁷ Eg. Forordningen for målretting og åpenhet om politiske annonser (under behandling i EU):

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0731>

³⁸ https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=NO&media_type=all

³⁹ <https://www.facebook.com/ads/library/api/>

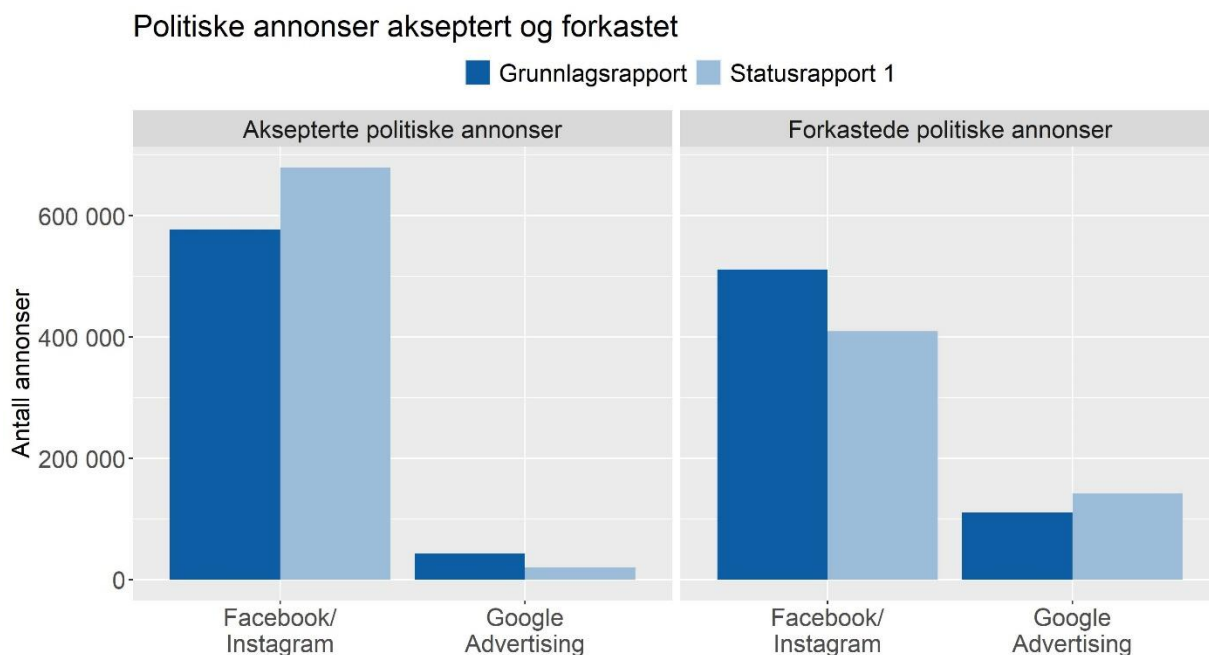
⁴⁰ <https://transparency.fb.com/nb-no/policies/ad-standards/siep-advertising/siep/>

bransjenormen (Permanent Task Force), da grensegangen mellom «cause-based» og «political» fremstår diffus. I kommune- og fylkestingsvalget brukte eksempelvis norske partier rundt åtte millioner kroner på annonser på TikTok, med et klart formål å mobilisere velgerne.⁴¹

Google har egne regler for nasjonale og europeiske valg i det indre markedet, der politisk reklame kun er tillatt fra verifiserte annonsører. Annen rapportering i dette kapitlet er mangelfull. Google svarer kort og ufullstendig på hvordan selskapet forholder seg til øvrig politisk reklame. Google avventer varslet regulering av politisk annonsering i EU, uten å spesifisere dagens praksis godt. Det vises til retningslinjer for «sensitive spørsmål» istedenfor å svare konkret på hvordan selskapet forholder seg til reklame knyttet til politiske enkelt saker.

Microsoft forbyr politiske annonser både for Microsoft Advertising og LinkedIn. LinkedIn har streng praksis på dette området. Plattformen har ikke tillatt politiske annonser siden 2018, inkludert annonser til støtte for en kandidat, et parti eller for å påvirke valg. Det er også restriksjoner knyttet til annonser for sensitive politiske saker. Advertising har tilsvarende streng praksis globalt, og forbyr annonser for valgrelatert innhold, politiske kandidater, partier, avstemningsforslag og politisk innsamling m.m. Hvorvidt dette håndheves godt i praksis er vanskelig å vurdere ut fra rapporteringen. Selskapet har ikke signert en rekke av forpliktelsene fordi disse vurderes som irrelevante i forlengelsen av Microsofts generelle forbud mot politisk annonsering.

Figur 5: Antall politiske annonser vurdert og akseptert (til venstre) og vurdert og forkastet (til høyre) i EU. Meta oppgir sine tall som antall aksepterte annonser om sosiale temaer, valg eller politikk (SIEP) med en «betalt av»-klausul, og antall annonser forkastet for på bakgrunn av SIEP-regelverk. Google oppgir sine tall som antall annonser knyttet til EU-valg fra verifiserte aktører (aksepterte annonser) og ikke-verifiserte aktører (forkastede annonser). Innrapporterte tall er justert til en periode på 181 dager. TikTok og Microsoft oppgir at de ikke tillater politisk reklame. SLI 6.2.1 og SLI 7.1.1.



⁴¹ <https://www.vg.no/nyheter/innenriks/i/nQ715m/se-forskjellen-knuser-alle-paa-tiktok>

Kapittel 4: Tjenestens integritet

Kapitlet om integritet kommer med tre forpliktelser; felles forståelse av manipuleringsmetoder, åpenhet om bruk av KI, og åpenhet om samarbeid. Det skal blant annet innføres tiltak for å avdekke og motvirke taktikker, teknikker og prosedyrer (TTP) for å spre desinformasjon, manipulere valg eller holdninger. Det skal også innføres tiltak mot koordinert ikke-autentisk adferd, eksempelvis falske kontoer og *botter*, og kunstig skapt innhold, eksempelvis deepfakes.

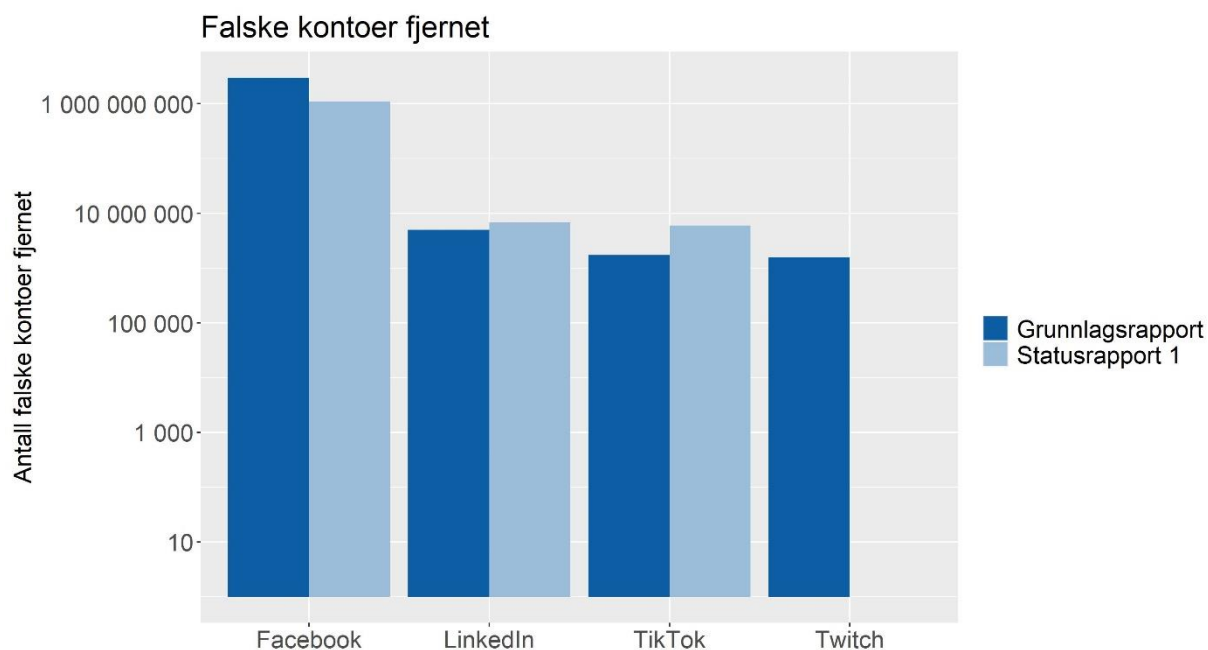
Meta oppgir å ha fjernet 1,5 milliarder falske kontoer globalt bare i tredje kvartal 2022. Dette sier noe om omfanget av ikke-autentiske kontoer. Meta har lang erfaring med tiltak mot koordinerte påvirkningsforsøk, og har en særlig beredskap i forbindelse med demokratiske valg. Det er innført nye retningslinjer for kunstig skapt innhold.

TikTok rapporterer å ha innført flere tiltak mot ikke-autentiske kontoer som *bots*, bruk av KI for likes og reaksjoner, samt mot kapring av kontoer. TikTok bruker en kombinasjon av KI og menneskelige moderatører i vurdering av innhold, og har innført nye regler mot kunstig skapt innhold.

Google har omfattende tiltak for å redusere risiko for manipulering og villedning både på sin søkemotor og på YouTube, og redegjør grundig for dette i sin rapportering. Sentralt er innretningen av algoritmer for å fremheve kvalitetssikret innhold og nedprioritere innhold som vurderes av lavere kvalitet, noe falske kontoer og misledende innhold typisk vil sortere under. Google har også verktøy for å stoppe spam. Selskapet har imidlertid ikke rapportert kvalitativt på dette kapitlet, utover noen helt overordnede tall fra egne kvartalsregnskap og for SR1 dessuten noe beskrivelse av håndhevelse av spamregler.

Microsoft viser til generelle brukervilkår for å forhindre manipulerende adferd, og at LinkedIn forbyr desinformasjon på sine flater. Medietilsynet vurderer rapporteringen som svært mangelfull, blant annet mangler flere av de konkrete TTP-ene data, særlig i GR. Dette er forbedret i SR1. Når det gjelder søkemotoren Bing, argumenterer Microsoft godt for at tiltak for å hindre viral spredning av desinformasjon er mindre relevant for en søkemotor enn for sosiale medier. Søkemotorene har ikke nyhetsfeed eller deling av innhold, men svarer på søketermer. Strategien er å prioritere autorativt innhold. Derimot er det relevant å stoppe forsøk på manipulering av disse søkealgoritmene. Microsoft oppfordrer til et mer fleksibelt rapporteringssystem, der tilbyderne kun rapporterer for relevante TTP-er. Microsoft iverksetter undersøkelser ved mistanke om forsøk på å manipulere søkealgoritmene til Bing. Interne forskerteam skal gjenkjenne trusler. Koordinert, ikke-autentisk adferd (coordinated inauthentic behaviours, CIB) skal fjernes. Det er verdt å merke seg at rapportering av tiltak mot generativ KI er tatt inn i statusrapportene til tilbydere av VLOP/VLOSE. Google og Microsoft Bing merker kunstig genererte bilder, mens TikTok krever at realistisk syntetisk innhold merkes av opplasteren.

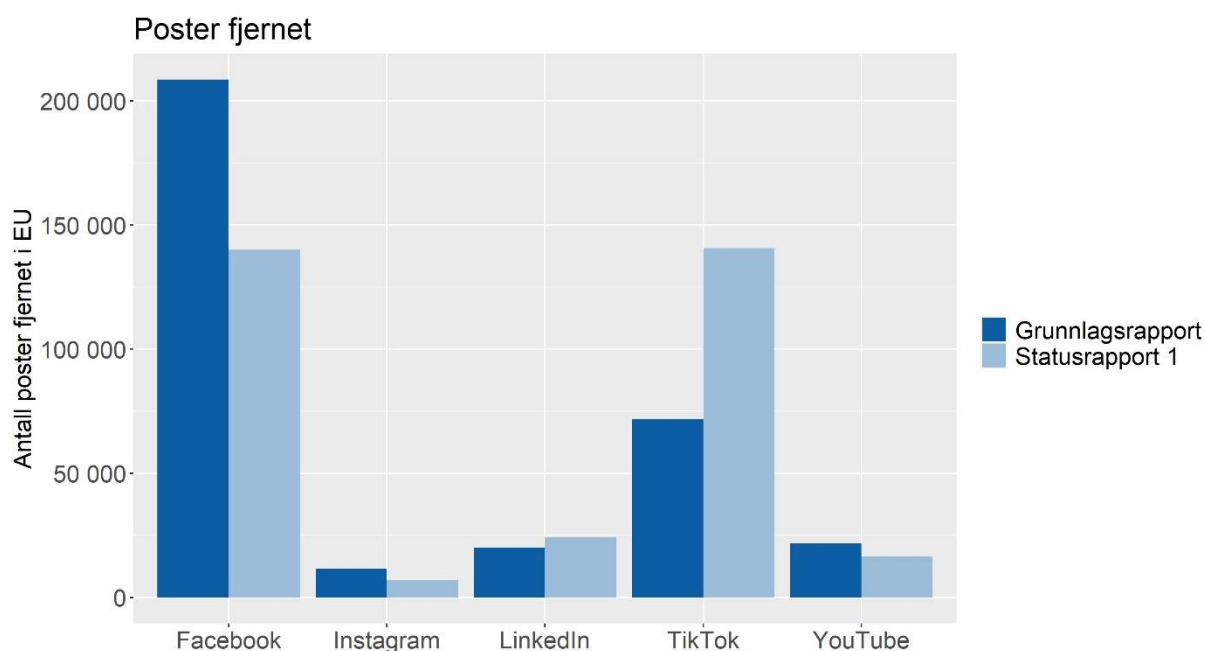
Figur 6: Antall falske kontoer fjernet på de ulike plattformene i EØS (TikTok) / EU (LinkedIn) / globalt (Facebook og Twitch). Tall justert til 181 dager. Legg merke til logaritmisk y-akse. SLI 14.2.1.



Kapittel 5: Styrke befolkningens motstandskraft

Kapitlet inneholder seks forpliktelser: kritisk medieforståelse, «safe design», enkel identifisering av desinformasjon, mulighet til å flagge feil- og desinformasjon, klageordninger, og begrensninger av desinformasjon i meldingsapplikasjoner.

Figur 7: Antall poster fjernet i EU for å begrense skadelig, falskt og misvisende innhold. Tall justert til 181 dager. Et interessant funn er at antall poster som fjernes fra Facebook og Instagram går ned, mens antallet fjernede innlegg på TikTok er doblet fra GR til SR1. SLI 18.2.1.



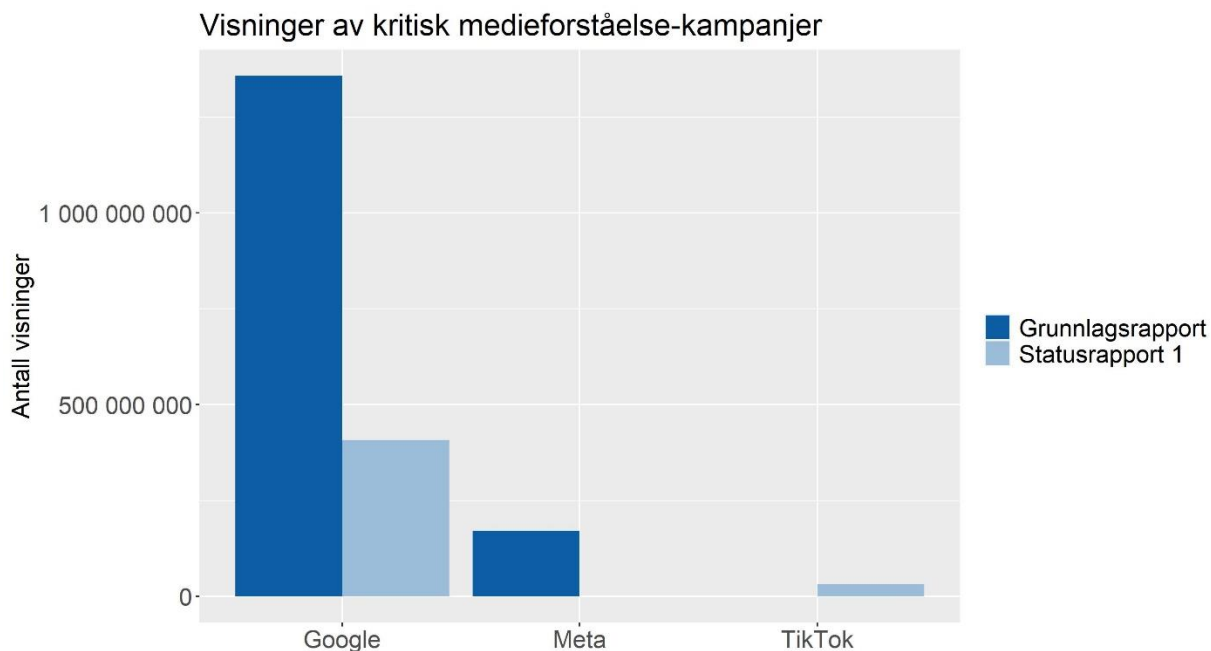
Meta oppgir å støtte en rekke informasjonskampanjer på Facebook og Instagram, og samarbeider med 26 faktasjekkorganisasjoner, inkludert faktisk.no. Meta merker feil- og desinformasjon, og henviser brukerne til troverdig innhold. Meta har revidert informasjonen om hvordan maskinlæringsmodellene fungerer, delt inn i 22 «systemkort». Oppdaterte regler for sanksjoner mot brukere som bryter med vilkårene er offentliggjort.

TikTok merker desinformasjon innen flere kategorier, som Covid, Holocaust og krigen i Ukraina. Innhold fra statskontrollerte medier merkes, og det lenkes til troverdige kilder. TikTok har etablert et samarbeid med Logically om faktasjekk i Norge. Antall faktasjekkede innlegg er begrenset, i Sverige ble 448 innlegg gjennomgått, ifølge grunnlagsrapporten.

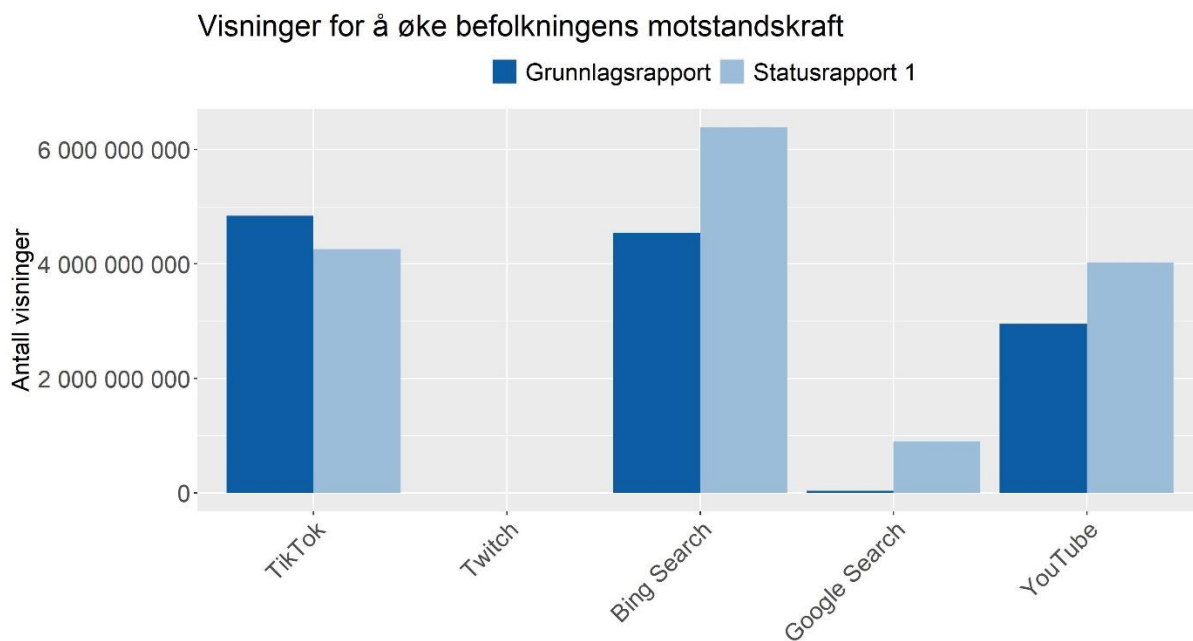
Google opplyser at YouTube har gjennomført kampanjen «Hit Pause» med god europeisk spredning, men tilsynelatende uten distribusjon i Norge. Ellers viser Google til generelle funksjoner som gir informasjon om søkeresultatenes kontekst og kilder. I SR1 er det oppdatering om kampanjer og tiltak for å motvirke russisk desinformasjon knyttet til krigen i Ukraina, men lite er relevant for norske brukere. Den kvalitative rapporteringen på videoer fjernet fra YouTube som følge desinformasjon er forbedret i statusrapporten.

Microsoft rapporterer detaljert om tiltak for å bedre brukernes kritiske medieforståelse, som NewsGuard, et verktøy for å vurdere troverdighet av nyhetskilder. Medietilsynet vurderer at det er svak og mangelfull rapportering på rekkevidden av kampanjer, selv om det nevnes kampanjer knyttet til Covid 19-pandemien og russisk invasjon av Ukraina på LinkedIn. I kjølvannet av lanseringen av KI-drevne tjenester som Bing Chat har Microsoft igangsatt nye kampanjer og forskningsprosjekter rundt ansvarlig KI.

Figur 8: Antall visninger av ulike kampanjer knyttet til kritisk medieforståelse. Google har rapportert tall for YouTube-kampanjen «Hit pause», mens Meta har rapportert tall for kampanjer knyttet til krigen i Ukraina, klimakrisen, sikkerhet og kritisk medieforståelse (kun i grunnlagsrapporten). TikTok har rapportert antall visninger av en annonse for en polsk kritisk medieforståelseskampanje (kun i statusrapporten). Det er verdt å merke at antall visninger av TikToks kampanje er begrenset, sammenlignet med Meta og YouTube. Tall justert til 181 dager. SLI 17.2.1.



Figur 9: Antall visninger av ulike elementer for å øke befolkningens motstandskraft mot desinformasjon. TikTok: antall visninger av videomerknader, søkeinnlegg og offentlige kunngjøringer knyttet til temaene covid-19, covid-19-vaksiner, Holocaust-fornektelse, apekopper og klimaendringer samt russisk innhold, i tillegg til antall visninger av sikkerhetscenter-side knyttet til covid-19 og valg (EØS-tall). Twitch: Antall visninger av en hjelpesenter-video om kritisk medieforståelse (globale tall). Bing Search (Microsoft): antall visninger av verktøyene NewsGuard Impressions, Knowledge Cards, Transparency HubViewership, Public Service Announcement og NewsGuard Extension Downloads (EØS-tall). Google Search: antall visninger av verktøyene "More About This Page" og "About this result" (EØS-tall). For YouTube er antall visninger av informasjonspaneler rapportert (EØS-tall). Tall justert til 181 dager. Meta rapporterte kun tall for ulike kampanjer (fremstilt i **Feil! F** **ant ikke referanseilden.**). SLI 17.1.1.




5 Nærmere vurdering av de ulike tilbydernes rapporter

I dette kapitlet gjennomgås de ulike tilbydernes grunnlagsrapporter og statusrapporter, og rapportene vurderes nærmere opp mot bransjenormens retningslinjer og forpliktelser. Rapportene er rangert etter dårlig (1/rød), middels (2/oransje), og god (3/grønn).

Meta

Grunnlagsrapport: kvalitativ 2,  kvantitativ 1 . Antall sider: 168.

Statusrapport 1: kvalitativ 2,  kvantitativ 1 . Antall sider: 137.

Rapportering for Norge: Nei 

Vurdering: Meta rapporterer på de fleste kvalitative indikatorene, men svarene er omfattende, repeterende og lite poengterte.⁴² Enkelte svar er lite troverdige, eksempelvis når det gjelder tilgang til data for forskere, der forskere oppgir at tilgangen er vanskeligere og arbeidet med CrowdTangle er nedprioritert.⁴³ Meta rapporterer identiske tall for flere indikatorer, og gjenbraker innhold i grunnlagsrapporten i statusrapporten, noe som kan indikere at selskapet ikke har satt seg godt nok inn i retningslinjene, eller har nedprioritert arbeidet.

Rapportering på de kvantitative indikatorene har store mangler. Noen indikatorer er utelatt i begge rapportene. I statusrapporten er tre indikatorer ubesvart, begrunnet med at Meta ikke rakk å rapportere før fristen. Meta rapporterer kvartalsvis, ikke på månedsbasis, og de kvantitative rapportene er omtrentlige. Noen ganger rapporteres det på globale tall, andre ganger på tall fra EU. Det rapporteres ikke for EØS/EFTA-landene, noe som er i strid med retningslinjene. Det rapporteres ufullstendig om antall falske kontoer, særlig på Instagram. Det varierer om det er rapportert på Facebook eller Instagram. Meta er den klart største tilbyderen av sosiale medier i det indre markedet og i Norge, og det bør forventes høyere nivå på rapporteringen. **Medietilsynets vurdering er at Meta etterlevelse av bransjenormen er svak, og at det er behov for betydelige forbedringer.**

Relevante funn

- 37–38 prosent av videredelingene avbrytes hvis et innlegg er vurdert og merket som feilinformasjon etter faktasjekk. Dette viser at merking av desinformasjon har effekt. Meta oppgir at mer enn 40 millioner innlegg i EU er faktasjekket.
- Meta oppgir at fem prosent av kontoene på Facebook og Instagram er falske, og har avdekket og utestengt 200 nettverk fra 68 land (42 språk) for «Coordinated Inauthentic Behaviour».
- Meta har etablert søkbar database for politiske annonser, men den fremstår som lite strukturert.

⁴² Meta rapporterer samlet for Facebook og Instagram, og behandles samlet i Medietilsynets vurdering av rapportene, hvis noe annet ikke er presisert.

⁴³ <https://www.cnet.com/news/misinformation/meta-plans-to-drop-tool-researchers-use-to-track-misinformation-online-report-says/>

- Meta omtaler sin API for forskertilgang til data i positive vendinger, i motsetning til forskeres erfaringer om at tjenesten CrowdTangle bygges ned. Det bør vurderes om praksisen er i overensstemmelse med DSAs artikkel 40 om forskertilgang.
- Meta har utviklet en maskinlæringsmodell for merking av kunstig skapt innhold, som svar på den eksplosive utviklingen av syntetisk produsert lyd, bilde, video og tekst.
- Meta rapportert identiske tall på flere av de kvantitative indikatorene, og leverer etter Medietilsynets vurdering den svakeste rapporteringen av de globale aktørene (VLOP/VLOSE) som har forpliktet seg til bransjenormen.

Tabell 2: Utvalgte nøkkeltall fra Meta-rapportene. Legg merke til ulik rapporteringsperiode, og at tallene ikke er justert; i grunnlagsrapporten er det hovedsakelig rapportert for et kvartal (Q2), mens det i den første statusrapporten er rapportert for første halvår (H1).

Nøkkeltall, Meta

KPI	Facebook	Instagram
Antall annonser fjernet grunnet desinformasjon (SLI 2.1.1)	Over 8 800 (EU; Q4 2022), over 24 000 (EU; H1 2023)	Over 8 800 (EU; Q4 2022), over 24 000 (EU; H1 2023)
Antall annonser fjernet for brudd på retningslinjer (SIEP-regelverk som omfatter samfunnsproblemer, valg, politikk) (SLI 7.1.1)	Over 260 000 (EU; Q4 2022), over 410 000 (EU; H1 2023)	Over 260 000 (EU; Q4 2022), over 410 000 (EU; H1 2023)
Antall falske kontoer fjernet (SLI 14.2.1)	1.500 mrd. (globalt; Q2 2022); 1.102 mrd. (globalt; H1 2023)	157 nettverk (globalt, Q2 2022), 50 kontoer (globalt, H1 2023)
Antall følgere av fjernede falske nettverk/kontoer (SLI 14.2.2)	Over 22000 (globalt, Q2 2022); over 550 000 (globalt, H1 2023)	Over 40 000 (Q2-Q4 2022), 1070 (H1 2023)
Andel falske kontoer av alle aktive brukere (SLI 14.2.3)	Rundt 5 % (globalt; Q2 2022), 4-5 % (globalt; H1 2023)	N/A
Antall innlegg fjernet for brudd på regler mot skadelig helseinformasjon eller valginnblanding (SLI 18.2.1)	Over 106 000 (EU; Q4 2022), over 140 000 (EU; H1 2023)	Over 5 900 (EU; Q4 2022), over 6 900 (EU; H1 2023)
Antall poster som er faktasjekket (SLI 31.1.1)	Over 28 mill. (EU; Q4 2022), over 40 mill. (EU; H1 2023)	Over 1.7 mill. (EU; Q4 2022), over 1.1 mill. (EU; H1 2023)
Antall artikler tilgjengelig skrevet av faktasjekkere (SLI 31.1.1)	150 000 (globalt; Q4 2022), 190 000 (globalt; H1 2023)	40 000 (globalt; Q4 2022), 52 000 (globalt; H1 2023)
Andel videreledinger som er avbrutt etter visning av faktasjekkmerking (SLI 31.1.2)	25 % (EU, des. 2022), 37 % (EU, 5. jan – 30. juni 2023)	38 % (EU, des. 2022), 38 % (EU, 5. jan – 30. juni 2023)
Antall poster innholdsvurdert daglig (SLI 31.1.3)	Over 2 mill. (globalt per januar 2023), (ikke rapportert i statusrapport 1)	Over 2 mill. (globalt per januar 2023), (ikke rapportert i statusrapport 1)
Antall månedlige aktive brukere (SLI 31.1.3)	2.96 mrd. (globalt, Q2 2022), 258 mill. (EU, per juni 2023)	Over 2 mrd. (globalt, Q2 2022), 257 mill. (EU, per juni 2023)

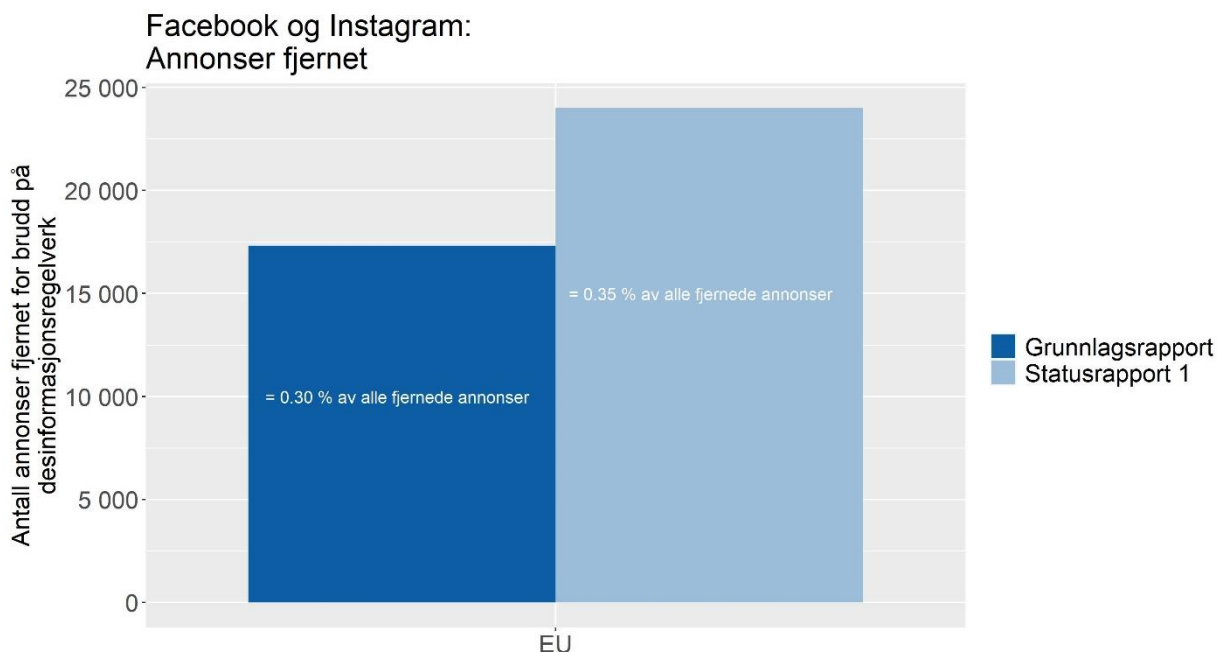
Kapittel 2. Kontroll av annonser: GR: Meta oppgir at innholdsleverandører (annonsører, skapere eller tredjepartsleverandører) ikke kan tjene penger på innlegg som bryter med [plattformenes standarder](#), eller er merket som feilinformasjon av faktasjekkere (QRE.1.1.1). Endringer i retningslinjene [publiseres](#) jevnlig (QRE 1.2.1) og annonsører kan velge innholdskategorier deres annonse ikke skal vises sammen med, ved hjelp av tjenesten Zefr (QRE 1.3.1). Meta er [akkreditert](#) av Media Rating Council, tilknyttet Global Alliance for Responsible Media (GARM), og kategoriserer innhold i risikogrupper høyt, middels og lavt, i tråd med GARMs rammeverk. Meta sanksjonerer kontoer som bryter retningslinjene, ikke er autentiske eller er koblet til sanksjonerte kontoer, og har et samarbeid med faktasjekkere (QRE.3.1.1). Meta rapporterer identiske tall på flere indikatorer i rapporten, blant annet 1.1.1 og 2.1.1, som er en svakhet.

SR1: Meta har innført få nye tiltak i rapporteringsperioden. Det er ikke rapportert på sentrale tall om annonsekontroll før fristen (SLI-ene 1.1.1, 1.1.2, 2.4.1). Meta rapporterer fremdeles omtrentlige tall, og har ikke inkludert tall for EØS/EFTA-landene.

Kapittel 3. Politiske annonser: GR: Meta har [retningslinjer for annonser om valg og om politikk](#) (QRE 4.1.1), der annonsør må autoriseres (basert på ID og tofaktorautentisering). Navn på annonsør merkes i annonsen, med flere begrensninger, blant annet for å minimere risiko for utenlandsk innflytelse (QRE 5.1.1). Meta har en rekke krav om politisk reklame (QRE 6.1.1), der navn på annonsør skal komme frem og merkes i annonsen (QRE 6.2.1). Det er opprettet et eget søkbart [annonsebibliotek](#) (QRE 6.2.2/10.2.1), men søkefunksjonen og brukergrensesnittet virker lite strukturert. Mediekontoer som er under statlig kontroll merkes «statskontrollert media». Meta har fjernet muligheten for målretting av politisk reklame basert på sensitive kategorier som etnisitet, religion og seksuell orientering (QRE 9.2.1).

SR1: Meta oppgir få endringer i vilkårene for politiske annonser, men skriver at det er innført flere innholdskategorier merket «hvorforser jeg dette?», og at annonser også er omfattet av merkingsordningen (QRE 9.1.1).

Figur 10: Antall annonser fjernet på Facebook og Instagram samlet for brudd på desinformasjonsregelverk i EU. Andelen dette utgjør av alle annonser fjernet totalt på de to plattformene er oppgitt i hvit skrift. Tall justert til 181 dager. SLI 2.1.1.



Kapittel 4. Integritet: GR: Facebook har fjernet 1,5 milliarder falske kontoer globalt i Q3 2022, og anslår at fem prosent av kontoene på plattformene er falske. Meta har avdekket og utestengt 200 nettverk fra 68 land (42 språk) for «Coordinated Inauthentic Behaviour» (QRE 14.1.1). Meta har retningslinjer for cybersikkerhet, deepfakes, spam, merkevarebruk og personvern, og blokkerer kontoer ved brudd. Tiltak er innført for å redusere bots, automatisk genererte likes/reaksjoner, følgere, kontoer eller sider, kapring av kontoer og spredere av desinformasjon (QRE 14.3). Det er også innført tiltak mot blant annet hatefulle ytringer mot minoriteter, generativ KI for desinformasjon og manipulasjon av Metas forsterkningsalgoritmer (QRE 14.3). Meta har utviklet [en maskinlæringsmodell](#) for å identifisere kunstig skapt innhold (QRE 15.1) og har et eget ansvarlighetsteam for KI som tar utgangspunkt i standarder fra OECD og EC, opp mot pilarene personvern, sikkerhet, rettferdighet, åpenhet og ansvarlighet (QRE 15.2). Det er verdt å merke seg at Meta har begynt å tilby generativ AI for annonsører, slik at de enklere kan lage annonser selv, uten at dette er nevnt eller problematisert i rapporten.⁴⁴ Meta oppgir å ha investert betydelige ressurser i å motvirke koordinerte påvirkningsoperasjoner, og at de ulike Meta-tjenestene samarbeider tett for å redusere risikoene. Det publiseres jevnlig rapporter for hvilke tiltak som er gjort (QRE 16.1.1).⁴⁵

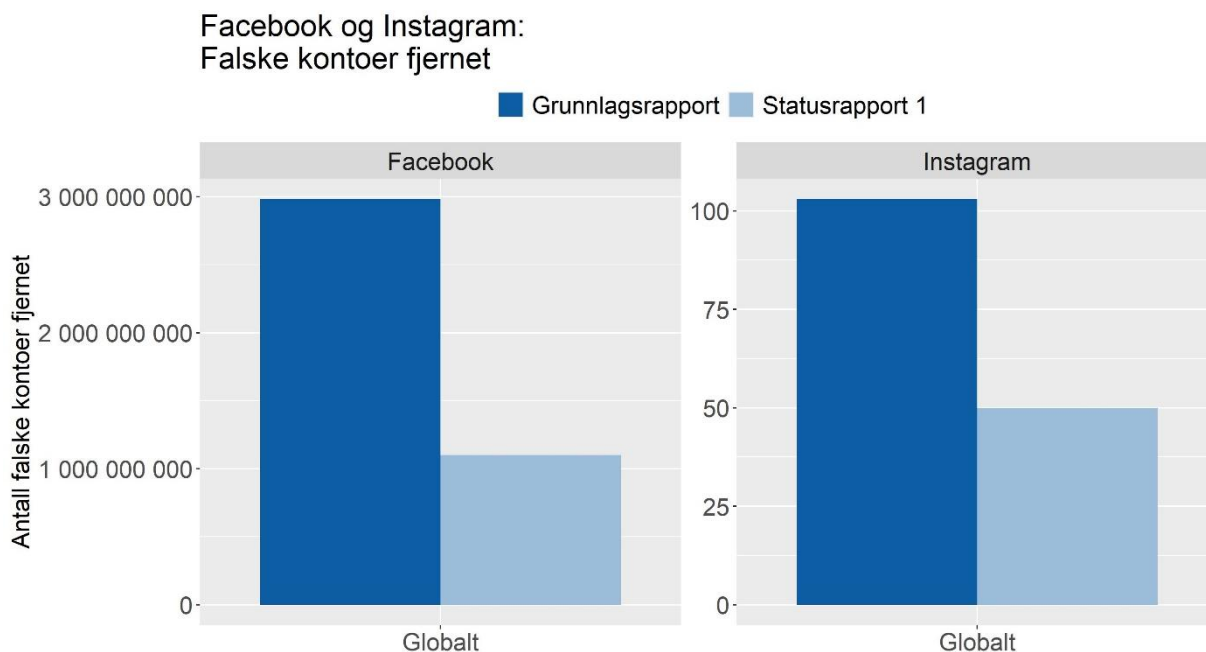
SR1: Meta rapporterer få endringer i arbeidet med integritet etter grunnlagsrapportene, men planlegger neste år å innføre tiltak mot KI-generert desinformasjon. Det vises til Metas egne kvartalsrapporter (Q1 og Q2), der henholdsvis seks påvirkningsoperasjoner og russiske desinformasjonskampanjer behandles (QRE 16.1.1). Meta oppgir å ha beredskap i

⁴⁴ <https://www.reuters.com/technology/meta-starts-rolling-out-generative-ai-tools-all-advertisers-2023-10-04/>

⁴⁵ <https://about.fb.com/news/2022/12/metas-2022-coordinated-inauthentic-behavior-enforcements/>

forbindelse med flere av høstens valg, blant annet valget i Slovakia, med betydelig risiko for russiske påvirkningsoperasjoner.⁴⁶

Figur 11: Antall falske kontoer fjernet globalt på Facebook (til venstre) og Instagram (til høyre). Legg merke til ulike y-akser. For Facebook oppgir Meta både antall falske kontoer globalt og antall nettverk tatt ned grunnet brudd på desinformasjonsregelverk. I figuren er bare førstnevnte tall tatt med. For Instagram er ikke antall falske kontoer globalt rapportert, og tallene i figuren representerer antall kontoer som tilhører de henholdsvis fire (Q3 2022) og to (H1 2023) nettverkene som er fjernet. Tallene er justert til 181 dager for sammenligningens skyld, på grunn av ulik rapporteringsperiode i Metas rapporter. SLI 14.2.1.



Kapittel 5. Tiltak for å styrke brukernes motstandskraft: GR: Meta oppgir flere informasjonstiltak om krigen i Ukraina og klima, inkludert en kampanje rettet mot ungdom (QRE 17.1.1). Selskapet har flere partnere for å fremme kritisk medieforståelse (EU, internasjonale og nasjonale organisasjoner, universiteter og faktasjekkere) og deltar i styringsgruppen i EU Digital Citizenship Working Group (QRE 17.2.1). Meta oppgir å ha effektive verktøy for å redusere skadelig innhold (QRE 18.1.1). Facebook har fire trinn for rangering av innhold, med bruk av et sett algoritmer for hvert trinn: 1) Hvilke innlegg (fra venner etc.) som er mest relevante. 2) Signaler for hvert innlegg (f.eks. interesse-match) 3) personlig prediksjon om engasjement 4) Relevansscore som bestemmer hva som vises i feeden (QRE 18.1.2). Instagram bruker egne, men tilsvarende rangeringssystemer. Meta har innført tiltak for å redusere spredningen av feilinformasjon (QRE 18.1.3), kombinert med faktasjekk-lenke (som bare fem prosent av brukerne klikker på). Innlegg som fjernes er knyttet til brukervilkårene (QRE 18.2.1). Meta samarbeider med flere forskningsinstitusjoner om forskning på desinformasjon (QRE 18.3.1). Grunnleggende informasjon om anbefalingssystemene er tilgjengelig (QRE 19.1.1), og det er mulig å justere parameterne, som vis mindre/mer, favoritt, kronologisk feed, preferanser og «hvorfor ser jeg dette» (QRE 19.2.1), selv om det er umulig for en bruker å måle effekten. Meta samarbeider med 26 faktasjekkere som dekker 22 språk, og merking av desinformasjon gir en lavere score og svekket distribusjon. Meta har en klageordning for avsendere av innlegg som merkes som

⁴⁶ <https://www.politico.eu/article/european-union-slovakia-elections-russian-interference-disinformation/>

feil- eller desinformasjon (QRE 23.2.1). Meta oppgir ikke samarbeidspartner i Norge, men det er kjent at faktisk.no utfører faktasjekk av innhold på Metas plattformer.

SR1: Meta har innført et sett med 22 «systemkort» for å forklare hvordan anbefalingssystemene på Facebook og Instagram fungerer (QRE 19.1.1).⁴⁷ Meta publiserte i februar oppdaterte regler for restriksjoner på Facebook. Første overtredelse fører til en advarsel. Andre til sjetten overtredelse fører blant annet til at brukeren ikke kan publisere i grupper. 7-10 overtredelser vil gi 1, 3, 7 og 30 dagers utestengelse. For alvorlige overtredelser vurderes sterkere virkemidler (QRE 24.1.1). Meta har gjort det enklere å varsle om ulovlig innhold og innhold som bryter med retningslinjene, og klageordninger for brukere som får innlegg fjernet eller begrenset som en følge av DSA. Det er også innført et system på Facebook der brukeren får advarsel om at et innlegg bryter med vilkårene før publisering (QRE 24.1.1).

Kapittel 6. Tilgang for forskere: Meta publiserer relevante rapporter på egne sider (transparensenteret), eksempelvis hvordan brudd på retningslinjene håndheves, visningsstatistikk, antall innlegg som slettes og annonsebibliotek (QRE 26.1.2), men Meta har ikke rapportert på SLI (26.1.1). Omtrent 5000 forskere (1000 akademiske brukerkontoer) har tilgang til et utvalg data via CrowdTangle (QRE 26.2.1 og 2). Meta deltar i en arbeidsgruppe med European Digital Media Observatory (EDMO) om forskertilgang etter DSAs artikkel 40 (27.1.1). En database over koordinerte ikke-autentiske påvirkningsoperasjoner (CIB) er etablert med en API for forskere (QRE 27.4.1). Det er etablert nye tilgangsmåter, som foreløpig er lite utbredt. Kun 70 forskere har tilgang til Metas API til målrettet markedsføring, 250 har tilgang til data for url-delning og 250 har tilgang til generell forsker-API (QRE 28.2.1). Forskere er kritisk til API-tilgangen for forskere, og det meldes om at tjenesten Crowdtangle bygges ned.⁴⁸ Dette bør undersøkes nærmere, og forskeres tilgang til data fra VLOP/VLOSE er også en forpliktelse i DSA, jf. artikkel 40.

Kapittel 7. Tilgang for faktasjekkere: Det rapporteres her på kriterier for å bli godkjent som faktasjekkere (QRE 30.2.1) og fremdrift for merking av faktasjekket innhold (QRE/SLI 31.1.1). Faktasjekkere har tilgang til et eget grensesnitt for å vurdere innhold (QRE 32.1.1). Metas moderatorer vurderer to millioner innlegg daglig på verdensbasis, men det fremgår ikke hvor mange menneskelige moderatorer Meta har, eller hvor mange som behersker ulike språk og kulturer. Dette skyldes at en slik indikator mangler i bransjenormen.


⁴⁷ <https://ai.meta.com/blog/how-ai-powers-experiences-facebook-instagram-system-cards/>

⁴⁸ <https://www.cnet.com/news/misinformation/meta-plans-to-drop-tool-researchers-use-to-track-misinformation-online-report-says/>

TikTok

Grunnlagsrapport: kvalitativ 2 , kvantitativ 3 . Antall sider: 159.

Statusrapport: kvalitativ 2 , kvantitativ 3 . Antall sider: 238.

Rapportering for Norge: Ja 

Vurdering av rapportene: TikTok etterlever i størst grad bransjenormen, og er særlig god på de kvantitative indikatorene, der detaljert rapportering på landnivå er med, inkludert for EØS/EFTA-landene. Det må, som tidligere nevnt, understrekes at etterlevelse av bransjenormen ikke er synonymt med at tiltakene mot desinformasjon er tilstrekkelige. Rapporten er svak når det gjelder samarbeid med faktasjekkere og tilgang for forskere. Det er problematisk at TikTok ikke rapporterer på kapitlet om politiske annonser. TikTok har ikke rapportert i SR1 på antall innlegg som er fjernet for brudd på retningslinjene om desinformasjon, og dette er uheldig. Grunnlagsrapporten tok også utgangspunkt i feil tidsperiode.

Relevante funn:

- TikTok fjerner få annonser sammenlignet med andre tjenester.
- TikTok rapporterer en dramatisk vekst når det gjelder falske kontoer i Sverige og Island, og i antall følgere av falske kontoer i Norge og Island. Utviklingen er tilsvarende i EØS. Tallene bør verifiseres gjennom uavhengig kontroll. Det bør undersøkes hva som er årsaken, for å se på behovet for tiltak i form av kampanjer og undervisningsopplegg rettet mot brukere av plattformen.
- TikTok rapporterer ikke på kapitlet om politiske annonser, men hevder at slike annonser ikke tilbys gjennom tjenesten. Tematiske annonser (issue based) tilbys imidlertid, og disse kan være vanskelige å skille fra politiske annonser. Politiske partier bruker plattformen aktivt for å nå unge velgere, og det bør settes spørsmålsteget ved TikToks egenvurdering på dette området.
- TikTok samarbeider ikke med faktisk.no, men med Logically. TikTok faktasjekker lite innhold sammenlignet med Facebook, Instagram og YouTube. Metodikken for faktasjekking er også forskjellig fra Metas, og kvaliteten på faktasjekkingen i Norge og Norden bør undersøkes nærmere.

Tabell 3: Utvalgte nøkkeltall fra TikTok-rapportene. Legg merke til ulik rapporteringsperiode; i grunnlagsrapporten er det hovedsakelig rapportert for et kvartal (Q4), mens det i den første statusrapporten er rapportert for et halvår (H1).

Nøkkeltall, TikTok

KPI	Norge	EØS
Antall annonser fjernet grunnet desinformasjon (SLI 2.1.1)	65 (Q4 2022), (ikke rapportert i statusrapport 1)	5 802 (Q4 2022), (ikke rapportert i statusrapport 1)
Antall falske kontoer fjernet (SLI 14.2.1)	7 021 (Q4 2022), 27 052 (H1 2023)	875 197 (Q4 2022), 5 979 617 (H1 2023)
Antall som følger fjernede falske kontoer (SLI 14.2.2)	210 593 (Q4 2022), 1 284 990 (H1 2023)	18 352 005 (Q4 2022), 49 149 995 (H1 2023)
Andel falske kontoer av alle aktive brukere (SLI 14.2.4)	N/A	0.67 ⁴⁹ % (EU; Q4 2022), 1.10 % (EU; H1 2023)
Antall visninger av advarsler på videoer knyttet til temaene covid-19, covid-19-vaksine, holocaustfornektelse og apekopper. For statusrapporten er også russisk innhold inkludert (SLI 17.1.1)	19.7 mill. (Q4 2022), 37.9 mill. (H1 2023)	1.6 mrd. (Q4 2022), 2.8 mrd. (H1 2023)
Klikkrate for lenker tilhørende merkelappene nevnt over (SLI 17.1.1)	0.06 % (Q4 2022), 0.09 % (H1 2023)	0.11 % (Q4 2022), 0.15 % (H1 2023)
Antall videoer fjernet for brudd på regler mot skadelig desinformasjon (SLI 18.2.1)	59 (Q4 2022), 2 055 (H1 2023)	36 577 (Q4 2022), 142 711 (H1 2023)
Antall videoer merket som ikke-verifisert av faktasjekkere (SLI 21.1.2)	584 (Q4 2022), 1 225 (H1 2023)	27 084 (Q4 2022), 77 542 (H1 2023)
Andel videredelinger som er avbrutt etter visning av faktasjekk-merking (SLI 21.1.1)	23.08 % (Q4 2022), 22.41 % (H1 2023)	28.39 % (Q4 2022), 29.92 % (H1 2023)
Andel videoer fjernet av moderatører etter henstilling fra andre brukere (SLI 24.1.1)	60.00 % (Q4 2022), 14.66 % (H1 2023)	36.53 % (Q4 2022), 36.83 % (H1 2023)

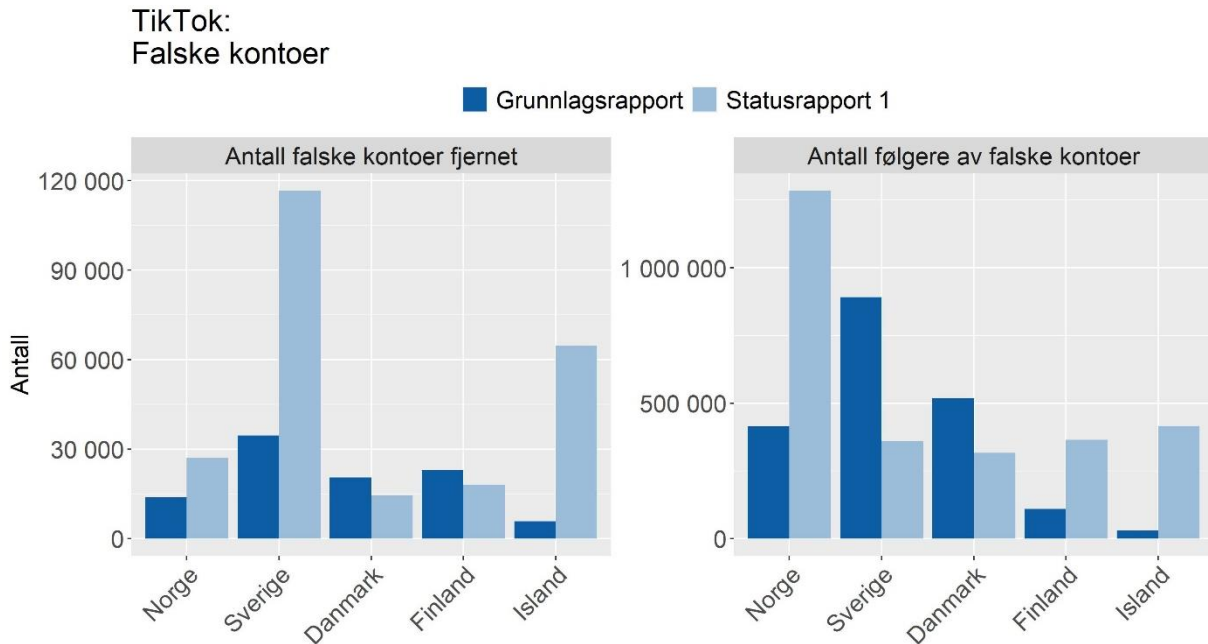
Kapittel 2. Kontroll av annonser. GR: TikTok oppgir å ha strenge retningslinjer for annonser, tilbyr ikke inntektsdeling med annonsører eller brukere, og rapporterer ikke på dette. Det er mulig for annonsører å bestemme hvilken kontekst annonsen skal vises i, ved TikToks innstillinger (Inventory Filter) og ulike former for sikker plassering (Brand Safety), med teknologi fra Zepr, IAS, DoubleVerify, TAG og NewsGuard (QRE 1.3.1, 1.4.1 og 1.5.1). TikTok planlegger å inkludere flere kategorier desinformasjon i brukervilkårene (QRE 2.1.1). Når en annonsør bryter retningslinjene, får annonsøren beskjed via TikToks Ad Manager, med klagerett (QRE 2.4.1), men TikTok har ikke rapportert på denne SLI-en. TikTok deltar i flere internasjonale nettverk for å redusere desinformasjon og finansieringen av den, som IFCN, GARM (QRE 3.1.1 og 3.2.1). SR1: TikTok har lagt til fire nye kategorier innhold som fjernes: Medisinsk feilinformasjon, farlig feilinformasjon, manipulert innhold, farlige konspirasjonsteorier (QRE 1.1.1).

⁴⁹ Tallet 0.0067 er rapportert under teksten "Fake accounts as a % of monthly active users". Medietilsynet har tolket dette som et desimaltall, og at korrekt prosent er 0.67 %.

Kapittel 3. Politiske annonser: GR: TikTok oppgir at politiske annonser er forbudt, med særlig strenge krav når det gjelder promotering av partier og politikere. TikTok tillater ikke politiske annonser, og rapporterer derfor ikke på dette (QRE 6.2.1). Det er imidlertid tillatt med «issue based» annonser og annonser fra myndighetsorganer og frivillige organisasjoner som ikke er politisk motivert (QRE 1.1.1 og 4.1.1). Skillet mellom politiske og temabaserte («issue based») annonser er imidlertid uklart. TikTok oppgir at det utarbeides en database for kommersielt innhold for å møte forpliktelsene etter DSA (QRE 10.2.1). SR1: Det er etablert en søkbar annonsedatabase, med et API for forskere, etter krav i DSA (QRE 10.2.1).

Kapittel 4. Integritet. GR: TikTok har innført tiltak mot ikke-autentiske kontoer (inkludert bots) og falske likes og reaksjoner, kontokapring og hatefulle ytringer (QRE 14.1.1 og 14.1.2). TikTok bruker en kombinasjon av teknologi og moderatorer i vurderingen av innholdet, og sletter kontoer som anses som ikke-autentiske (QRE 14.2.1). Selskapet har egne retningslinjer for å sørge for at KI-modellene er i overensstemmelse med regelverket (QRE 15.2.1). SR1: TikTok oppgir at det er gjort en omfattende gjennomgang av sine brukervilkår, og forbyr kunstig skapt innhold som ikke er merket, eller er redigert slik at innholdet er misvisende (QRE 14.1.1). Rapportering på antall falske kontoer og antall følgere av falske kontoer i Norden viser en sterk økning av falske kontoer i Norge, Sverige og Island. Det er også en sterk økning i antallet som følger falske kontoer i Norge og Island (se figur 12). Dette er oppsiktsvekkende tall, som bør kontrolleres nærmere for å finne årsaken, og tiltak for å redusere risikoen bør vurderes. Brukere av TikTok er i hovedsak barn, ungdom og unge voksne, og dette medfører et særlig behov for å følge med på risikoer på plattformen.

Figur 12: Antall falske kontoer fjernet (til venstre) og antallet som følger disse (til høyre) på TikTok oppgitt i grunnlagsrapporten (mørkeblått) og statusrapport 1 (lyseblått). Veksten i antall falske kontoer som er fjernet er svært høy for Sverige, Island og Norge. Når det gjelder antallet som følger falske kontoer, er veksten svært stor for Norge, Finland og Island. Legg merke til ulik y-akse. Tall justert til 181 dager. Merk den markante økningen i antallet som følger falske kontoer i Norge fra GR til SR1. Dette bør undersøkes nærmere. SLI 14.2.1 og SLI 14.2.2.



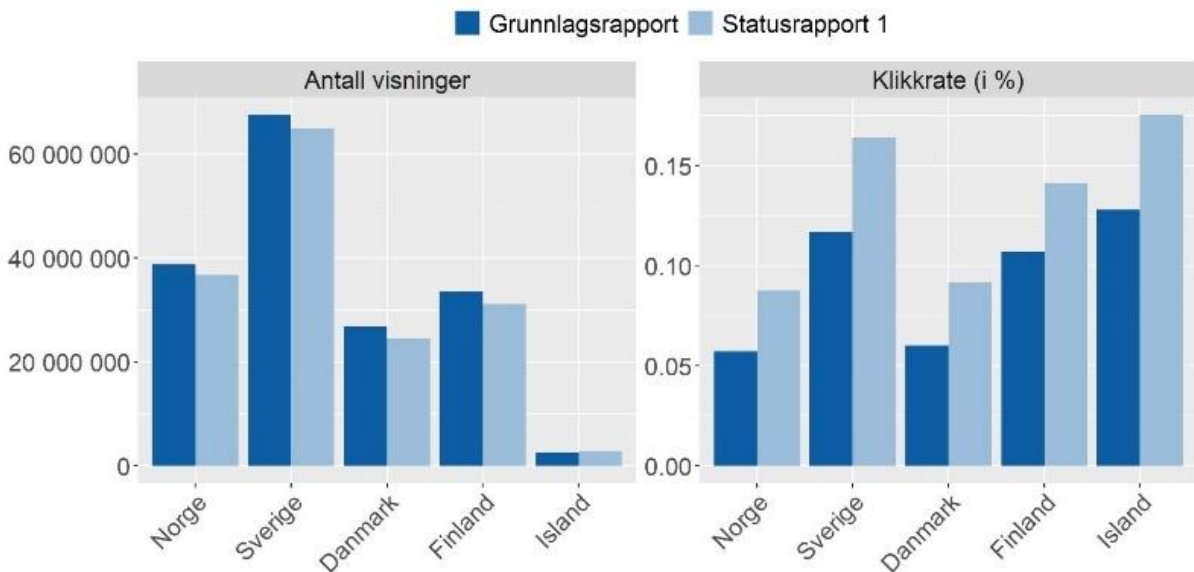
Kapittel 5. Tiltak for å styrke brukernes motstandskraft. GR: TikTok har siden 2020 merket desinformasjon om Covid, Holocaust, apekopper og (senere) krigen i Ukraina. Videoer med uverifisert innhold er merket og brukerne oppfordres til å oppsøke troverdig informasjon, med kobling til troverdige kilder for informasjon og offentlige kunngjøringer. Slike videoer tas ikke inn i feeden «for deg», for å begrense spredningen. Dette gjelder også innhold fra statskontrollerte medier, med vekt på russiske medier. TikTok har et eget sikkerhetscenter (QRE 17.1.1), og bidrar også med KMF-tiltak utenfor plattformen, særlig knyttet til demokratiske valg og tematikk som er særlig utsatt for feil- og desinformasjon (som Covid og Ukraina) (QRE 17.2.1 og 17.3.1). TikTok fjerner innhold på plattformen og har lagt inn flere sikkerhetskontroller for «for deg»-feeden. Innhold fjernes ved automatisk og manuell moderering. TikTok foretar stikkprøver innenfor enkelte tematiske områder, med vekt på videoer som «trender» og har også et samarbeid med faktasjekkere. Brukerne kan melde fra om desinformasjon.

TikTok har et samarbeid med Irrational Labs, som varsler brukerne om at de bør vurdere om innholdet de ønsker å dele er troverdig. TikTok hevder at sikkerhet er bygget inn i plattformens design (QRE 18.1.1). Retningslinjene forbyr feilinformasjon innen kategorier som hatefulle ytringer, nødsituasjoner, medisinske forhold, demokratiske valg og konspirasjoner. Innhold innen disse kategoriene fjernes. (QRE 18.2.1). TikTok orienterer om anbefalingssystemene i brukervilkårene og i sikkerhetscenteret. Hva brukerne ser i «for you»-feeden er basert på personalisert målretting, med et sett maskinlæringsmodeller for 1) hva brukeren oppgir å være interessert i eller ikke 2) hvilke innlegg brukeren liker og følger 3) hvor lenge en bruker ser en video innenfor en kategori 4) lokasjon og språkgruppe. TikTok har en funksjon for å forklare hvorfor en bruker eksponeres for innholdet («why this video?»).

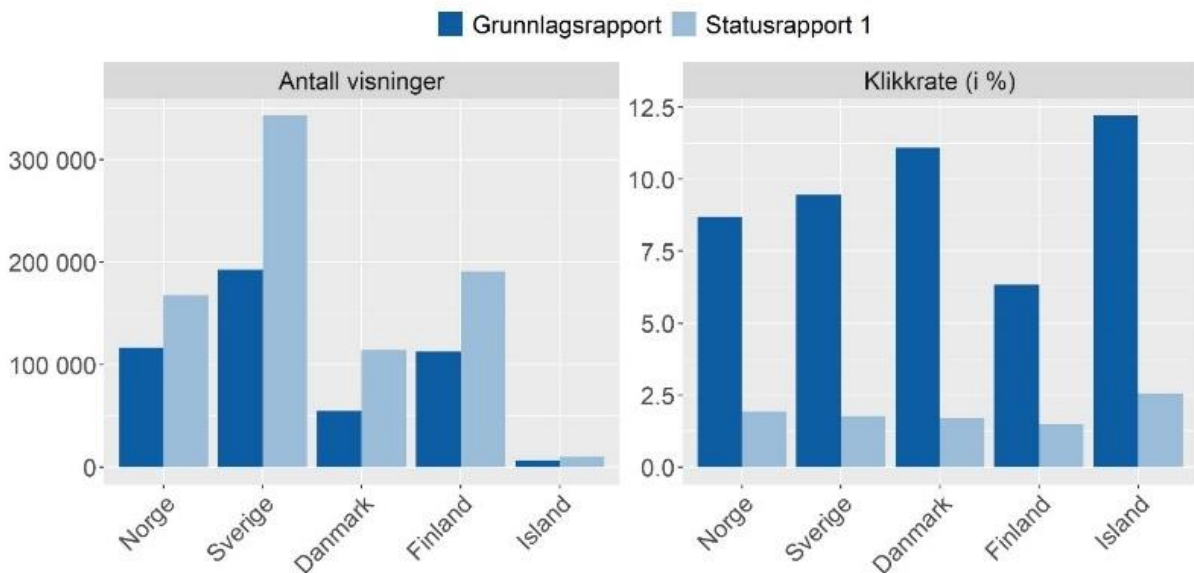
Brukerne kan også filtrere bort videoer innen innholdskategorier (fra juli 2022) og skru på en funksjon for å skjerme brukeren mot upassende innhold. TikTok oppgir at tjenesten forsøker å skjerme brukere under 18 år, og vil utforske dette ytterligere i lys av DS QRE 19.2.1). TikTok har åtte IFCN-godkjente faktasjekkere, som dekker ti av de offisielle språkene i EU. Det er etablert nye samarbeid med faktasjekkere i Sverige, Ungarn, Ungarn, Polen og Romania, og det er også opprettet et team for moderator i disse landene (QRE 21.1.1). Det skal også opprettes tilsvarende team i Danmark, Portugal, Hellas og Belgia. Norge er ikke nevnt. TikTok har valgt å ikke underskrive forpliktelse 22 om å lede brukere til troverdig innhold, da det vurderes at dette dekkes av øvrige forpliktelser. TikTok har en løsning for å flagge innhold som bryter med brukervilkårene (QRE 23.1.1) og en klagemekanisme (QRE 23.2.1).

Figur 13: Videomerking («Video Notice Tags»; øverst): Antall visninger av videoer merket med tagger som viser videre til troverdige kilder for informasjon om covid-19 og covid-19-vaksiner, holocaustfornektelse og apekopper (til venstre), og raten av visningene som førte til at brukere klikket seg gjennom taggene (til høyre). Varsel i søkeresultat («Search Interventions»; nederst): antall visninger av et informasjonsvindu med oppfordring til å verifisere fakta og med lenker til ressursider, som dukker opp ved søk knyttet til temaene covid-19 og covid-19-vaksiner, holocaustfornektelse, apekopper og klimaendringer (til venstre), og raten av visningene som førte til klikk på informasjonspanelet (til høyre). Tall justert til 181 dager. SLI 17.1.1.

TikTok:
Videomerking



TikTok:
Varsel i søkeresultat

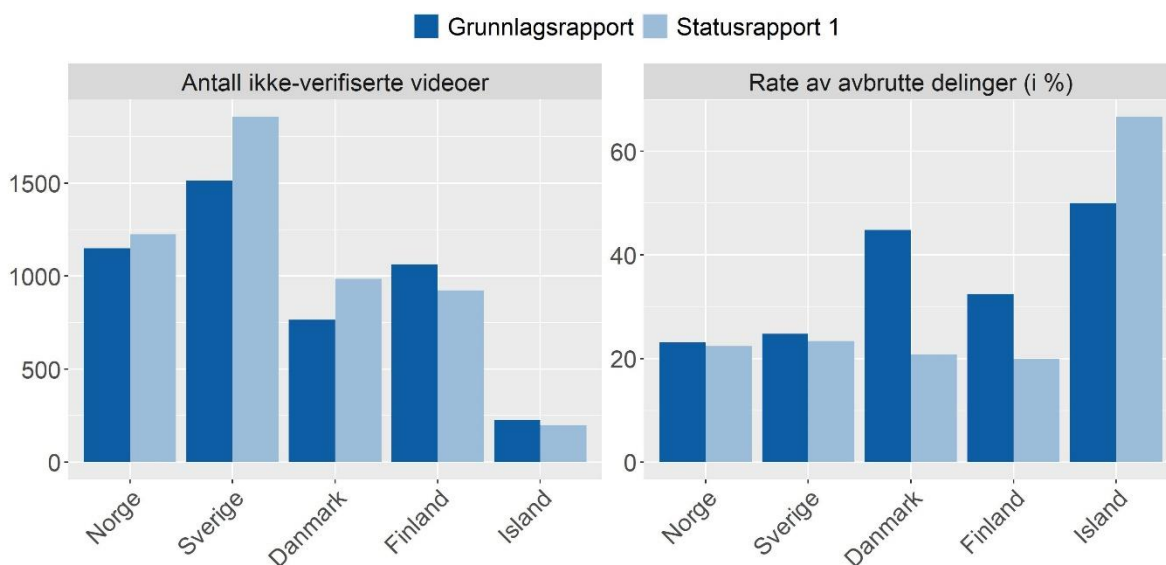


SR1: I TikToks oppdatering av brukervilkårene er det lagt vekt på å informere om tiltak mot uønsket påvirkning av demokratiske valg. Innhold fra statskontrollerte medier merkes i større grad enn før. Det henvises til troverdige kilder om Covid, Holocaust, krigen i Ukraina og klima på 25 språk, inkludert norsk og islandsk. Det samme gjelder ved demokratiske valg (QRE 17.2.1). Eksempelvis ble Medietilsynet kontaktet av TikTok om formidling av Stopp.Tenk.Sjekk-kampanjen før kommunevalget i september, men TikTok har ikke orientert om hvorvidt kampanjen ble brukt eller om eventuelle effekter av den. Innhold fra RT, Sputnik og andre russiske medier som er omfattet av EU-sanksjoner er begrenset eller fjernet (17.1.1), og det er inngått samarbeid med 50 eksperter i 20 land om merking av innhold fra statskontrollerte medier (QRE 21.3.1). Samarbeid med faktasjekkere er etablert i ni land siden GR, blant annet i Norge, men det fremgår ikke hvilken faktasjekker som benyttes i Norden og Norge (18.1.1). Medietilsynet er kjent med at TikTok har engasjert Logically. TikToks modell for faktasjekk gjør at faktasjekkerollen ligner moderatorrollen, da faktasjekket innhold i større grad fjernes enn merkes. TikTok oppgir at de har oppgradert maskinlæringsmodellene (KI) mot «skadelig feilinformasjon», blant annet i Norge, forankret i oppdaterte brukervilkår. Som en følge av DSA, vil TikTok innføre muligheten for å velge en feed som ikke er basert på anbefalingsalgoritmer (QRE 19.1.1). Brukerne har mulighet til å varsle om feilinformasjon, der de nye kategoriene i brukervilkårene er tatt med. Hvis innhold begrenses, gis det mulighet til å klage, som er en plikt etter DSA. For varsling av ulovlig innhold er det opprettet en egen rapporteringskanal, som også er en konsekvens av DSA, og det er opprettet et eget moderatorsteam i TikTok for å behandle sakene (QRE 23.2.1 og 24.1.1). TikTok skriver at moderatorene skal vurdere om innholdet er ulovlig.

Figur 14: Antall TikTok-videoer merket med ikke-verifisert innhold (til venstre) og raten av delinger av disse som ble avbrutt etter tilsynkomst av merking som ikke-verifisert innhold (til høyre). Raten for EØS totalt er rapportert til 28.39 % (grunnlagsrapport) og 29.92 % (statusrapport 1). Tall justert til 181 dager. SLI 21.1.1 og SLI 21.1.2.

TikTok:

Ikke-verifiserte videoer





Kapittel 6. Tilgang for forskere. TikTok oppgir at det utvikles en API for å gi tilgang til forskere (QRE 26.1.1). TikTok har et eget sikkerhetsråd i Europa, som består av ledere fra akademia og organisasjonslivet. Det er også etablert partnerskap med forskere, med vekt på Storbritannia (QRE 28.1.1).

Kapittel 7. Tilgang for faktasjekkere. TikTok har utarbeidet en database for faktasjekket innhold og lager også trendrapporter. Enkelte faktasjekkere har mandat til å undersøke utbredelsen av desinformasjon på TikTok og andre plattformer. TikTok har egne retningslinjer for moderatorene (QRE 30.1.3). TikTok har standardiserte avtaler for faktasjekkere (QRE 30.2.1). Rapporten viser at TikTok ikke har faktasjekket innhold i Norge, og at antall faktasjekkede videoer er svært begrenset i andre land (448 i Sverige, der TikTok har satset på dette).

Google

GR: kvalitativt 2 , kvantitativt 2 . Antall sider: 229.

SR1: kvalitativt 2 , kvantitativt 2 . Antall sider: 253.

Rapportering fra Norge: Ja 

Vurdering: Google rapporterer for søkemotoren Google Search (Søk), videodelingsplattformen YouTube og reklametjenesten Google Advertising (Annonser). YouTube og Search er definert som VLOP/VLOSE i DSA. Rapportene er omfattende, kvalitativt og kvantitativt, også på norsk nivå, men uten å dekke alle parametere. Rapportene har gode, detaljerte og ofte lett forståelige tall, der EØS/EFTA-landene er inkludert. GR inneholder feil rapporteringsperiode (kvartalstall), men dette er endret i statusrapporten. Microsoft rapporterer for tre plattformer, men gjennomgående uten å rapportere SLIer for alle tre. Noen tall er lagt inn i den kvalitative rapporteringen, som kan være lett å overse. Google har flere konkrete tiltak for å forhindre forsterket algoritmisk og finansiert forsterkning av desinformasjon. Den kvalitative rapporteringen bærer preg av endel gjentakelser.

Relevante funn:

- Maskinlæringsmodellene i Search og i YouTube skal rangere og prioritere troverdig innhold. Desinformasjon motvirkes primært ved at anbefalingssystemene forsterker innhold av høy kvalitet.
- 35 826 YouTube-kanaler var involvert i 44 separate aksjoner i koordinerte påvirkningskampanjer i rapporteringsperioden for SR1 (januar-juni 2023).
- Over 500 timer innhold lastes opp på YouTube hvert minutt, som krever en kombinasjon av automatiserte og manuelle vurderinger av innhold.
- I 2022 avbrøt Google 1950 tilfeller av skjulte russiske informasjonsoperasjoner på plattformene sine. 90 prosent av innholdet var på russisk, som sier noe om målgruppen for operasjonene.
- Google har distribuert en ny kampanjevideo for å motvirke russisk desinformasjon om Ukraina.

Tabell 4: Utvalgte nøkkeltall fra Google-rapportene. Legg merke til ulik rapporteringsperiode; i grunnlagsrapporten er det hovedsakelig rapportert for et kvartal (Q3), mens det i den første statusrapporten er rapportert for et halvår (H1). SR1 inneholder to nye kategorier, manipulerende praksis og sjokkerende innhold.

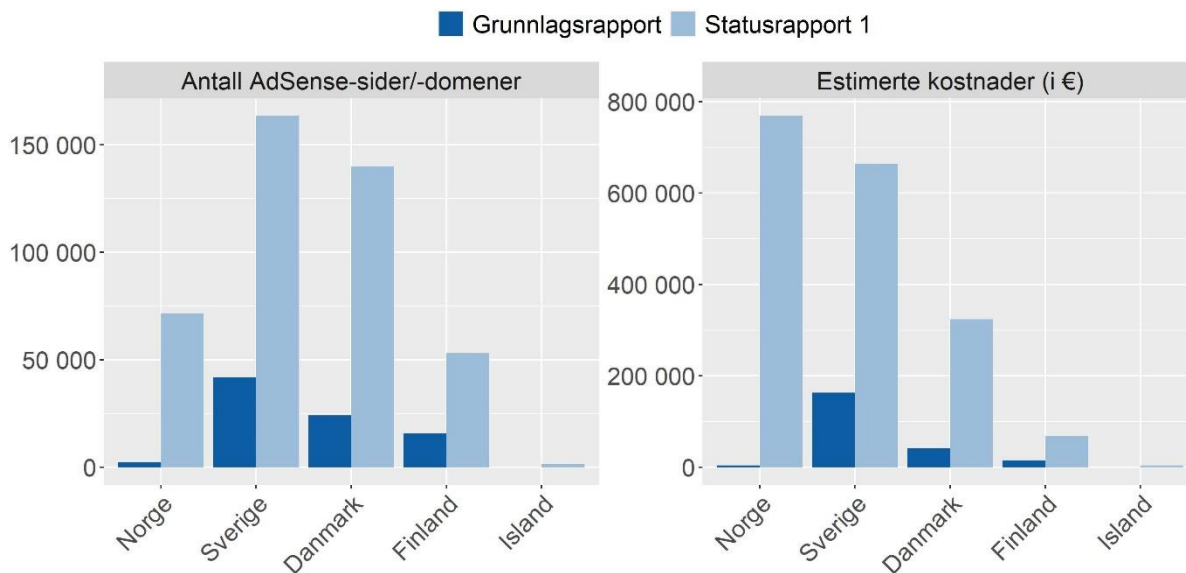
Nøkkeltall, Google

KPI	Norge	EØS
<i>Google Advertising</i>		
Antall AdSense-sider det er gjort tiltak mot grunnet brudd på desinformasjonsreglement (SLI 1.1.1)	1 196 (Q3 2022), 71 323 (H1 2023)	2.9 mill. (Q3 2022), 20.2 mill. (H1 2023)
Antall AdSense-domener det er gjort tiltak mot grunnet brudd på desinformasjonsreglement (SLI 1.1.1)	14 (Q3 2022), 16 (H1 2023)	1 662 (Q3 2022), 2 820 (H1 2023)
Verdi av blokkerte bud (AdSense-sider) (SLI 1.1.2)	1 765 € (Q3 2022), 768 250 € (H1 2023)	12.8 mill. € (Q3 2022), 30.2 mill. € (H1 2023)
Verdi av blokkerte bud (AdSense-domener) (SLI 1.1.2)	182 € (Q3 2022), 138 € (H1 2023)	503 587 € (Q3 2022), 1.9 mill. € (H1 2023)
<i>Google Search</i>		
Antall brudd på desinformasjonsreglement som førte til fjerning av innhold (SLI 14.2.1)	N/A	400 (globalt, Q3 2022), 41 840 (globalt, H1 2023)
Antall visninger av ulike attributter som skal hjelpe brukeren å evaluere informasjon (SLI 17.1.1)	1.46 mill. (Q3 2022), 6.8 mill. (H1 2023)	102.6 mill. (Q3 2022), 900.7 mill. (H1 2023)
Antall faktasjekkartikler tilgjengelig (SLI 21.1.1)	447 (Q3 2022), 481 (H1 2023)	122 570 (EØS, Q3 2022), 211 154 (globalt, H1 2023)
<i>YouTube</i>		
Antall videoer fjernet grunnet brudd på desinformasjonsreglement (SLI 18.2.1)	Over 80 (Q3 2022), 124 (H1 2023)	Over 11 000 (Q3 2022), 16 558 (H1 2023)
Antall kanaler involvert i brudd på desinformasjonsreglement (SLI 14.2.1)	N/A	Over 10 000 (globalt, Q3 2022), 7 366 92 (globalt, H1 2023)
Antall visninger av informasjonspaneler ved siden av søkeresultat (SLI 17.1.1)	N/A (grunnlagsrapport), 4.2 mill. (H1 2023)	1.5 mrd. (globalt, Q3 2022), 4.0 mrd. (EØS, H1 2023)
Antall visninger av kritisk medieforståelse-kampanjen «Hit Pause» (SLI 17.2.1)	N/A (grunnlagsrapport), 2.8 mill. (H1 2023)	Over 180 mill. (22.11 – 15.12.22), 408 mill. (H1 2023)

Kapittel 2. Kontroll av annonser: GR: Google tilbyr annonseinntektsdeling gjennom programmet AdSense, ved at nettsiden der annonsen publiseres får en prosentandel av annonseinntektene. Google Advertising har flere tiltak for å hindre inntjening basert på spredning av desinformasjon, ut fra flere kategorier: 1) Innhold som er beviselig galt eller i betydelig grad egnet til å underminere deltakelse i eller tillit til en demokratisk prosess/valg. 2) Manipulerende innhold relatert til politikk, sosiale spørsmål, saker av offentlig betydning (QRE 1.1.1). 3) Visse typer annonser knyttet til klimaendringer, Covid, krigen i Ukraina (QRE 1.2.1). Annonsører kan selv velge vekk visse typer innhold de ikke ønsker å annonsere ved siden av (QRE 1.3.1./1.6.1). Annonseinhold kan ikke være villedende, det må fremgå hvem som står bak, og det er ikke tillatt med manipulasjon for å svindle eller villedende rundt politiske spørsmål.

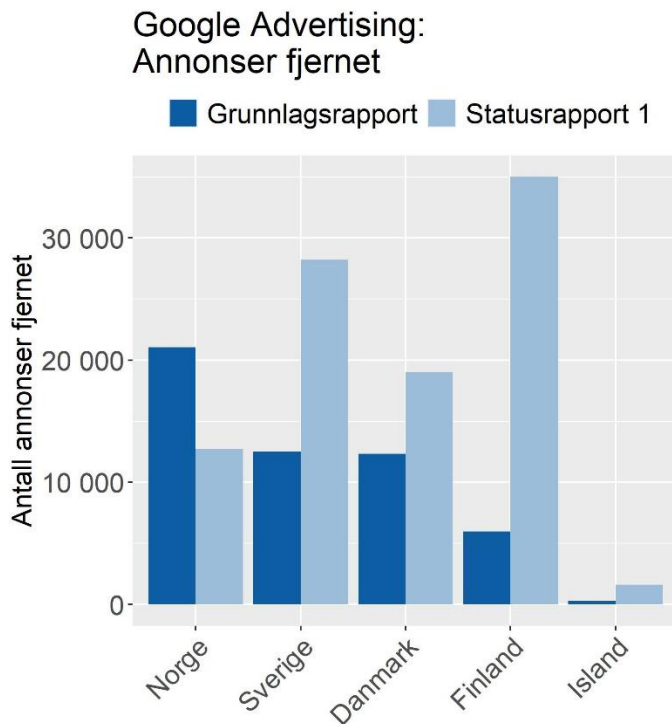
Figur 15: Antall AdSense-sider og -domener (venstre) hvor tiltak mot desinformasjon er gjort hos Google Advertising. Estimerte kostnader av de blokkerte spørringene er gitt i grafen til høyre. Tall justert til 181 dager. SLI 1.1.1.

Google Advertising:
AdSense-sider og -domener med tiltak mot desinformasjon



(SR1): Tiltakene er i stor grad videreført. Kvalitativ rapportering er utvidet med informasjon om innhold sanksjonert for villedende praksis og sjokkerende innhold. Det rapporteres nå på riktig tidsintervall (seks måneder).

Figur 16: Antall annonser fjernet grunnet brudd på desinformasjonsregelverk på Google Advertising. Tall justert til 181 dager. SLI 2.1.1.



Kapittel 3. Politiske annonser: GR: Google tilbyr politiske annonser, men med visse restriksjoner: 1) Annonser som oppfordrer til hat og diskriminering basert på etnisk opphav, religion, seksuell orientering eller andre karakteristika assosiert med systematisk diskriminering eller marginalisering. 2) Politisk reklame med skjult/villedende avsender. Valgannonser innenfor EU er kun tillatt fra verifiserte annonsører⁵⁰. Google avventer forordningen for politisk reklame (QRE 4.1.1).⁵¹ Google Ads har iverksatt nye retningslinjer som forbyr annonser som utnytter sensitive hendelser eller gir tilgang til hacket materiale knyttet til politiske aktører. 3) Annonser som omhandler politikk, samfunnsproblemer eller offentlige saker, må nå oppgi riktig identitet og vesentlig informasjon. (QRE 5.1.1.).

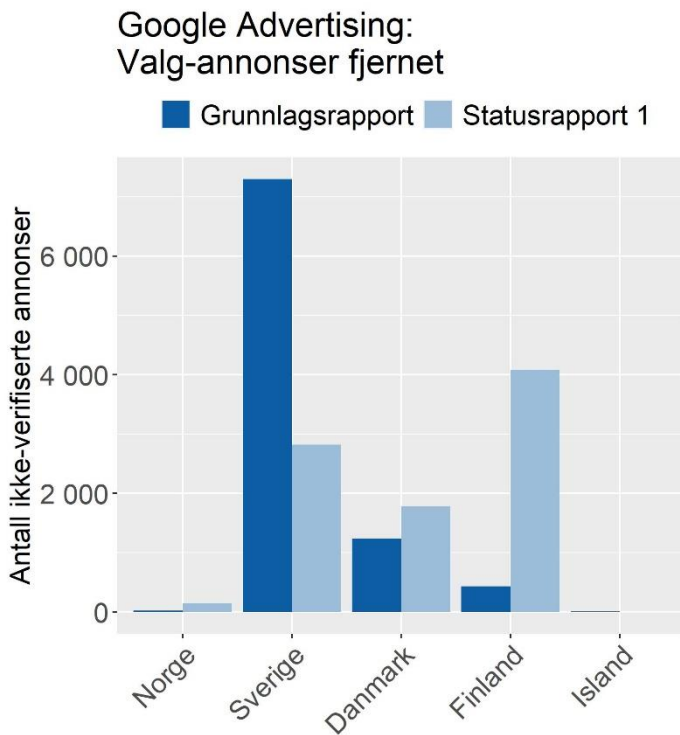
Google rapporterer at det kreves åpenhet rundt hvem annonsørene er og hvor de befinner seg. For de fleste annonseformater blir det automatisk generert en «betalt av»-merking basert på verifisert informasjon. Noen annonsesjangere og funksjoner krever at annonsørene merker med «betalt av» direkte i annonsen (QRE 6.2.1). Google har et annonsebibliotek for politiske annonser, og utarbeider en rapport om EU-valgannonser, inkludert hvor de er vist og hvordan de var målrettet mot kjønn, alder og geografisk plassering. Rapporten er tilgjengelig for alle (QRE 8.1.1).

SR1: Google viderefører sin praksis for politiske annonser og avventer EUs forordning for politisk reklame.

⁵⁰ <https://support.google.com/adspolicy/answer/6014595?sjid=4769325654731154241-EU#zippy=%2CEuropean-union-eu-election-ads>

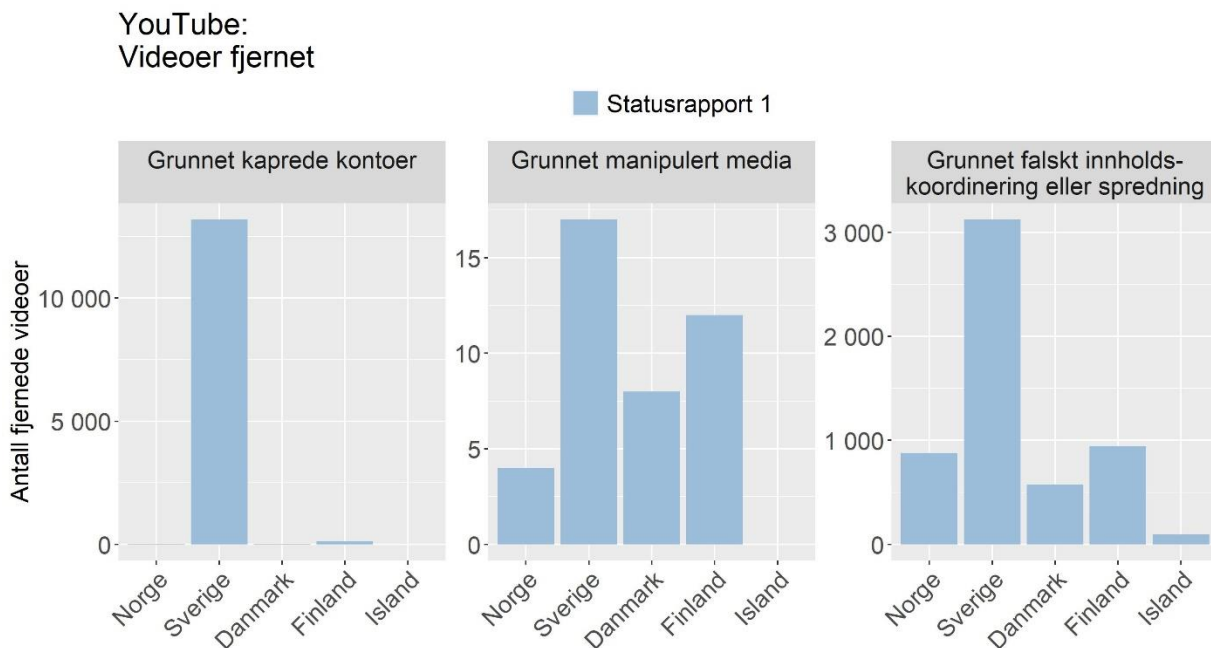
⁵¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0731>

Figur 17: Antall annonser knyttet til valg i EU som ble fjernet på grunn av ikke-verifisert annonsør på Google Advertising. Tall justert til 181 dager. SLI 7.1.1.



Kapittel 4. Integritet: GR: Maskinlæringsmodellene er designet for å fremheve troverdig informasjon. Google viser til at bransjenormens fremhevede teknikker, taktikker, praksiser (TTP) for desinformasjonskampanjer, slik som bot-er, falske sider, kontokapring eller falske likerklikk, ikke er relevante for en søkemotor (QRE 14.1.1). Search bruker rangeringssystemer for å fremme pålitelig informasjon, inkludert tiltak mot falske domener, manipulering, falske medier og andre villedende praksiser.

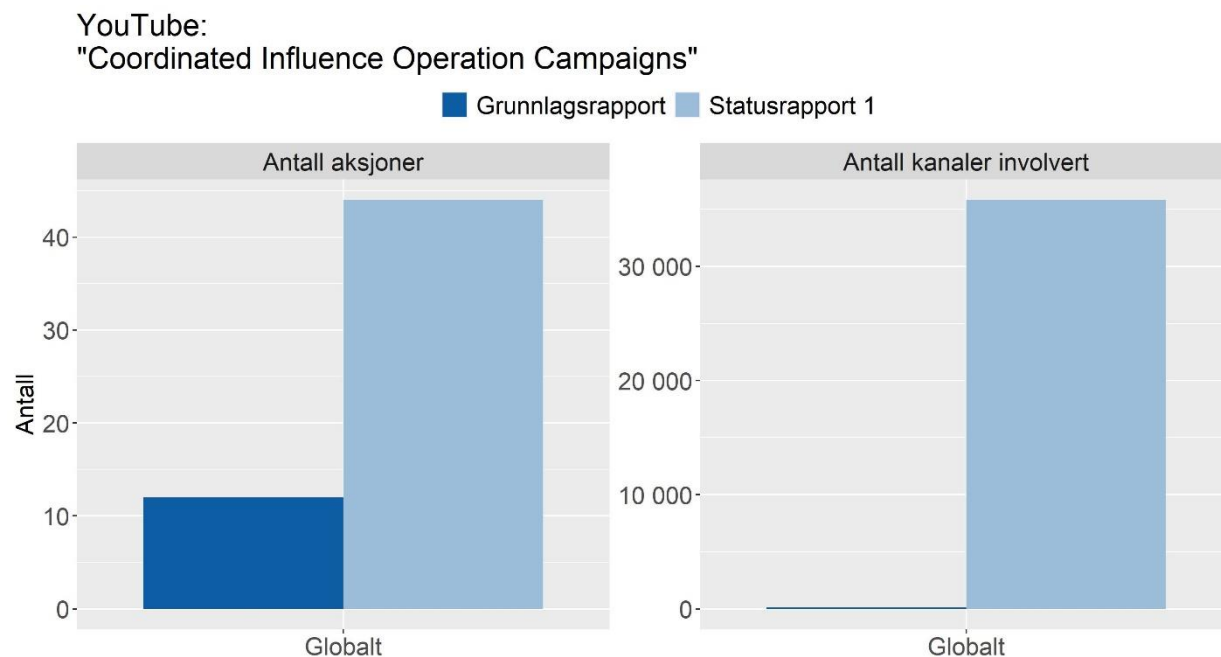
Figur 18: Antall fjernede videoer på YouTube på grunn av kaprede kontoer (til venstre), manipulert media (i midten) og falsk koordinering eller spredning av innhold (blant annet forsøk på å manipulere plattformenes algoritmer; til høyre). Legg merke til ulik y-akse. Tall kun rapportert i statusrapporten (dekker 181 dager). SLI 14.2.1; TTP 5, TTP 7 og TTP 9.



Search videreutvikler KI-verktøy mot spam (SpamBrain). Dette favner flere typer innhold av lav kvalitet, men har også til hensikt å fange opp desinformasjon. Manipulert medieinnhold brukt til å villede, svindle eller lure ved å gi feilaktig inntrykk av hendelser eller hendelsesforløp som ikke har funnet sted, er forbudt (QRE 14.1.1./14.1.2). YouTube har maskinlæringsmodeller for å fremme innhold fra troverdige kilder, og retningslinjer som dekker innholdet i mange av de nevnte TTPer. Disse dekker imidlertid ikke sentrale virkemidler som ikke-autentiske sider, grupper eller domener, eller bevisst målretting av innhold mot sårbare mottakere. YouTube viser til at menneskelige moderatører og KI sikrer effektiv moderering, i kombinasjon med flagging av skadelig innhold. Varsler fra sertifiserte varslere (trusted flaggers), som myndighetskontakter og organisasjoner, prioriteres. Effekten er vanskelig å anslå på grunn av svak rapportering. Google har et eget team (Threat Analysis Group - TAG) som overvåker koordinerte informasjonsoperasjoner, og henviser til egne rapporter⁵² i stedet for rapportering etter bransjenormen. Det er utarbeidet egne KI-prinsipper for ansvarlig utvikling av kunstig intelligens. De skal forhindre villedende bruk av KI-manipulert innhold (QRE 15.1.1.) I et case-eksempel beskrives hvordan russiske myndigheter benytter et bredt spekter av informasjonsoperasjoner, fra åpne, statlig støttede medier til skjulte plattformer og kontoer, for å forme den offentlige oppfatningen av krigen.

⁵² <https://blog.google/threat-analysis-group/tag-bulletin-q1-2023/>

Figur 19: Antall aksjoner og antall kanaler involvert i YouTube-kampanjene «Coordinated Influence Operation Campaigns». Tall justert til 181 dager. SLI 16.1.1.



SR1: Google problematiserer at bransjenormen ikke skiller mellom rapporteringskrav for søkemotorer og plattformer, særlig når det gjelder TTPer, men oppgir en ambisjon om å rapportere i større detalj i fremtiden (QRE 14.1.1). Det har vært en sterk økning i antall YouTube-kontoer brukt i koordinerte påvirkningsoperasjoner (SLI 16.1.1). Search vil etablere en ny tjeneste for å kode metadata på bilder som genereres av Google AI (QRE 15.1.1). En ny rapport om informasjonstrusler i kjølvannet av Ukraina-krigen er publisert (QRE 16.1.1), med gode eksempler på samarbeid for å overvåke desinformasjonskampanjer på YouTube (QRE 16.2.1).

Kapittel 5. Tiltak for å styrke brukernes motstandskraft. **GR:** Merking av «About this result» er utvidet til flere språk.⁵³ Funksjonen gir tilgang til informasjon om søkeresultat og kilder, men er ikke intuitiv for alle. Ulike verktøy gir mer kontekst til søkeresultatets troverdighet. (QRE 17.1.1). Search har lansert et opplæringsprogram i kritisk medieforståelse høsten 2022, sammen med offentlige biblioteker, kalt «Super Searchers» (QRE 17.2.1).

YouTube-kampanjen «Hit Pause», mot spredning av desinformasjon, er distribuert i 20 EU-land i 2022, og skal spres i alle EØS-land i 2023.⁵⁴ YouTube har egne retningslinjer⁵⁵ mot medisinsk villedende informasjon og desinformasjon som kan forstyrre demokratiske prosesser. Search viser til troverdig kontekst gjennom Fact Check Explorer, som henviser videre til faktasjekkhet innhold.⁵⁶ Brukere kan dessuten varsle om krenkende innhold (QRE 23.1.1). Det er en kombinasjon av maskinell og menneskelig moderering av innhold (over 500 timer innhold lastes opp på YouTube hvert minutt).

⁵³ <https://blog.google/products/search/about-search-results/>

⁵⁴ <https://www.youtube.com/@HitPause>

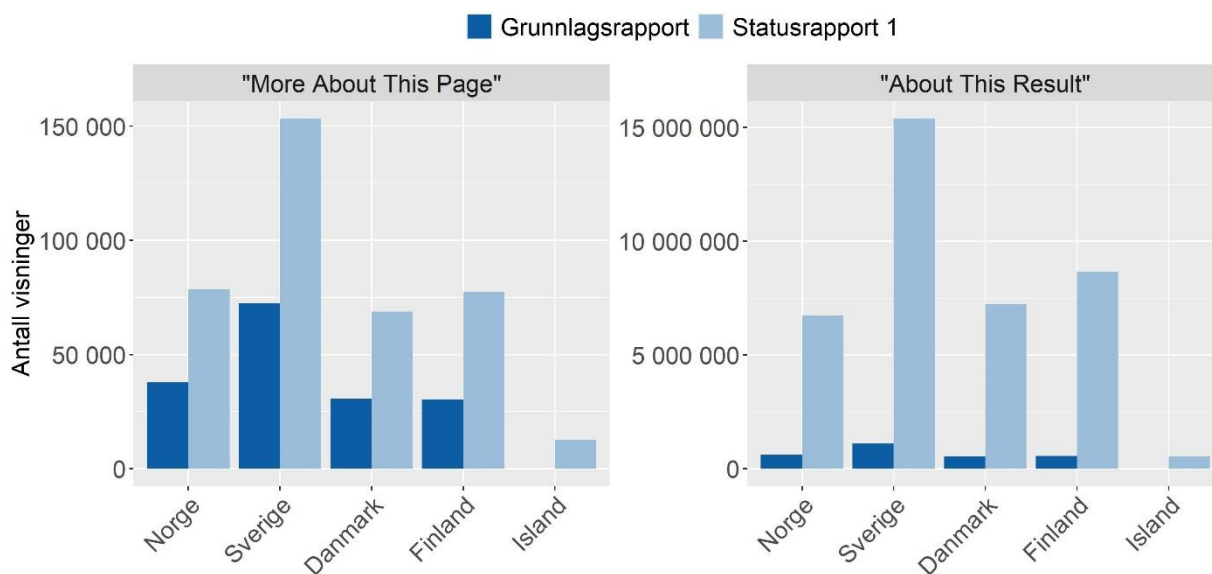
⁵⁵ <https://www.youtube.com/howyoutubeworks/policies/community-guidelines/>

⁵⁶ <https://toolbox.google.com/factcheck/explorer>

SR1: Det rapporteres langt bedre enn i statusrapporten en i grunnlagsrapporten om YouTube's fjerning av desinformasjon som bryter med brukervilkårene.⁵⁷ Tjenesten «About this result» er utvidet, og det rapporteres på utbredelse av Hit Pause (SLI 17.2.1.). Tall på landnivå er tatt med. Avdelingen Jigsaw forsker på informasjonsangrep, og har laget en ny kampanje om å motvirke desinformasjon rettet mot polakker, tsjekkere og slovakere knyttet til krigen i Ukraina.

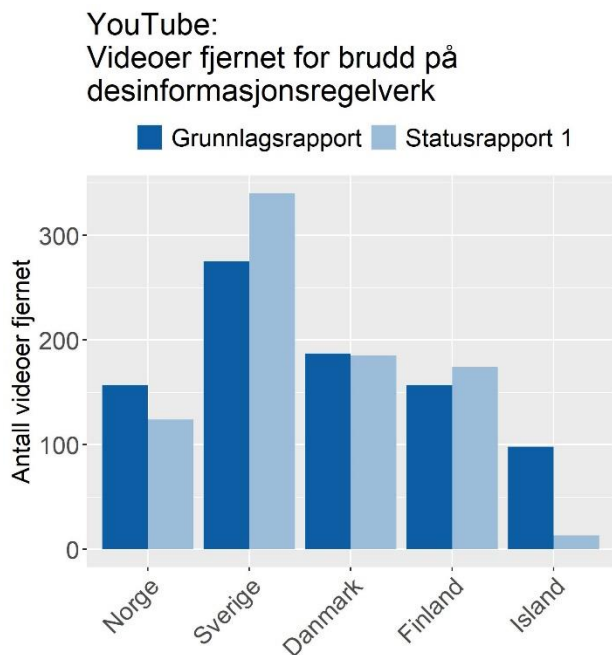
Figur 20: Antall visninger for to ulike verktøy for å øke kritisk medieforståelse i Google Search: «More About This Page»-lenke til tilleggsinformasjon (til venstre) og «About this result»-panel (til høyre). Tall justert til 181 dager. SLI 17.1.1.

Google Search:
Visninger av verktøy for å øke kritisk medieforståelse

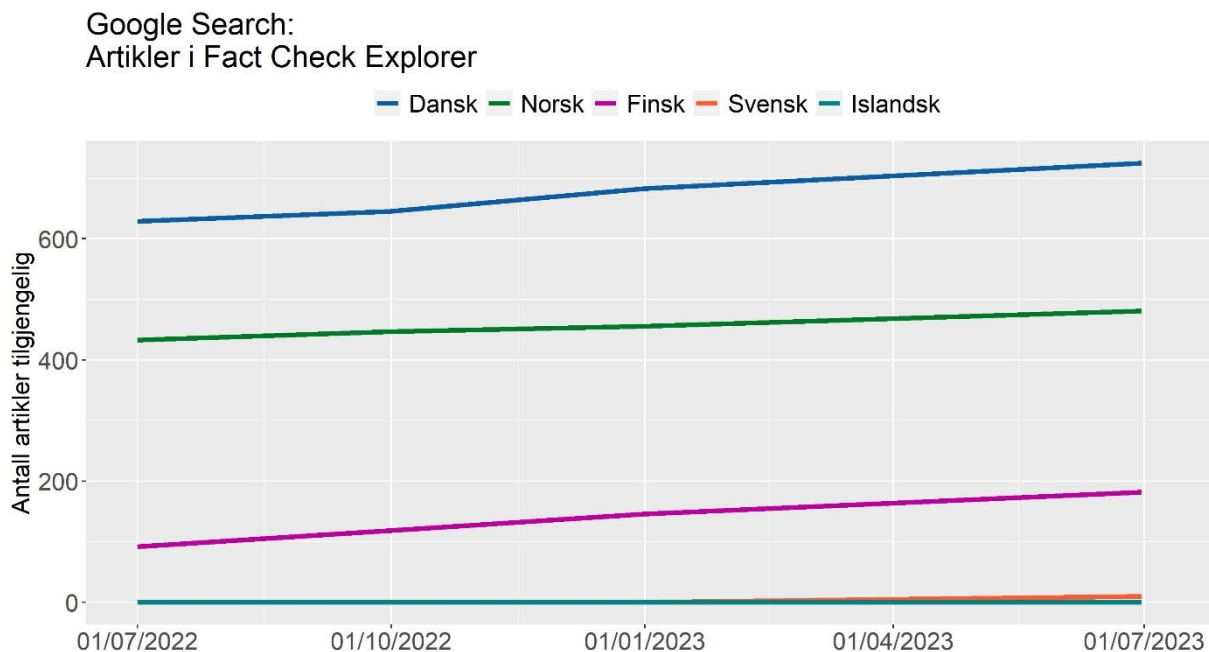


⁵⁷ https://support.google.com/youtube/topic/10833358?hl=en&ref_topic=2803176

Figur 21: Antall videoer fjernet grunnet brudd på desinformasjonsregelverk¹ på YouTube. Tall justert til 181 dager. SLI 18.2.1.



Figur 2222: Antall artikler tilgjengelig i Fact Check Explorer i Google Search på ulike nordiske språk. Google innrapporterte at ingen artikler er tilgjengelig på islandsk. Totalt antall artikler tilgjengelig globalt var 211 154 i slutten av første halvår 2023. Fact Check Explorer er en egen søkedatabase som lar brukeren søke i faktasjekket informasjon om et emne. SLI 21.1.1.



Kapittel 6. Tilgang for forskere. GR: Rapporter fra Search og YouTube er tilgjengelig på Google Trends (representativt utvalg søketermer). Search tilbyr også verktøy som Fact Check Explorer, med eget API for oversikt over faktasjekkede sider/artikler (26.1.2). YouTube har i

2022 opprettet et eget forskerprogram som skal tilby forskere data, verktøy og støtte.^{58 59} Google har også API-ene Fact Check Explorer og FactCheck Claim Search, med oversikt over artikler vurdert, oppsummering, vurdering m.m. (QRE 26.2.1).⁶⁰




SR1: Oppdatering om deltakelse i European Digital Media Observatory (EDMO) Working Group for the Creation of an Independent Intermediary Body. Den nye arbeidsgruppen ble lansert i mai 2023, og er ment å spille en rådgivende rolle for Bransjenormen og bestemmelsen om forskertilgang i henhold til Artikkel 40 i DSA⁶¹ (27.1.1.).

Kapittel 7. Tilgang for faktasjekkere. GR: Google samarbeider ikke med norske faktasjekkere direkte, men indirekte gjennom IFCN – International Fact Checking Network⁶². Faktisk.no er medlem av dette nettverket. Google annonserte i 2022 at selskapet sponser IFCN med \$13.2M USD over 2,5 år.⁶³ Søk og YouTube sanksjonerer ikke innlegg som er faktasjekket som feilinformasjon, og rapporterer ikke på denne indikatoren. Google har bidratt med 25 millioner euro til European Media and Information Fund, opprettet for å fremme kritisk medieforståelse og bekjempe desinformasjon (QRE 30.1.3) SR1: Google redegjør for landspesifikke tiltak (ingen norske) og direkte samarbeidspartnere (ingen norske).

Microsoft

GR: kvalitativ 2 , kvantitativ 2 . Antall sider: 207.

SR1: kvalitativ 2 , kvantitativ 2 . Antall sider: 190.

Rapportering fra Norge: Ja . (Advertising og Bing), Nei /Ja  (LinkedIn, som mangler norske tall i GR, men har inkludert norske tall i SR1).

Vurdering: Microsoft rapporterer for Microsoft Advertising, søkemotoren Bing og nettverksplattformen LinkedIn. De to sistnevnte tjenestene er kategorisert som VLOP/VLOSE i DSA. Rapporteringen er omfattende og grundig, dels på landnivå, og flere egne initiativer som overlapper med målene i bransjenormen, blant annet en tydelig satsing på å oppdage og analysere utenlandske informasjonsoperasjoner. Microsoft har endret ordlyd på SLIene (f.eks. SLI 1.1.1.), som gjør sammenligning på tvers av ulike rapporter utfordrende. Det rapporteres på norsk nivå for søketjenesten Bing og Advertising, men ikke for LinkedIn i grunnrapporten. Det er til dels svært mangelfull rapportering for LinkedIn i GR, men den er betydelig forbedret i SR1. Rapportperiode varierer mellom 1. -31. desember 2022 eller medio desember 2022 – medio januar 2023 (GR.)

⁵⁸ <https://research.youtube/>

⁵⁹ <https://trends.google.com/trends/>

⁶⁰ <https://toolbox.google.com/factcheck/apis>

⁶¹ <https://edmo.eu/2023/05/15/launch-of-the-edmo-working-group-for-the-creation-of-an-independent-intermediary-body-to-support-research-on-digital-platforms/>

⁶² <https://blog.google/outreach-initiatives/google-news-initiative/how-google-and-youtube-are-investing-in-fact-checking/>

⁶³ <https://www.poynter.org/fact-checking/2022/google-and-youtube-partner-with-poynters-international-fact-checking-network-on-13-2-million-grant-for-the-global-fact-checking-community/>

Relevante funn:

- Kvalitetssikret innhold fra troverdige kilder prioriteres i søkeresultatene i Bing, men uten å fjerne søkeresultater som vurderes som falske, som Microsoft understreker vil være i strid med ytringsfrihet.
- Fire grunnleggende prinsipper ligger til grunn for Microsofts tjenester: Ytringsfrihet, prioritering av troverdig innhold (kvalitetssikret av tredjepart), avstå fra profitt fra utenlandske påvirkningsoperasjoner, og forebygging av slike operasjoner.
- LinkedIn forbyr desinformasjon og feilinformasjon både som organisk og sponset innhold.
- Microsoft tilbyr uavhengige faktasjekkere å merke faktasjekket innhold som troverdig/ikke troverdig gjennom "ClaimReview"-merking. Det rapporteres ikke om samarbeid med faktasjekkere i Norge.
- Microsoft identifiserer og blokkerer politiske annonser fra sine annonsenettverk, inkludert begrensninger fra visse domener. Det er utarbeidet en liste over politiske partier, kandidater og avstemninger, som Microsoft ikke tillater kjøp av annonser eller økt synlighet i søkeresultat.
- Det er innført flere retningslinjer og verktøy etter lansering av KI-drevne Bing Chat.
- Mellom juli og desember blokkerte eller fjernet LinkedIn rundt 58,1 millioner falske kontoer (globalt). 87,4 % av falske kontorer ble blokkert av automatiserte systemer, 12,6 % stoppet av manuelle undersøkelser, og til sammen 99,6 % av de falske kontoene ble stoppet før et medlem av LinkedIn varslet.⁶⁴

⁶⁴ <https://about.linkedin.com/transparency/community-report#fake-accounts>

Tabell 5: Utvalgte nøkkeltall fra Microsoft-rapportene. Legg merke til ulike rapporteringsperioder; i grunnlagsrapporten er stort sett rapportert på en måned, mens det i den første statusrapporten er rapportert for et halvår.

Nøkkeltall, Microsoft

KPI	Norge	EØS
<i>Microsoft Advertising</i>		
Antall sider med annonser det ble gjort tiltak mot grunnet brudd på desinformasjonsreglement (SLI 1.1.1)	4 (16.des-16.jan 2022), 19753 (H1 2023)	9 334 (16.des 22-16.jan 23), 1.6 mill. (H1 2023)
Antall domener med annonser det ble gjort tiltak mot grunnet brudd på desinformasjonsreglement (SLI 1.1.1)	N/A	96 (globalt; 16.des 22-16.jan 23), 1 007 (H1 2023)
Antall unike annonser blokkert grunnet brudd på desinformasjonsreglement (SLI 2.1.1)	37 (19.des 22-16.jan 23), 102 069 (H1 2023)	7 897 (19.des 22-16.jan 23), 2.9 mill. (H1 2023)
<i>LinkedIn</i>		
Antall annonser begrenset grunnet brudd på desinformasjonsreglement (SLI 1.1.1, 2.1.1)	N/A (des. 2022), 0 (H1 2023)	19 (EU; des. 2022), 7 (EØS; H1 2023)
Antall falske kontoer fjernet (SLI 14.2.1)	N/A (des. 2022), 88 195 (H1 2023)	849 465 (EU; des 2022), 6.8 mill. (EØS; H1 2023)
Antall poster fjernet om desinformasjon (SLI 18.2.1)	N/A (des. 2022), 113 (H1 2023)	3 443 (EU; des 2022), 24 327 (H1 2023)
<i>Bing Search</i>		
Antall søk behandlet med «defensive search»-algoritme for å motvirke desinformasjon og skadelig innhold (SLI 14.2.1)	N/A (des. 2022), 37 337 (H1 2023)	38.5 mill. (globalt; des 2022), 1.5 mill. (H1 2023)
Antall visninger av faktasjekkinger ved søk (SLI 21.1.1)	200 (des 2022), 575 (H1 2023)	9 911 (des 2022), 24 081 (H1 2023)

Kapittel 2. Kontroll av annonser: GR: LinkedIn forbyr desinformasjon og feilinformasjon på sin plattform, og fjerner poster med beviselig galt eller villedende innhold egnet til å gjøre skade, eksempelvis om Covid og deepfakes⁶⁵ (QRE 1.1.1). LinkedIn tilbyr ikke annonseinntektsdeling. Første siling av annonser er automatisert, mens flaggede annonser gjennomgås manuelt (QRE 2.3.1.). LinkedIn har svært lave tall for antall annonser stoppet. *Advertising* har retningslinjer (fra 2022) for å stoppe publisering og spredning av skadelig desinformasjon og plassering av annonser knyttet til desinformasjonsinnhold.⁶⁶ Annonser om desinformasjon forbyr, i tillegg til annonsering på sider som inneholder desinformasjon, ekstremt politisk innhold m.m. Det er etablert rutiner for å hindre annonsering av innhold som bryter med retningslinjene. Annonsører har mulighet til å ekskludere visse sider/domener de ikke ønsker at deres annonser skal vises ved siden av (QRE 1.3.1), som desinformasjon og sensitivt politisk innhold. NewsGuard og GDI er engasjert for å vurdere om domener/sider regnes som desinformasjon (QRE 1.1.1/1.2.1). Microsoft har et internt team for

⁶⁵ <https://about.ads.microsoft.com/en-us/policies/disallowed-content/information-integrity-and-misleading-content#information-integrity> og

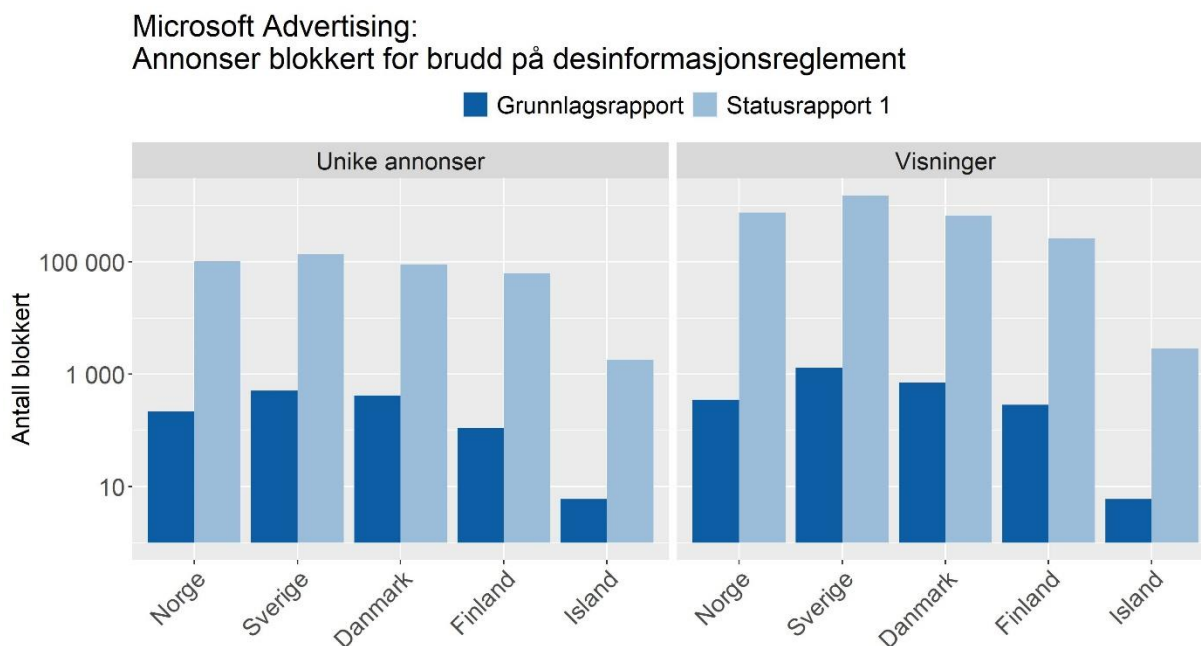
<https://www.linkedin.com/help/linkedin/answer/a1340752>

⁶⁶ <https://about.ads.microsoft.com/en-us/policies/disallowed-content/information-integrity-and>

trusselforebygging, som samarbeider med teamet for maskinl ring for   identifisere TTP-er (teknikker, taktikker og praksiser) for spredning av desinformasjon (QRE 3.1.1).

SR1: Microsoft poengterer at lavt antall annonser fjernet for brudd p  desinformasjonsregelverket henger sammen med LinkedIn's profil, med jobbrelatert innhold og forholdsvis dyre annonser (QRE 1.1.1.).

Figur 23: Unike annonser er enkeltannonser blokkert for brudd p  desinformasjonsregler. Visninger er de individuelle annonsevisningene som de unike annonsene ville ha generert hvis de ikke hadde blitt blokkert. Tallene er justert til 181 dager. Legg merke til logaritmisk y-akse. SLI 2.1.1.



Kapittel 3. Politiske annonser. **GR:** *LinkedIn* har ikke tillatt politiske annonser siden 2018. Dette inneb rer at tjenesten ikke godtar annonser til st tte for en kandidat, et parti eller for innhold ment   p virke valg. Ogs  restriksjoner p  annonsering knyttet til sensitive politiske saker.⁶⁷ *Advertising* har tilsvarende streng praksis⁶⁸ (QRE 4.1.1.). Hvorvidt dette h ndheves godt i praksis er ikke mulig   vurdere ut fra rapporteringen. Microsoft har avst tt fra   signere en rekke av forpliktelsene fordi selskapet mener disse er lite relevante for tjenestene.

Microsoft oppgir at verifiseringsprosessen best r av en blanding av automatiserte og manuelle systemer for   kontrollere at annonsering p  LinkedIn er i tr d med retningslinjene (QRE 7.3.1.). I tillegg kan brukere rapportere innhold. Microsoft Advertising har ogs  restriksjoner p  annonser for politiske saker (QRE 7.3.1).

SR1: I all hovedsak likelydende rapportering som i grunnlagsrapporten. Rapporten utdyper hvordan Microsoft har innf rt flere prosesser for   identifisere og blokkere politiske annonser fra   vises p  sitt annonseringsnettverk, inkludert begrensninger fra visse domener (QRE 7.3.2).

⁶⁷ <https://no.linkedin.com/legal/ads-policy?>

⁶⁸ <https://about.ads.microsoft.com/en-us/policies/restricted-categories/political-content>

Kapittel 4. Integritet. GR: *LinkedIn* forbyr manipulerende adferd (QRE 14.1.1) og desinformasjon, inkludert fra statlige aktører (QRE 14.2.1), og fjerner slikt innhold. Tiltak omfatter blokkering av botter, falske profiler, sider og reaksjoner, deepfakes og annet ikke-autentisk innhold. Rapporteringen er svært mangelfull, data som er oppgitt er overfladiske, og norske tall mangler helt.⁶⁹ Microsoft argumenterer overbevisende for at viral spredning av desinformasjon er mindre relevant for søkemotorer enn for sosiale medier. *Bing* legger vekt på å prioritere troverdige kilder i sine søkealgoritmer (QRE 14.1.1), samt å stanse forsøk på manipulering av søkeresultater. *Bing* iverksetter det som kalles “defensive search” ved mistanke om forsøk på å manipulere søkealgorithmsene, som kan innebære å endre vurderingen av en nettsides kvalitet og troverdighet. Microsoft har egne forskerteam som skal identifisere trusler, blant annet knyttet til desinformasjon og Covid.

SR: Microsoft har oppdatert brukervilkårene som følge av introduksjonen av søketjenester som tar i bruk generativ KI (*Bing Chat*). Nye sikkerhetstiltak skal håndtere TTP-er (taktikker, teknikker og prosedyrer) for å unngå misbruk ved nye, generative søk (QRE 14.1.1.) Microsoft oppgir at de genererte svarene i *Bing Chat* er «forankret» i tredjeparts søkeresultater man kan klikke seg videre til, og hevder derfor at funksjonaliteten i *Bing Chat* ikke er vesensforskjellig fra tradisjonelle *Bing Search*. Rangeringsalgoritmer og spampolitikk er hovedforsvaret mot manipulasjon og misbruk av både *Chat* og *Search*. I tillegg kommer tiltak for å hindre manipulasjon i generative KI-funksjoner på *Chat*, slik som «jailbreaking», altså forsøk på å omgå reglene for tjenesten. Microsoft erkjenner at generativ KI på *Bing* kan brukes til å spre desinformasjon i andre kanaler enn deres egne. Dette er åpenbart av stor betydning, da KI-generert tekst og bilder gir muligheter til å produsere desinformasjon på mer sofistikerte måter enn før og i en helt annen skala. Microsoft viser her til brukervilkårene, der det blant annet heter at man godtar «ikke å delta i aktiviteter som er uredelige, falske eller villedende».⁷⁰ Dette vil være ufullstendig for å stoppe misbruk av tjenestene. Det kan også stilles spørsmål ved Microsoft sidestilling av *BingChat* og *BingSearch*.

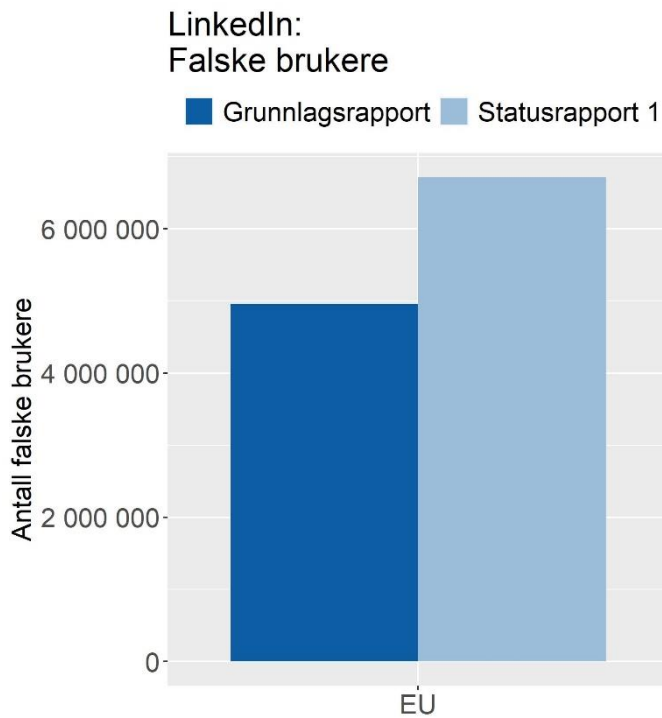
Bing overvåker trender for manipulerende innhold i høyrisiko-områder og setter i verk tiltak for å oppdage og spore informasjonsoperasjoner rettet mot demokratiske land (QRE 14.2.1). De samarbeider med tredjeparter som *NewsGuard* og *Global Democracy Index (GDI)*. Microsoft investerer i KI-teknologi for å oppdage nettverk av falske kontoer på *LinkedIn*. Mellom juli og desember blokkerte eller fjernet *LinkedIn* 58,1 millioner falske kontoer globalt⁷¹. Microsoft oppfordrer til et mer fleksibelt rapporteringssystem der man kun rapporterer for relevante TTP-er, da flere er lite relevant for søkemotorer.

⁶⁹ <https://fujomedia.eu/wp-content/uploads/2023/09/CoP-Monitor-Report.pdf>

⁷⁰ <https://www.bing.com/new/termsfuse>

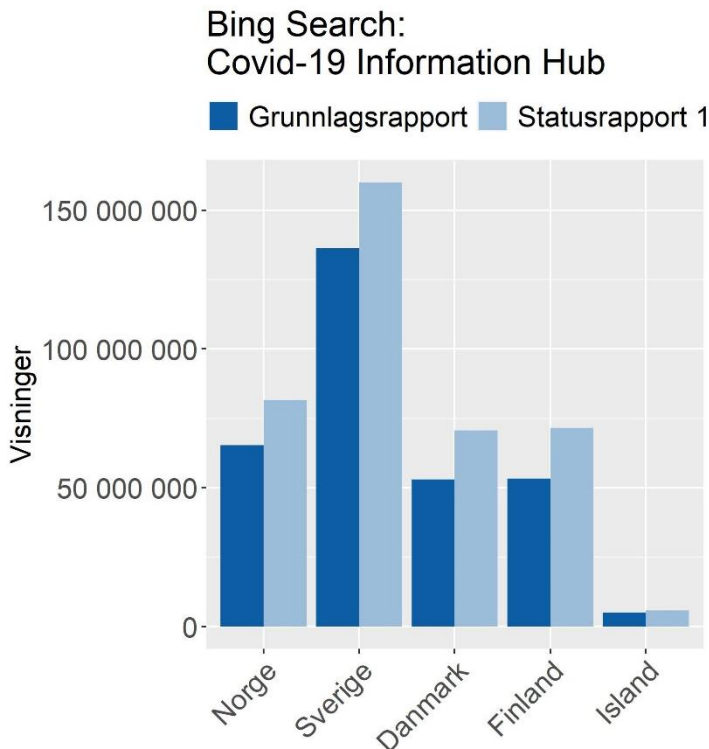
⁷¹ <https://about.linkedin.com/transparency/community-report#fake-accounts>

Figur 24: Antall falske brukere i EU fjernet fra LinkedIn. Tallene er justert til 181 dager. SLI 14.2.1.



Kapittel 5. Tiltak for å styrke brukernes motstandskraft. GR: LinkedIn har satset på tiltak mot Covid og den russiske invasjonen av Ukraina, der kuraterte nyheter fra troverdige kilder fremheves (QRE 17.1.1). Bing legger troverdig offentlig informasjon øverst i søkeresultater. Ved spesielle omstendigheter som valg eller pandemi etableres informasjonssentraler for å sikre kvalitetssikret innhold om emnet (QRE 17.1.1.).

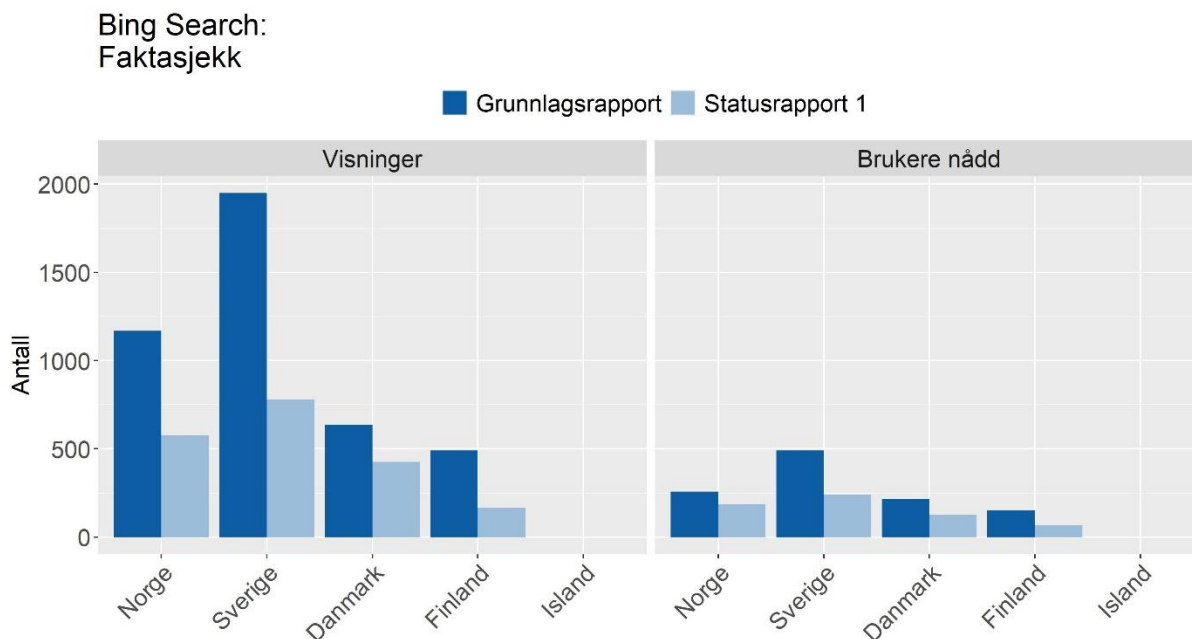
Figur 25: Antall visninger i Bing Search av verktøyene NewsGuard Impression, Knowledge Cards, Transparency Hub Viewership og Public Service Announcement knyttet til Covid-19 Information Hub – en informasjonssentral som gir tilgang til kvalitetssikret og troverdig informasjon om pandemien. Tallene er justert til 181 dager. SLI 17.1.1.



NewsGuard har troverdighetsvurdert over 7500 nyhetssteder og andre informasjonssider på engelsk, som er omsatt til et graderingsystem (grønn/rød) og skal utvides til flere språk. Microsoft har også en tjeneste som hjelper brukere å finne faktasjekket informasjon og som advarer når søk leder til innhold som i faktasjekk har vist seg å være falskt. «Faktaark» dukker opp ved søkeresultater med kvalitetssikret informasjon om et emne (QRE 21.1.1). Gjennom oppkjøp av Miburo i juli 2022 har Microsoft skaffet et internt team for å forske på informasjonspåvirkningsoperasjoner, Digital Threat Analysis Center (DTAC), som jevnlig publiserer rapporter. I tillegg finansierer og samarbeider Microsoft med eksterne aktører som Global Disinformation Index, NewsGuard og Alliance for Securing Democracy og inkorporerer data og forskning fra disse aktørene i utviklingen av egne produkter (QRE 18.3.1). Gjennom «Project Origin» jobber Microsoft med tiltak for å redusere spredning av desinformasjon gjennom syntetisk medieinnhold. Prosjektet utvikler en måte å sertifisere kilden til innhold på, som et vannmerke, og skal kunne avsløre om innhold er blitt endret (QRE 20.1.1). LinkedIn understreker at desinformasjonsinnhold ikke merkes, men fjernes. Derfor er dataene for faktasjekk-merket innhold lik null under kvantitativ rapportering.

For søkemotoren Bing Search kan brukere vurdere pålitelighet gjennom en egen Fact Check-funksjon (QRE 21.1.1.). I tillegg fungerer ordningen ClaimReview som et merkingssystem som påvirker søkealgoritmene. Uavhengige, verifiserte faktasjekkorganisasjoner kan merke egne artikler som faktasjekket for søkemotorer og sosiale medieplattformer. Innhold merket med ClaimReview «leses» dermed som faktasjekket innhold når Bing henter nettsteder for indeksering (QRE 21.1.1).

Figur 26: Antall visninger og brukere nådd for faktasjekk hos Microsoft. Bing Søk anvender ClaimReview, en merkingsordning for faktasjekket innhold. Tabell viser antall ganger en ClaimReview-merket lenke, altså en verifisert og faktasjekket lenke, dukket opp på første side av søkeresultatene for en bruker i landet. Tall justert til 181 dager. SLI 21.1.1.



SR: Etter lanseringen av KI-drevne tjenester, har Microsoft oppdatert sin satsing på forskning på kunstig intelligens og ansvarlig bruk av tjenestene dette muliggjør. Flere forskningsprosjekter om desinformasjon er gjennomført av Microsoft Research og Microsofts AI for Good Lab. I mars lanserte de en KI-chatbot for å gi troverdig informasjon om covid-19, som skulle gi brukere pålitelig informasjon om pandemien. Forskere har også gjort sammenlignende studier for å vurdere tradisjonelle søk opp mot søk basert på store språkmodeller. I tillegg samarbeider Microsoft med Princeton University for å forske på desinformasjonstrender, spesielt etter viktige hendelser, for å forstå hvordan folk reagerer på falsk informasjon på internett (QRE 18.3.1.). Microsoft har iverksatt merking av opphav til KI-genererte bilder i sin KI-bildegenerator Image Creator (C20). Selskapet har også forpliktet seg til å utvikle tilsvarende kildemerking for KI-generert lyd og annet bildeinnhold, og skal jobbe med å forbedre disse systemene.⁷² Microsoft har lite informasjon om samarbeid med faktasjekkere i GR, men oppgir i SR at de er i dialog med nye, mulige samarbeidspartnere.

Kapittel 6. Tilgang for forskere: **GR:** LinkedIn gir API-tilgang til data knyttet til blant annet innhold og engasjement, markedsføring og salg, men Microsoft kjenner ikke til at det er brukt for å undersøke desinformasjon. Bing Search gir forskere og allmennheten tilgang til MS MARCO, en samling datasett knyttet til dyp læring innen søk, utledet fra Bing-forespørsler og tilknyttede data. Dette er åpent for alle, og er benyttet i flere forskningsartikler om desinformasjon (QRE 26.1.1). Søkemotoren tilbyr også bruk av Bing API-er for bildesøk, nettsøk og nyhetssøk.

⁷² <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/07/Microsoft-Voluntary-Commitments-July-21-2023.pdf>

SR: Microsoft oppgir at LinkedIn utvider API-tilgangen for data knyttet til desinformasjon. Det tilbys nye verktøy for forskere som jobber med ansvarlig KI (QRE 26.1.1).

Kapittel 7. Tilgang for faktasjekkere: GR: Rapporterer ikke om samarbeid med norsk faktasjekkjeneste. LinkedIn oppgir å ha to avtaler med faktasjekk-organisasjoner, og at det jobbes med å utvide til flere språkområder fra dagens engelsk, spansk og portugisisk (SLI 30.1.1.) Bing Search tilbyr uavhengige faktasjekkere å merke faktasjekkinnhold som troverdig/ikke troverdig gjennom "ClaimReview"-merking, et system som i utgangspunktet skal være åpent for faktasjekkere uavhengig av landbakgrunn (QRE 30.1.2).

SR: LinkedIn oppgir å ha inngått et samarbeid med Reuters, global nyhetstjeneste med over 2500 journalister spredt på 200 lokasjoner verden rundt.

Twitter

Rangering: GR: kvalitativ 1 , kvantitativ N/A . Antall sider: 75.

Twitter har trukket seg fra bransjenormen, og har ikke levert SR.

Norsk rapportering: Nei 

Vurdering: Den kvalitative rapporteringen er svært mangelfull. Twitter oppgir å ha 21 kategorier for annonseforbud, men har ikke rapportert på antall annonseforespørsler som er avvist, verken totalt eller nasjonalt. Twitter mener de fleste forpliktelsen ikke er relevante, noe som er lite troverdig. Twitter oppgir ellers at de vil gjenåpne sitt senter for annonsetransparens, som ble lagt ned i 2019. Det er ikke rapportert på kvantitative indikatorer. Twitters rapport viser først og fremst at tjenesten ikke forholder seg til bransjenormen. Det er også kjent at selskapet har kuttet drastisk i staben, som har gått ut over den europeiske delen av virksomheten, og de som arbeidet med samfunnsansvar og moderering.

Grunnlagsrapporten fra Twitter er på ingen områder i tråd med bransjenormen. X/Twitter er definert som en VLOP i DSA, som har plikt til å identifisere og redusere systemiske risikoer. Det er lite som tyder på at Twitter har innført tiltak for å dempe de systemiske risikoene om desinformasjon.

Etter at Twitter trakk seg fra bransjenormen, har EC uttalt at Twitter vil «stresstestes» opp mot de systemiske risikokravene i DSA.⁷³ EC har som nevnt vurdert globale plattformers rolle i spredning av russisk desinformasjon om Ukraina, der Twitter står for mye av veksten av russisk desinformasjon.⁷⁴

Medietilsynet vurderer at Twitter ikke etterlever bransjenormen. EC har eksklusiv tilsynskompetanse med VLOP/VLOSE i DSA, og må vurdere om Twitter bryter forpliktelsene i DSA, og hvordan et eventuelt brudd skal håndheves.

⁷³ <https://www.euronews.com/next/2023/06/23/eu-visits-silicon-valley-thierry-breton-puts-twitter-under-stress-test-over-blocs-new-law>

⁷⁴ <https://apnews.com/article/disinformation-musk-x-twitter-european-union-9f7823726f812bb357ee4225b884354f>

Twitch

Rangering: GR: kvalitativ 1●, kvantitativ 1●. Antall sider:16.

Twitch skal sende første statusrapport i januar 2024.

Rapportering for Norge: Nei ●

Vurdering: Rapporteringen er mangelfull. Som live-streaming-tjeneste der det meste skjer live over tid, som gaming, argumenterer Twitch for at plattformen er mindre egnet til spredning av virale desinformasjonskampanjer. Plattformen oppgir å blokkere brukere som gjentatte ganger deler desinformasjon eller konspirasjonsteorier som kan forherlige blant annet vold. Det er også forbud mot boter, spammere, falske kontoer, kontokapring etc.

Den kvalitative rapporteringen er lite relevant. Det er ingen konkret informasjon om inntjeningsmodeller for annonsører. Informasjonen er for mangelfull til å gi en reell vurdering. Det er ingen tall på norsk nivå.

Twitch er ikke er regnet som VLOPs, og har dermed mulighet for forenklet rapportering. Som mindre plattform leverer tjenesten heller ikke statusrapport halvårlig, men skal etter planen levere ny rapport ett år etter grunnlagsrapporten.

Tabell 6: Utvalgte nøkkeltall fra Twitch-rapport.

Nøkkeltall, Twitch

KPI	Globalt
Antall brukere utestengt for all fremtid (SLI 1.1.1)	4 (månedssnitt Q4 2022), N/A (H1 2023)
Antall falske kontoer/botnets (SLI 14.2.1)	270 921 (månedssnitt Q4 2022), N/A (H1 2023)
Antall tilfeller av bruk av falske følgere/abonnenter (SLI 14.2.1)	6 (månedssnitt Q4 2022), N/A (H1 2023)
Antall kapringer av kontoer (SLI 14.2.1)	32 (månedssnitt Q4 2022), N/A (H1 2023)
Antall interaksjoner med et hjelpesenterverktøy knyttet til kritisk medieforståelse (SLI 17.1.1)	2 800 (månedssnitt Q4 2022), N/A (H1 2023)



ISBN: 978-82-8428-042-4



Nygata 4, 1607 Fredrikstad



69 30 12 00



post@medietilsynet.no

 Medietilsynet