DIGITAL HVERDAGS-SEXISME

Hvordan forebygger og håndterer man den hårde tone i onlinedebatter

RAPPORT AF ANALYSE & TAL OG KVINFO
APRIL, 2025

KVINFO
KØN OG LIGESTILLING

&# Analyse & Tal

TrygFonden

Forord

Den digitale samtale udgør en stadig større del af vores demokratiske debat. Alligevel oplever danskere, at den hårde tone online afskrækker dem fra at deltage. Derfor står vi over for væsentlige udfordringer med at sikre en respektfuld og inkluderende dialog online.

Denne rapport belyser digital hverdagssexisme, og hvordan man i praksis forebygger og håndterer centrale udfordringer, som moderatorer og andre hver dag bliver konfronteret med på sociale medier. Her kan moderation være vejen frem til at mindske sexisme og hverdagssexisme i den offentlige debat. Rapporten tager således også afsæt i moderatorers daglige arbejde med at modvirke særligt sexisme og hverdagssexisme på digitale platforme.

Projektet *Digital hverdagssexisme* viser, at moderation af den hårde tone er en meget central, men også udfordrende opgave. Den nuværende moderationspraksis er ressourcekrævende og fyldt med dilemmaer, som i praksis betyder, at det er svært at sætte aktivt ind over for digital hverdagssexisme og effektivt at forbedre debatkulturen.

Samtidig viser vores test af automatiserede moderationsværktøjer et lovende potentiale, om end teknologien endnu kræver videreudvikling, før den kan implementeres effektivt i praksis.

Rapporten giver således indblik i både udfordringer og muligheder inden for digital moderation og peger på vigtige udviklingspunkter for at styrke en inkluderende online debatkultur uden diskrimination og hverdagssexisme.

Tak til TRYGFONDEN for at støtte projektet.

Henriette Laursen
Direktør, KVINFO

og

Anna Ørtoft
Partner, Analyse & Tal

Indledning

Sociale medieplatforme har grundlæggende ændret måden, vi fører offentlig samtale på i Danmark, og det har medført nye udfordringer for den demokratiske debat. Hadefuld tale og diskrimination online er blevet almindeligt forekommende. Især kvinder, kønsminoriteter og etniske minoriteter bliver ramt med det resultat, at nogle trækker sig fra den offentlige debat.

Det er et åbent og stadig mere presserende spørgsmål, hvordan vi får skabt et inkluderende og trygt digitalt rum for alle. Flere projekter har fokuseret på de helt grove former for chikane, trusler og vold. Omdrejningspunktet i dette projekt er den mere subtile sexisme, som optræder i hverdagen.

Sexisme er diskrimination på baggrund af køn. På sociale medier kommer sexisme til udtryk på mange måder – fra grove udtryk såsom digitale trusler og overgreb til mere subtile former såsom nedladende stereotype kommentarer og sexistiske 'jokes'. Denne subtile form kaldes *hverdagssexisme*.

Vi ved fra forskning, at det er vigtigt at sætte ind over for hverdagssexisme for også at komme den mere grove sexisme til livs. Hverdagssexismen er nemlig med til at rykke normerne for, hvad der er legitim adfærd.

Hverdagssexisme er derfor vigtig at forebygge og håndtere, da den både kan have konsekvenser i sig selv og samtidig bane vejen for grovere sexistiske udtryk og handlinger, både online og offline.

Det kræver viden og værktøjer at forebygge og håndtere den digitale hverdagssexisme. Hverdagssexisme optræder typisk i kommentarspor på sociale medier og er indlejret i vores sprog, kultur og normer. Det er også den form for sexisme, man typisk først opdager, når den går ud over en selv – og som de fleste affejer eller helt overser.

I denne rapport udfolder vi udfordringerne med at få øje på digital hverdagssexisme og modvirke den systematisk. Vi viser, hvordan moderatorer spiller en central rolle i dette arbejde. Og vi kommer med anbefalinger til, hvad der skal til for at tage livtag med digital hverdagssexisme og derved gøre det digitale rum mere trygt og inkluderende.

Ved at dykke ned i de eksisterende moderationspraksisser har vi udviklet et mere nuanceret vidensgrundlag for moderation på sociale medier i en dansk kontekst. Vi har udforsket potentialet i både manuel og AI-baseret moderation samt udviklet og testet nye strategier og udviklet konkrete redskaber til håndtering af digital hverdagssexisme. Disse resultater kan bruges til at give et mere kvalificeret grundlag for at skabe tryggere rum for de grupper, som er underrepræsenteret i den digitale offentlige debat.

Ulige deltagelsesmuligheder i det digitale rum

Næsten fire gange så mange borgere deltager i den offentlige debat digitalt end borgere, der deltager i debatter i fysiske rum (Institut for Menneskerettigheder, 2024).

Det er derfor bekymrende, at tre ud af fire danskere oplever, at tonen i debatten er blevet for hård på de sociale medier (Center for Sociale Medier og Demokrati, 2024), da det kan have konsekvenser for deltagelse i den offentlige debat online.

Mange offentlige debatter forgår på Facebook – som 60 procent af den danske befolkning besøger dagligt (DR Medieanalyse, 2023).

65 procent af Facebook-brugere afholder sig fra at skrive kommentarer i debatter på grund af debattonen. Kvinder føler sig i højere grad afskrækkede fra at deltage i debatten, da de oftere modtager sproglige angreb, sexistiske kommentarer og udsættes for groft sprogbrug (Institut for Menneskerettigheder, 2022).

Langt flere kvinder er udsat for hadefulde angreb end mænd. Det viser rapporten "Angreb og had i den offentlige debat på Facebook", som baserer det på 73 millioner Facebook-opslag og kommentarer fra danske medie- og politikersider i perioden 2021-2024. Undersøgelsen viser, at kønsbaseret had i 73 procent af tilfældene rettes mod kvinder, 27 procent mod mænd og 5 procent mod kønsminoriteter. Herudover estimerer rapporten, at 68 procent af angreb i den offentlige debat er skrevet af profiler med mandlige navne og kun 32 procent med kvindelige navne (Analyse & Tal et al., 2025).

Sexisme skaber barrierer for den demokratiske debat

En ny undersøgelse viser også, at den hadefulde tale online især retter sig mod kvinder. Hadet kan have en meget grov og sexistisk karakter og er ofte meget seksuelt ladet. Det bevirker, at kvinder og minoriteter trækker sig fra at ytre sig og udtrykke sig, som de gerne vil (Amnesty International, 2025).

En tidligere undersøgelse viser, at næsten hver femte danske kvinde, der beskriver sig selv som "moderat til aktiv internetbruger", oplever digital chikane. Ud af de kvinder, der oplyste at have været udsat for digitale trusler og chikane mindst en gang, oplevede 45 procent, at det havde karakter af misogyni eller sexisme. Mange oplevede også, at det havde store negative personlige konsekvenser for dem: 49 procent oplevede lavere selvværd eller tab af selvtillid. 36 procent oplevede stress, frygt eller panikanfald (Amnesty International, 2017).

Undersøgelserne peger altså på, at køn har en stor indflydelse på borgeres oplevelse af den digitale debat på Facebook. Og at det skaber særlige barrierer for kvinder i forhold til at deltage i den offentlige debat. Det udfordrer idealet om et demokrati, som alle uanset køn, kønsidentitet, seksuel orientering og etnicitet har lyst til og mulighed for at deltage i.

Moderation kan afhjælpe sexisme

Moderation er et redskab, der bruges til at monitorere og skabe bedre dynamik, retning eller stemning i opslag og kommentarspor. Det kan indebære at kommentere eller fjerne indhold, der er stødende, irrelevant eller af andre grunde uønsket.

Moderation af den offentlige samtale bruges i forskellige typer digitale formater. Dette kan være på Facebook-sider og i Facebook-grupper, men også digitale fora på Reddit eller i kanaler på Discord.

Danskere viser bred opbakning til brug af moderation i den offentlige digitale debat (Center for Sociale Medier og Demokrati, 2024). 59 procent er "enige" eller "overvejende enig" i, at onlinedebatten har behov for moderation. 32 procent svarer "hverken enig eller uenig" eller "ved ikke". Kun 9 procent af danskerne er "uenig" eller "overvejende uenig" i, at der er behov for moderation af offentlige debatter online.

Hvad kan du læse om i denne rapport

Moderation af digitale platforme er en udfordring, som rigtig mange står overfor. Det gælder både i foreninger, virksomheder, mediehuse, NGO'er og politiske partier, men også blandt de enkelte brugere. Moderation er et arbejde, som varetages af mange forskellige aktører, men oftest af løst ansatte eller frivillige.

Denne rapport undersøger moderationsværktøjer og de udfordringer, der opstår med moderation af kommentarer, som indeholder sexisme og hverdagssexisme.

Formålet er at bidrage med anbefalinger til, hvordan man med moderation kan skabe et trygt digitalt rum, som er inkluderende for alle, samt mindske hverdagssexisme online og dermed også forebygge grovere sexisme og overgreb.

I rapporten præsenterer vi værktøjer til moderatorer, der kan gøre dem i stand til at identificere hverdagssexisme og sexisme på baggrund af en faglig vurdering. Hertil kommer vi med anbefalinger til organisering af arbejdet samt konkrete værktøjer til at forebygge og modvirke digital hverdagssexisme og skabe større tryghed.

Værktøjerne kan anvendes bredt af moderatorer i den digitale debat, mens en del af anbefalingerne retter sig mod mere professionelt organiseret moderation i for eksempel mediehuse eller politiske sekretariater.

Kapitel 1 har særligt fokus på, hvordan man identificerer den digitale hverdagssexisme ud fra en faglig vurdering.

Kapitel 2 har særligt fokus på, hvilke moderationsstrategier der er mest effektive for at håndtere digital hverdagssexisme og andre former for uønsket indhold.

Kapitel 3 har særligt fokus på barrierer for moderation af digital hverdagssexisme i relation til større moderationsteams i mediehuse og lignende.

Kapitel 4 har særligt fokus på automatiserede moderationsløsninger med søgenøgle og kunstig intelligens.

Hovedresultater

1 **Hverdagssexisme er en central udfordring i digitale debatter**
Den spredes potentielt hurtigere og når bredere ud digitalt og normaliseres dermed i og af debatten online. Det udgør både et demokratisk problem i sig selv og skaber grobund for grovere sexistisk adfærd, hvorfor det er meget væsentligt at forebygge og håndtere gennem moderation.

2 **Hverdagssexisme er vanskeligt at identificere**
- både generelt og for moderatorer. Der mangler viden og tydelige retningslinjer til at identificere og håndtere sexisme og hverdagssexisme.

3 **Den nuværende moderationspraksis er ressourcekrævende og baserer sig primært på usynlig moderation gennem sletning af kommentarer.**
Dette står i kontrast til forskningsresultater, der viser, at synlig moderation med kommentarer er mere effektiv til at forbedre den digitale debatkultur.

4 **Moderatorer i større moderationsteams står over for tre centrale barrierer, der hindrer effektiv, synlig moderation:**

1) Arbejdsforhold og organisering
Skæve arbejdstider betyder, at moderatorer ofte sidder alene uden mulighed for sparring med kolleger og de ofte løse ansættelser giver tillige usikre rammer og mandat.
2) Arbejdet baseres på bias og mavefornemmelser
Uden klare retningslinjer må moderatorer oftere basere deres vurderinger på bias og mavefornemmelser, hvilket skaber usikkerhed og manglende håndtering og forebyggelse af sexisme.
3) Ideal om neutralitet og frygt for eskalering
Moderatorer fanges i dilemma om neutralitet og frygt for eskalering af konflikter i kommentarsporet. Uden retningslinjer og klart mandat afholder det dem fra at gribe ind i problematiske debatter.

5 **De testede automatiserede moderationsværktøjer viser potentiale.**
men det kræver yderligere udvikling, før de effektivt kan understøtte moderatorernes arbejde i praksis.

Anbefalinger

Til mediehuse og andre organisationer, der leder og anvender moderatorer på deres Facebook-sider

Opkvalificer moderatorer i viden om og håndtering af hverdagssexisme for at mindske subtile og grovere former for sexisme

Skab fælles viden, forståelse og retningslinjer om digital sexisme og hverdagssexisme ved brug af sexismetrekanten. Dette værktøj kan bruges til at analysere og kategorisere hverdagssexisme og sexistisk adfærd. Ved at sætte ind over for hverdagssexisme skaber man ikke kun et mere trygt miljø for brugerne i ens kommentarspor, men reducerer også voldsommere former for sexistiske ytringer. Det kan skabe mere tryghed, større deltagelsesmuligheder samt en strømlinet tilgang til moderation.

Styrk professionalisme, og implementer en systematisk moderationsstrategi og klare retningslinjer for at skabe strømlinet og effektiv moderation

Implementer en moderationsstrategi, der etablerer tydelige faglige rammer og retningslinjer samt standardiserede procedurer for håndtering af forskellige overtrædelser. En sådan strategi vil medføre mere ensartet moderation på tværs af moderatorer og teams, samt en mere effektiv udnyttelse af moderatorernes ressourcer og kompetencer.

Brug synlig moderation for at fastslå rammer for den digitale debatkultur og forebygge uønsket adfærd i kommentarsporet

Anvend *synlig moderation*, hvor årsagen til indgreb forklares tydeligt. Denne transparens har større effekt end at fjerne indhold uden begrundelse. Når brugere kan se begrundelsen for moderationen, lærer de fællesskabets retningslinjer at kende og tilpasser deres adfærd derefter. Dette skaber på sigt et bedre debatmiljø, i modsætning til *usynlig moderation*, hvor indhold blot forsvinder uden mulighed for læring og uden potentiel ændring af normer og kultur for den digitale adfærd.

Øg ressourcerne og giv moderatorerne et tydeligt mandat

Øg ressourcerne til moderation og forbedre organiseringen af moderatorernes arbejde, så de får systematisk adgang til sparring og oplæring og mindre usikkerhed i ansættelsen. Formuler et tydeligt mandat og retningslinjer, der sikrer moderatorerne beslutningskompetence og at de kan handle på organisationens vegne og ikke som individer. En stærkere ledelsesmæssig forankring og prioritering er med til at skabe det nødvendige grundlag for, at moderatorer i praksis kan forebygge og håndtere digital sexisme og hverdagssexisme.

Metode

Digital hverdagssexisme undersøger, hvordan moderation kan mindske hverdagssexisme i den offentlige digitale debat.

Vi tager udgangspunkt i eksisterende forskning i, hvordan kultur, strukturerer og normer muliggør hverdagssexisme og sexisme (Muhr, 2019; Reinicke, 2018; Einersen et al., 2021).

På baggrund af denne viden og en undersøgelse af moderatorers praksisser på blandt andet to mediehuse giver rapporten et indspil til, hvordan man skaber bedre rammer for moderation, og hvordan moderatorer i praksis kan forebygge og modvirke digital hverdagssexisme.

Undersøgelsen er baseret på fire analytiske tilgange beskrevet nedenfor.

Litteraturstudie

En systematisk gennemgang af relevant forskning for at kortlægge eksisterende metoder til moderation og interventioner i forhold til diskrimination og sexisme. I denne gennemgang er studier relateret til både fysiske og digitale rum gennemgået med fokus på at identificere relevante tilgange og strategier anvendt i praksis (bilag 2).

Kvalitativ undersøgelse af moderatorers praksis og udfordringer

Et feltstudie har undersøgt moderatorers praksis og udfordringer i to mediehuse. Dette inkluderer fem dages feltarbejde med observation af live-moderation og 15 dybdegående interviews. Den kvalitative undersøgelse fokuserer på rammerne for moderation samt de redskaber og strategier, som moderatorerne anvender for at håndtere hårde debatter med diskrimination, herunder sexisme og hverdagssexisme (bilag 3).

Evaluering af digitale moderationsmetoder

En evaluering af to typer automatiserede moderationsløsninger. Vi sammenligner en AI-baseret misogyni-algoritme med vores egenudviklede søgenøgle til identifikation af hverdagssexisme. Metoderne er blevet evalueret ved at sammenligne deres resultater med menneskelige vurderinger af næsten 2.000 Facebook-kommentarer for, om de indeholdt hverdagssexisme eller ej.

Udvikling af anbefalinger til moderation baseret på GenderLabs

Afprøvelse af nye moderationsstrategier til moderation af hverdagssexisme baseret på litteratursøgning fra forskning og praksis samt samskabelsesmetoden GenderLab udviklet af KVINFO og Copenhagen Business School til at skabe praksisnære og kulturforandrende tiltag (Christensen et al., 2021). Vi har afholdt GenderLabs og oplæg for henholdsvis moderatorer i mediehuse, de politiske sekretariater og for en gruppe med ledelsesrepræsentanter hos et af mediehusene (se bilag 1).

Advisory board

Projektet *Digital hverdagssexisme* er blevet rådgivet af et advisory board, der samler en række eksperter indenfor sexisme, digitale krænkelser, sociale medier, allegoritmer og hadefuld tale.

Leon Derczynski er associate professor hos ITU og er ledende forsker hos Strømberg NLP med ekspertise i machine learning og digital sprogbehandling. Han har udviklet en algoritme der detekterer og måler typer af hadtale på nettet, herunder had mod kvinder og hverdagssexisme.

Sara Louise Muhr er professor på Copenhagen Business School og har i mere end 15 år forsket i ledelse og inklusion, herunder haft særligt fokus på køn, sexisme og hverdagssexisme.

Miriam Michaelsen er forperson for Medierådet for børn og unge samt stifter og tidligere bestyrelsesleder for foreningen Digitalt Ansvar, med særligt fokus på tidssvarende lovgivning og e ektiv efterforskning af digitale krænkelser.

Lumi Zuleta er specialkonsulent på området for ligebehandling ved Institut for Menneskerettigheder med særlig viden indenfor digital hadtale.

Rune Vammen Lesner er seniorforsker hos VIVE og har en ph.d. i økonomi. Han forsker primært i konsekvenser af uligheder på arbejdsmarkedet grundet køn og forældres socioøkonomiske baggrund.

1

Hvordan udspiller hverdagssexisme sig digitalt

Hverdagssexisme i en digital kontekst

Sexisme er et komplekst fænomen og findes i mange former – fra åbenlyse krænkelser til subtile hverdagshændelser. Sexisme defineres i sin simpleste form som et udtryk for diskrimination på baggrund af køn.

Europarådet definerer sexisme som enhver adfærd, hvor nogen gennem billeder, handlinger eller ord udtrykker, at en person eller gruppe har mindre værdi på grund af deres køn. Det kan både foregå i offentlige og private sammenhænge, online og offline (Europarådet, 2019). Sexisme kan være både intentionel og uintenderet med den effekt, at en person eller en gruppes værdighed eller rettigheder bliver krænket.

Sexisme er med til at skabe et intimiderende, fjendtligt, degraderende, ydmygende eller krænkende miljø for eksempel på arbejdspladser eller digitalt. Ofte har sexisme rod i ubevidste bias og kønsstereotyper, som vi alle bærer rundt på. Derfor er det afgørende at skabe adfærdsmæssige og kulturelle forandringer.

Hverdagssexisme i en digital kontekst er en form for sexisme, der spænder fra subtile og ofte normaliserede sexistiske udtryk og handlinger, som forekommer online på sociale medier, gaming-platforme eller andre online rum.

Det kan omfatte hadefulde kommentarer, kønsstereotyper og diskriminerende sprogbrug, som både går direkte på en person i kommentarsporet eller retter sig mod en gruppe, som for eksempel kvinder, transpersoner eller muslimske mænd.

Intersektionalitet

I arbejdet for at modvirke sexisme og hverdagssexisme er det vigtigt at have fokus på, at det ikke alene vedrører kvinder. I praksis overlapper forskellige former for diskrimination og skaber særligt udsatte grupper og personer i forhold til identitetsmarkører såsom seksuel orientering, kønsidentitet, etnicitet, klasse, alder osv. Man kan for eksempel være særligt udsat, hvis man både er etnisk minoritet og transperson. Begrebet intersektionalitet indrammer denne bevidsthed (Crenshaw, 1991).

Mænd bliver også udsat for sexisme, og kvinder kan være udøvere og bærere af en sexistisk kultur. Hvis man ikke har dette for øje, kan man risikere at overse konkrete tilfælde af sexisme og misse vigtige aspekter af den måde, sexisme manifesterer sig og fungerer på.
For eksempel ser vi en tendens til at nedgøre mænd ved at brug af skældsord, der nedgør det feminine – som i dette eksempel:

"Der er da ikke nogen, som vil spilde sin stemme på jer med den v..... som formand. Det er godt gjort, at de to gamle BORGERLIGE partier har sådanne to slapsvanse som formænd".

Sexisme handler således ikke alene om diskrimination af kvinder, men om diskrimination af alle uanset køn. At reducere sexisme til en kvindesag er med til at negligere, hvordan mænd og non-binære personer også oplever sexisme og selv er udsat for snævre kønsstereotyper. (Groes, 2023; Reinicke, 2018)

Sexismetrekanten

Sexismetrekanten illustrerer, hvordan de forskellige former for sexisme hænger sammen. Herunder at hverdagssexisme kan lægge grund til mere grove former for sexisme som overgreb og voldtægt.

Sexismetrekanten er oprindelig udviklet af Everyday Sexism Project Danmark og belyst teoretisk af flere (Everyday Sexism Project; Kelly, 1989).

Den skaber rum til at forstå, hvordan én form for kønsbaseret krænkelse kan lede til andre former for seksuel vold. Eksempelvis kan accept af hverdagssexisme i en organisation skabe rum for, seksuel chikane eller isoleret set grovere sexisme og/eller overgreb. I en digital kontekst vil accept af hverdagssexistiske kommentarer kunne skabe grobund for grovere former for verbal sexisme og digitale overgreb.

Dryp-dryp effekten

Ofte opfattes hverdagssexistiske handlinger som ubetydelige i sig selv, men de kan have store konsekvenser. Det kan anskueliggøres ved at sammenligne med myggestik: Et enkelt myggestik er muligvis ikke så slemt. Men hvis du får rigtig mange myggestik, og måske endda hver eneste dag, så bliver det problematisk.

I forskningen anvendes begrebet *dryp-dryp-effekten* til at beskrive den måde, hvor små, men ofte forekommende sexistiske hændelser, der finder sted over en lang periode, til sidst bliver for meget. Hver for sig fremstår de små sexistiske handlinger eller udtryksformer måske harmløse, men sammenlagt kan det få konsekvenser (Einersen et al, 2021).

Sexismetrekanten

Digital sexismetrekant

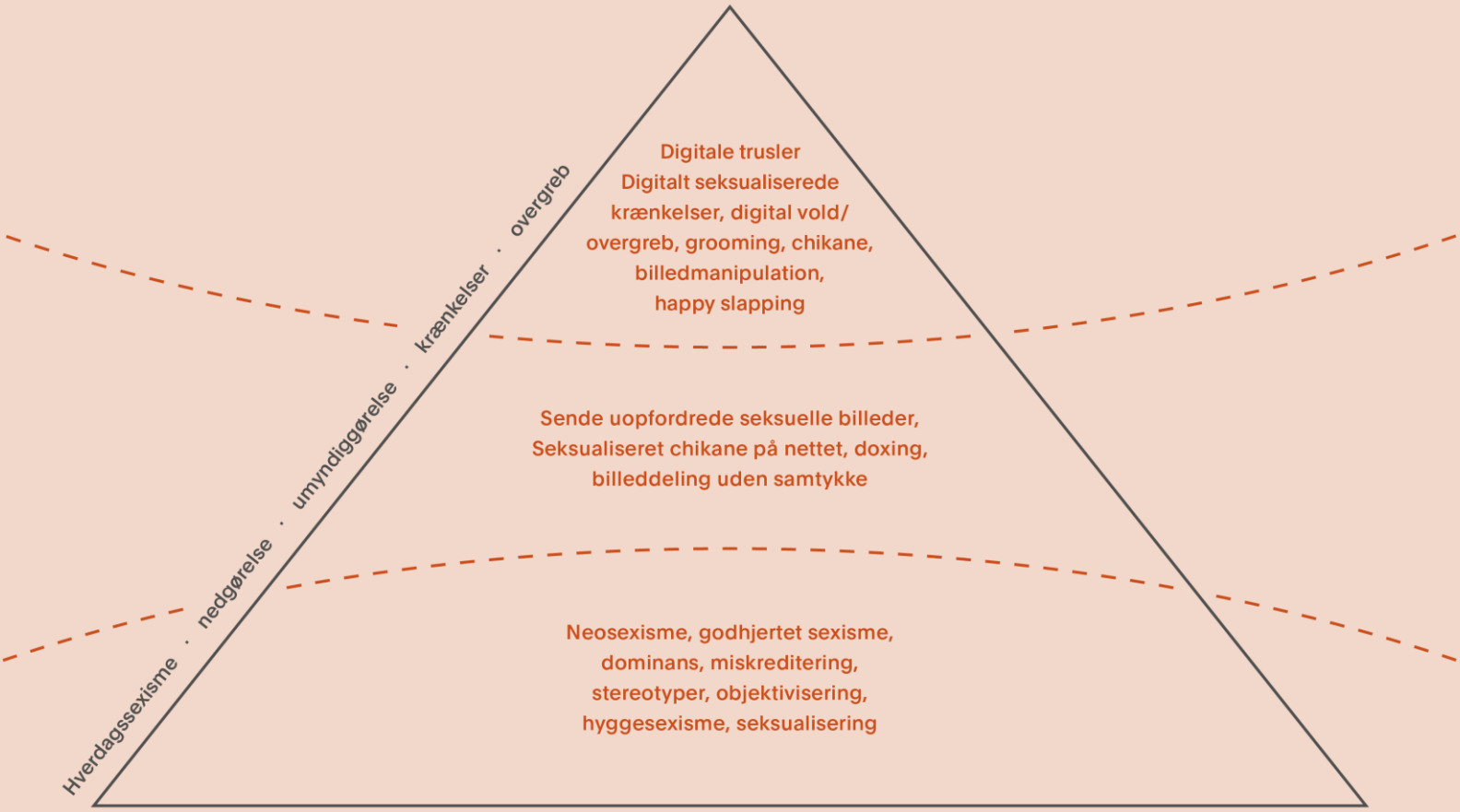
Sexismetrekanten ovenfor illustrerer primært den sexisme, som forekommer i et fysisk rum, gennem verbale eller fysiske krænkelser, og berører kun delvist det digitale. På baggrund af indsigter fra projektet har vi udviklet en sexismetrekant, som alene relaterer sig til det digitale rum. Her spiller nogle andre forhold ind, som er særlige for det digitale.

De fysiske rum er let afgrænselige og vi kan i højere grad se eller fornemme hvor afgrænsningerne går. Dette er modsatrettet ikke tilfældet med de digitale miljøer, hvor der også er andre forstærkende faktorer såsom likes, kommentarer eller delinger. Det har den effekt, at (hverdags)sexismen kan virke allestedsnærværende.

I det digitale rum antager sexisme altså andre og nye former i forhold til det fysiske rum. Vi illustrerer dette nedenfor i en videreudvikling af den oprindelige sexismetrekant. Her vises, hvordan forskellige typer af digital sexisme hænger sammen og bygger oven på hinanden.

At sætte ind over for digital hverdagssexisme kan være med til at skabe en ændring i debatkulturen, som også kan have betydning for grovere sexisme og diskrimination. Herudover kan en ændring i den digitale kultur også forandre adfærden i den fysiske verden.

På næste side uddybes betydningen af de forskellige former for digital hverdagssexisme.



Miskreditering

Underkendelse på baggrund af køn, ofte koblet til kompetencer. Personer eller grupper bliver miskrediteret ved, at andre nedgør, underkender eller udelukker dem med henvisning til deres køn. Eller med en indirekte miskreditering fordi de falder uden for de normative idéer om maskulinitet og femininitet (Anzovino et al., 2018).
Eksempler: "Man kan ikke overlade vigtige beslutninger til kvinder, fordi de er styret af deres hormoner", "Man skal ikke lytte til ham, for han er en tøffelhelt", "Man kan ikke overlade beslutningerne til en, som ikke engang kan finde ud af, om han/hun skal være mand eller dame."

Stereotypisering

Når man fastholder og reproducerer forsimplede og begrænsede samfundsidealer om køn og/eller kønsidentitet og seksuel orientering (Anzovino et al., 2018).
Eksempel: "Det eneste kvinder dur til, er at lave mad", "Det kan han ikke finde ud af, med de løse håndled".

Velment sexisme

Når man fx roser en kvinde ved at tillægge hende mandlige kvaliteter (Jha & Mamidi, 2017).
Eksempler: "Hun har sgu nosser", "Du skal nok lære det/forstå det, når du får hår på brystet", "Hvem havde troet, at man kunne se så godt ud?" (om en transperson).

Dominans

Afsender giver udtryk for, at kvinder er mindre værd end mænd, eller at nogle former for maskulinitet er mere værd end andre (hegemonisk maskulinitet) (Anzovino et al., 2018).
Eksempler: "Overlad det til mændene. Kvinderne har ikke intelligensen", "Det kan han da ikke holde til. Han er jo bare en splejs", "Den slags arbejde kræver hår på brystet", "Du flæber som en lille pige".

Seksualisering og objektivisering

Udtalelser med seksuelle undertoner, eksplicitte opfordringer eller henvisninger til seksuel aktivitet, hvor man bliver objektiviseret. Ofte med henvisning til køn, enten direkte eller indirekte som i dette eksempel, der typisk vil rette sig til en kvinde.
Eksempel: "Hvor mange penge skal du have for at være min en hel eftermiddag?"

Neosexisme

Idéen om, at sexisme ikke findes, og myten om, at vi allerede har ligestilling (Tougas et al., 1999). Denne form for hverdagssexisme er en af de mest udbredte digitalt, ifølge Leon Derzynsky m.fl. Men også i den fysiske verden ser vi, hvordan neosexisme er en barriere for en inkluderende debat om emner om køn, ligestilling og diskrimination. Neosexisme udgør en gråzone, da nogle opfatter det som en legitim holdningstilkendegivelse, mens andre opfatter det som en måde at afspore debatten og udelukke stemmer. Besvarelse af neosexisme kan med fordel være faktabaseret.
Eksempel: "Hvorfor snakker vi igen om kønskvoter, når vi allerede har ligestilling i Danmark?"

Hvad kendetegner og forstærker digital hverdagssexisme

Digital spredning forstærker dryp-dryp-effekten

I digitale miljøer er likes, kommentarer og delinger med til at forstærke sexismens spredning og effekt. Digital hverdagssexisme spredes hurtigt og bredt. Det er vanskeligt at spore, hvem og hvor mange der er vidne til den (Mogensen & Rand, 2019).

Dryp-dryp-effekten betyder, at mange små hverdagssexistiske oplevelser samlet set hober sig op og på den måde kan ramme hårdt (Einersen et all 2021). Den digitale spredning forstærker dryp-dryp-effekten på digitale platforme, fordi hverdagssexistisk indhold når ud til langt flere end i det fysiske rum. Og dermed eksponeres den enkelte deltager også i højere grad.

Hvis moderatorer alene ser (og reagerer) på digital hverdagssexistisme som isolerede tilfælde, så får man ikke greb om problemets karakter og kan ikke gribe passende ind.

Digital hverdagssexisme eskalerer nemmere

Fraværet af ansigt-til-ansigt-møder kan være med til at eskalere konflikter – og samtidig give en minimal følelse af forpligtelse og ansvar over for dem, man deler et online rum med. Dette kan forstærke den digitale hverdagssexisme, når man ikke kan se og mærke hinanden.

Sproget er nøgle til forandring

Hverdagssexismen er indlejret i vores sprog, kultur og normer. Sproget er centralt i forhold til hverdagssexisme og digital hverdagssexisme, da sproget ikke alene afspejler verden, men også i en digital sammenhæng konstant er med til at skabe den verden, vi lever og interagerer i (Søndergaard, 1996).

Moderation kan derfor også skabe forandringer i forhold til digital hverdagssexisme. Ud over at moderere debatter og moderere tonen kan moderationen sætte en ny standard for en sproglig kultur uden sexisme. For at kunne lykkes kræver det viden om hverdagssexisme og de rette redskaber.

Digital hverdagssexisme bliver overset

Det er en velkendt problematik og en generel tendens, at det kan være svært at genkende og modvirke sexisme og hverdagssexisme i praksis (Einersen et al., 2021).

Gennem feltarbejde og interviews ser vi også i *Digital hverdagssexisme*, at moderatorer har meget svært ved at identificere og håndtere hverdagssexistiske ytringer online. Blandt andet fordi de ser mange voldsomme diskriminationsformer i deres kommentarspor generelt i deres hverdag.

”Sexisme er blevet så indgroet i vores kommentarer, at vi ikke ser det. Vi spotter det ikke med det samme, når vi skal læse en masse kommentarer. Racistiske udtryk virker derimod mere voldsomme, og dem reagerer vi også hurtigere på”. (Moderator, kvinde i 20’erne)

Moderatorerne fortæller, at mens hadefulde kommentarer og eksplicitte trusler ofte er tydelige at genkende, så er generelle sexistiske kommentarer vanskeligere at se, da de optræder hyppigere og bliver en del af, hvad man opfatter som et normaliseret kommentarspor.

Det er problematisk, fordi hverdagssexisme kan have konsekvenser for den enkelte. Men også fordi hverdagssexisme lægger grunden til, at grovere former for sexisme opstår.

Eksempel på digital hverdagssexisme

”Jonna en skam, at du ALDRIG kan få samme værdi som en mand, selv om du bruger maskuline udtryk om dig selv. Det er kun ord. Det ændrer ikke ved, at du er en kvinde”.

Værktøjer til at genkende digital hverdagssexisme

De seks undertemaer i den digitale hverdagssexisme har vi identificeret på baggrund af data fra Facebook, feltarbejde og interviews med moderatorer i mediehuse, forskning på feltet, samt de til projektet udviklede søgenøgle og algoritme.

Temaerne udgør det nederste lag i den digitale sexismetrekant og er vigtige opmærksomhedspunkter for moderation af hverdagssexisme.

Vi har på den baggrund udviklet en konkret 'kodebog', der sammen med den digitale sexismetrekant kan anvendes af moderatorer til at identificere og håndtere sexisme og hverdagssexisme (se bilag 4).
De forskellige definitioner giver mulighed for at analysere, hvordan debatter om ligestilling og køn bliver afsporet og overtaget af en grov tone, som er sexistisk og/eller hverdagssexistisk. Ligesom det giver mulighed for at udvikle moderationsstrategier, der kan anvendes i praksis for at modsvare de forskellige former for digital hverdagssexisme.

Debatter om køn kræver moderation

Ikke alle Facebook-kommentarer kræver samme grad af moderation. Indhold om køn kræver markant flere ressourcer og mere opmærksomhed end andre emner. Det viser vores feltarbejde og interview med moderatorer fra mediehuse.

Særligt indhold med transkønnede, nonbinære og kvinder, der træder lidt uden for normerne, fremkalder sexistiske kommentarer og hård debat.

Flere moderatorer beskriver, at de hårde kommentarer kommer ved debatter om emner, der "kan kategoriseres som woke".

Køn og ligestilling er en angrebstung debat

Moderatorernes observation om, at emner om køn skaber debat og hadefulde angreb på mediers Facebook-sider, bliver bekræftet af undersøgelser og forskning om hadtale og diskrimination på sociale medier.

Debatten om køn er meget angrebstung. Det viser Analyse & Tals undersøgelse "Angreb og had i den offentlige debat på Facebook", som er baseret på 73 millioner Facebook-opslag og kommentarer fra danske medier og politiker-sider i perioden 2021-2024.[1] Angreb defineres i undersøgelsen som stigmatiserende, nedsættende, krænkende, stereotypiserende, ekskluderende, chikanerende eller truende ytringer. 21 procent af kommentarer om kvinder indeholder angreb. For mænd er andelen 15 procent og for kønsminoriteter 12 procent (Analyse & Tal et al., 2025).

Undersøgelsen viser også, at de angrebstunge kommentarspor om ligestilling, kønsminoriteter og internationale konflikter er blevet markant hårdere og større mellem 2021 og 2024. Emnerne *ligestilling* og *kønsminoriteter* er i perioden steget med omkring 30 procent, mens andelen af angreb er steget med knap 50 procent (Analyse & Tal et al, 2025).

De kønsbaserede debatter er altså hårde kommentarspor, som kræver flere ressourcer til moderation end andre.

Moderatorer forbereder sig, men har individuelle strategier

Når moderatorer skal formulere Facebook-posts, med for eksempel deling af artikler eller andet indhold, kan de med vinkling og ordvalg påvirke, hvordan en debat udvikler sig. En moderator kan således se en potentielt sprængfarlig artikeloverskrift og foreslå en anden til den digitale version. Ofte er det nemlig overskriften og underrubrikken, som brugerne tager udgangspunkt i, når de debatterer – og ikke selve artiklen.

Moderatorerne som vi har været i kontakt med forholder sig på forskellige og ofte individuelle måder til risikoen for de angrebstunge debatter og muligheden for at påvirke debatten fra start.

Nogle moderatorer forsøger aktivt at undgå en ophedet debat gennem ændring af vinkling eller ordvalg. Disse moderatorer fortæller, at de for at undgå hårde kommentarspor forsøger at underspille kønsudtryk og begreber, som kan opfattes som "woke" eller "queer".

Andre moderatorer opfatter selve debatten som væsentlig og holder principielt fast i ordvalg. For nogle moderatorer kan det således være en vigtig tilkendegivelse at skrive "de/dem" om en nonbinær kunstner i overskriften – også selv om de ved, at det vil skabe debat om binære kønsopfattelser og diskrimination af nonbinære personer.

Angreb smitter og kræver fokus fra start

For flere af de moderatorer, som vi har fulgt under feltarbejdet, er det vigtigt at arbejde med overskrifter for fra starten at kunne styre debatten i en mere hensigtsmæssig retning.

Vigtigheden af arbejdet med at mindske angreb i kommentarsporet på Facebook understreges af Analyse & Tals undersøgelse. Hvis et opslag på en mediesides Facebook-side indeholder et angreb, kommer der 310 procent flere angreb i deres kommentarspor, end hvis et opslag ikke indeholder et angreb (Analyse & Tal et al., 2025).

Flere moderatorer har observeret, at hvis en af de første kommentarer på et indhold er nedladende, bliver det ofte anslaget til resten af debatten. Derfor prioriterer de at holde øje med og moderere i kommentarsporet, så snart nyt indhold lægges op.

Dette bakkes ligeledes op af Analyse & Tals undersøgelse, der viser, at angreb smitter fra kommentar til kommentar. Hvis en bruger skriver et angreb i kommentarsporet, er der mere end dobbelt så stor sandsynlighed for, at der kommer et angreb som svar (Analyse & Tal et al., 2025).

Angreb har altså en slags boomerang-effekt i den digitale debat. Og starten på en debat er retningsgivende for resten af kommentarsporet.

Anbefalinger

Til moderatorer og til mediehuse og andre, der leder og anvender moderatorer på deres Facebook-sider

Som vi har vist lever hverdagssexismen – den subtile sexisme – i bedste velgående digitalt. Og den har antaget nye former, som kan gå helt under radaren hos for eksempel moderatorer.

Netop fordi den er svær at opdage, er den også svær at modvirke. Men hvis mere subtile former for sexisme overses og overhøres, så er det ikke muligt at komme den digitale sexisme og diskrimination til livs. Derfor må man forstå sexismens store spændvidde på den digitale arena for at kunne håndtere og modvirke den.

Viden og forståelse af, hvad sexisme og hverdagssexisme indebærer, og anerkendelse af eksistensen og omfanget af sexisme som et systemisk problem er første skridt for at kunne forebygge og håndtere det.

Brug den digitale sexismetrekant til at forstå og genkende sexisme
Når der er en klar og fælles definition af sexisme, er det lettere at få øje på og anerkende sexisme som et problem. Her kan sexismetrekanten hjælpe til at genkende og forstå sexisme, og hvorfor digital hverdagssexisme kræver moderation. Det er afgørende for at kunne spotte digital hverdagssexisme i kommentarsporene og dermed kunne forebygge og håndtere den.

Brug de konkrete definitioner og eksempler fra kodebogen for at genkende digital hverdagssexisme
Fordi hverdagssexisme oftest er normaliseret kan det være svært i praksis at genkende og spotte de forskellige former for kommentarer. Eksempler og definitioner kan give en fælles forståelse og ensretning for moderatorer, som et første skridt på vej mod at modvirke digital hverdagssexisme gennem moderationsstrategier.

Prioriter fokus på angrebstunge debatter og sæt tidligt ind
Digital hverdagssexisme kan nemt eskalere grundet det grænseløse digitale rum. Her kan viden om, hvilke emner der er angrebstunge og derfor ressourcekrævende, øge muligheden for at sætte tidligt ind og mindske sexisme og hverdagssexisme i kommentarsporet.

Skab fælles strategi og retningslinjer for oplæg og opstart af debatter
En fælles strategi for opstart af debatter (for eksempel formulering af overskrift, tone og ordvalg) samt håndtering af eskalering af debattonen og angreb i kommentarsporet (se mere i kapitel 2 om moderationsstrategier) er vigtig for at forebygge og mindske digital hverdagssexisme og angreb.

2

Moderationsstrategier til at bekæmpe digital hverdagssexisme med

Moderation af hverdagssexisme er forebyggende

Den subtile form for sexisme er svær at spotte og sætte ind overfor. Men som vi har set på de foregående sider er hverdagssexismen både skadelig i sig selv og skaber grundlaget for mere grove former for sexisme og overgreb, digitalt og fysisk.

Hvis vi skal modvirke sexisme, så skal vi altså også ramme de mere subtile former for sexisme, kaldet hverdagssexisme.

Den grove form for sexisme er ifølge vores interviews og samtaler med moderatorer en virkelig ressourcekrævende opgave. Derfor er det oplagt at sætte systematisk ind over for hverdagssexismen og derved på sigt medvirke til at reducere indsatsen over for grov sexisme.

Det kræver en målrettet indsats at modvirke hverdagssexisme, fordi den er subtil og ofte en grundlæggende, men usynlig del af kulturen – det som vi alle oplever eller er vidne til , men som vi ofte også er blevet blinde overfor. Som vi har set, giver moderatorerne også udtryk for, at det er vanskelig at identificere den digitale hverdagssexisme.

Når man ikke får sat effektivt ind over for hverdagssexismen, vil der ske et gradvist skred i, hvad der opfattes som normalt og naturligt. Det kan derfor også være vanskeligt for moderatorer at vurdere, hvornår noget enkeltstående er over stregen og i strid med den gode tone.

Hvis hverdagssexismen ikke udfordres, bliver den yderligere normaliseret og kan blive en integreret del af kulturen på digitale platforme.

Moderationspraksisser

Viden om sexisme er afgørende for at kunne 1) genkende den og 2) at kunne handle på den. Det viser forskning om sexisme og kulturforandringer generelt (blandt andet Krøier et al., 2024). Det bekræftes også af vores GenderLABs og feltarbejde med moderatorer.

Et afgørende første skridt til at forebygge, håndtere og mindske sexisme digitalt er således viden om og forståelse af sexismens grundlæggende mekanismer og udtryksformer. Moderation er med til at sikre, at den digitale debat ikke bliver stødende, skadelig eller direkte ulovlig.

Kort fortalt går moderation ud på at påvirke en digital samtale for at skabe den ønskede dynamik, retning og stemning. Det kan ske ved at kommentere eller fjerne indhold, der er stødende, irrelevant eller af andre grunde uønsket. Moderator kan også vælge at blokere brugere. Der er flere forskellige måder at moderere online. De væsentligste er listet her til højre.

Overordnet er de fleste danskere positivt indstillet over for moderation af debatter online. Det viser en undersøgelse fra 2024. (Center for Sociale Medier og Demokrati, 2024)

Den usynlige, reaktive og servicerende moderation er den hyppigst anvendte moderationsform i forhold til sexisme og hverdagssexisme. Det viser vores interviews, feltarbejde og GenderLABs. Men forskning viser, at synlig moderation bedre er i stand til at mindske sexisme og skabe en god samtale digitalt.

På de kommende sider præsenterer vi forskellige moderationspraksisser og går i dybden med aktiv moderation med fokus på empatisk modtale som en strategi til at bekæmpe hverdagssexisme digitalt.

Typer af moderation

Usynlig moderation
Moderator sletter eller skjuler kommentarer. Moderator kan også sende en personlig besked med forklaring på sletning, men denne besked er usynlig for de andre borgere.

Reaktiv moderation
Moderator reagerer på kommentarer fra borgere, der eksplicit afkræver svar fra moderator.

Servicerende moderation
Moderator fikser og ordner det tekniske, men viser ikke aktivt og synligt, hvad der overtræder den gode tone eller retningslinjerne.

Faciliterende moderation
Moderator støtter et community ved at sætte både indhold og brugere i spil.

Dialogisk, aktiv og synlig moderation
Moderator stiller spørgsmål, blander sig og er synlig som moderator ved for eksempel at sætte en tydelig ramme for debatten.

Personlig moderation
Med dialogisk moderation , som adresserer den enkelte kommentar specifikt og/eller henvender sig personligt til den, der har skrevet kommentaren.

Generel moderation
Moderator er aktiv og skriver kommentarer af generel karakter til kommentarer med en generisk henvisning til for eksempel retningslinjer.

Community moderation
Her er det op til borgerne selv at markere eller rapportere problematisk indhold. Det kan resultere i, at indhold forbliver synligt og ikke-modereret i længere tid.

Usynlig moderation er den primære praksis

Usynlig moderation er den hyppigste moderationspraksis. Det viser vores feltarbejde, interviews og GenderLABs.

Ofte forsøger moderatorerne at fastholde et billede af, at alle ytringer er velkomne. Samtidig fortæller moderatorer, at de generelt sletter kommentarer "hellere én gang for meget end én gang for lidt".

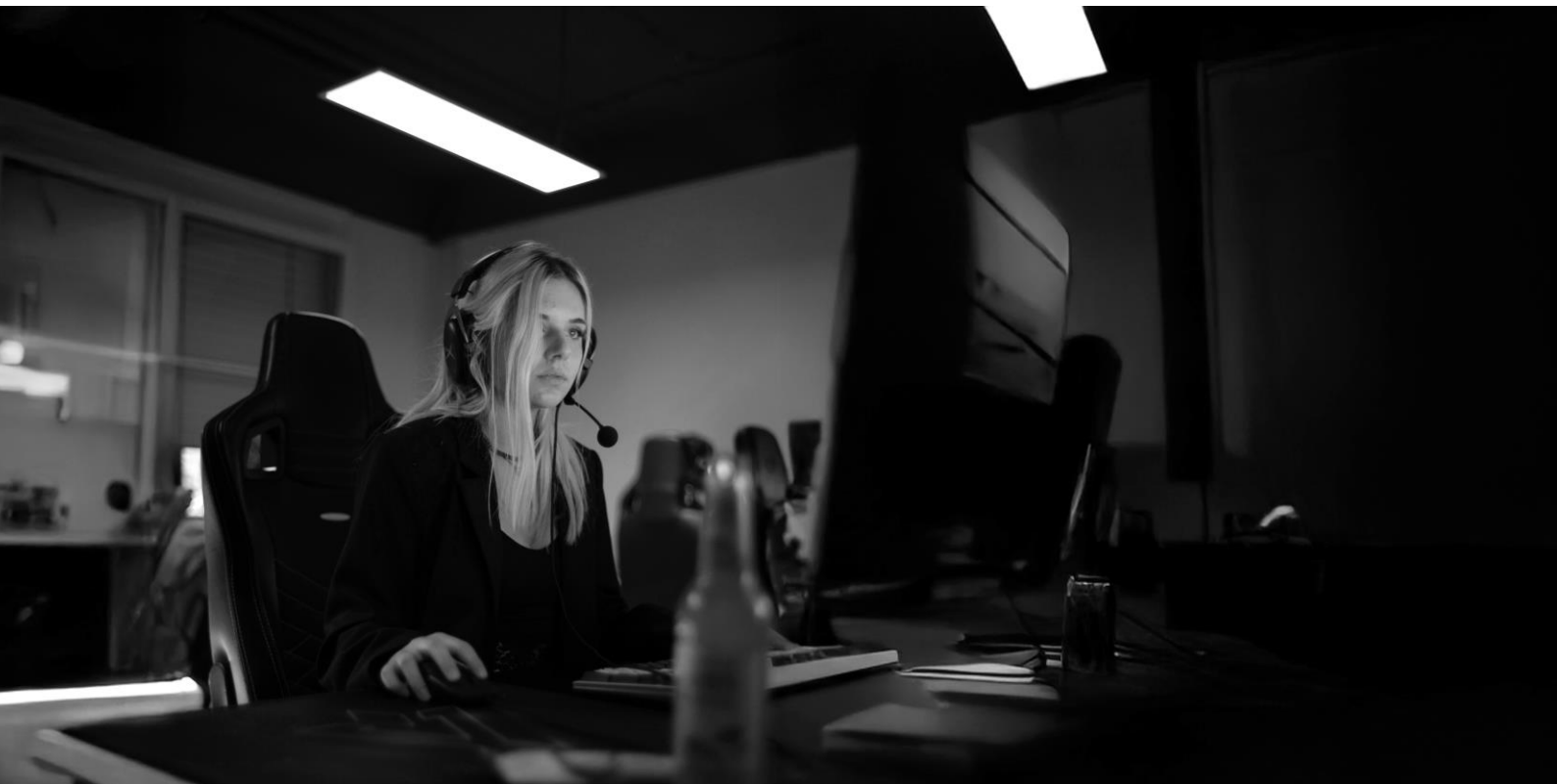
Det gælder dog ikke i forhold til sexisme og hverdagssexisme. Mange moderatorer har nemlig vanskeligt ved at identificere sexisme og hverdagssexisme digitalt. De oplever det også som svært at sætte ind overfor. Forskning viser, at det er langt lettere for moderatorer at identificere racisme end sexisme (Hawkins, 2023).

Når moderatorerne er usikre, bringer de oftere den usynlige, reaktive og servicerende moderation i spil. Det kan for eksempel være ved at støtte op om positive kommentarer ved at like eller markere med et hjerte.

Hvis der kommer et direkte spørgsmål om, hvorfor noget er modereret, forsøger moderator primært at svare i generelle vendinger.

Konsekvensen af usynlige moderation er, at sexistiske kommentarer, som får lov til at blive stående på platformen, virker legitime (Gillespie, 2018).

Derudover kan borgerne have svært ved at gennemskue og afkode reglerne for debatten. Det kan endda virke som om, at der slet ikke er nogen, som tager ansvar for at sikre en god debat.



Synlig moderation viser størst effekt

Flere studier viser, at synlig moderation kan skabe et mere inkluderende og trygt debatmiljø i grupper på sociale medier (Masullo et al., 2022).

Aktiv moderation gennem dialog har større effekt på diskriminerende kommentarer end, hvis moderator blot sletter eller skjuler kommentarer.

Effekten af synlig moderation er især undersøgt på Facebook og Reddit. Et eksperiment på Reddit viser, at grupper med en synlig moderator har et langt mere positivt sprog end uden moderator. Synlig moderation fører til mere *selvmoderation* – for eksempel at privatpersoner anmelder grov tale og opfordrer til bedre sprog (Gibson, 2019).

Dialogisk moderation

Synlig moderation går aktivt, dialogisk og tydeligt ind med modtale, hvis tonen går over stregen og for eksempel er sexistisk eller hverdagssexistisk.

Modtale kan benyttes som en direkte respons på hadefulde ytringer og sexistisk tone. Her er formålet at konfrontere ytringen i stedet for blot at fjerne den (Hangartner et al., 2021).

Modtale virker bedst, hvis den er rettet mod specifikke personer, og hvis den indeholder en grad af empati (Muhr ,2019). Dette viser sig også i vores interviews, GenderLABs og feltarbejde: De moderatorer, der anvender modtale, havde god erfaring med at kombinere det med forståelse og empati – afhængigt af hvor grov kommentaren var.

Personlig eller generel moderation

Synlig, aktiv moderation har størst positiv effekt, når den henvender sig direkte til den person, der skriver (kaldet *high person centered*). Modsat mere generelle og upersonlige kommentarer (kaldet *low person centered*).

Det kan måske virke udfordrende at henvende sig direkte til afsenderen af en ubehagelig kommentar. Men en undersøgelse fra 2022 viser, at personlige kommentarer har mere positiv effekt end generelle kommentarer (Masullo et al., 2022). . Det er dog vigtigt, at moderator i sit modsvar viser emotionel forståelse over for den person, moderator henvender sig til.

Studiet viser også, at personlige kommentarer fra moderator har den bedste effekt på brugernes syn på nyhedssiden, dens håndtering af uanstændige angreb og bedre kontakt til Facebook-brugerne generelt (Ibid.).

Empatisk modtale som moderationsredskab

Empatisk modtale kan være en effektiv måde at udfordre en ubehagelig kommentar og forsøge at påvirke tonen, kulturen og normerne for den digitale debat (Hangartner et al., 2021).

Som redskab kan modtale balancere hensynet til ytringsfriheden ved at kommentere på en holdningsytring i stedet for at slette den. Ulempen kan dog være, at man risikerer at bidrage til en diskussion, som avler mere had eller sexisme. Derfor skal modtalen være velovervejet – og indeholde empati.

Modtale er en direkte reaktion på hadefuld eller skadelig tale med henblik på at konfrontere ytringen. Hensigten med modtale er at reducere had.

Her definerer vi modtale som en direkte respons fremsat af en moderator eller en aktiv allieret med det formål at:

- få den pågældende til stoppe
- få den udsatte til at føle sig i et mere sikkert rum
- delegitimere hverdagssexisme, så andre også finder det mere illegitimt at sprede hverdagssexisme i kommentarsporet.

Empati har bedre effekt

Det er vigtigt, at modtalen ikke fordømmer eller udskammer, men er empatibaseret.

Ved at fremkalde empati kan moderator menneskeliggøre dem, som er berørt af den hadefulde eller sexistiske kommentar – og minde afsender om, at folk kan blive skadet eller såret af det, afsenderen skriver.

Ved at kommentere empatibaseret er moderator desuden selv en rollemodel: Man praktiserer den empati og konstruktive tone, som man selv efterlyser i debatten.

Empatisk modtale rettet mod det stille flertal

Ofte fokuserer moderatorer på de aktive brugere og deres ytringer. Men med empatisk modtale kan man også nå den passive majoritet, der blot følger med i for eksempel en debat uden at ytre sig.

Med empatisk modtale kan moderator sprede en inkluderende tone og fremme et venligt miljø, der påvirker alle brugere både de aktive og de stille. Og det kan få færre til at trække sig fra debatten (dangerous-speech.org/counterspeech).

Blandt de interviewede moderatorer og andre, som benytter sig af modtale, er der tydelig enighed om, at et centralt formål er at forsøge at påvirke den stille majoritet.

Empati og tydelighed

Empatisk modtale er modelleret på baggrund af særligt erfaringer med modtale mod digital racisme. Et studie af kommentarer på Twitter har testet nedbringelsen af racistisk hadtale ud fra tre strategier for modtale: empati, advarsel om konsekvenser og humor.

Resultaterne viser, at empati har den bedste virkning på at mindske forekomsten af racistisk hadtale, samt øge afsenderens sletning af egne racistiske kommentarer (Hangartner et al., 2021).

De moderatorer vi har interviewet, som anvender modtale, har også erfaring med, at det virker i forhold til at modvirke og dæmpe sexisme og hverdagssexisme. Men også, at det kan være vanskeligt at praktisere, da det meget er op til den enkelte moderators erfaring og 'mod', om man kan og vil bringe modtale i spil. Hvilket bevirker, at mange afholder sig fra det.

Anvendelse af modtale kræver eksempler og træning

Moderatorerne fortæller således, at de i meget lille udstrækning anvender modtale. Men samtidig, at modtale er meget effektfuldt, hvis det bliver "gjort ordentligt", hvilket ifølge moderatorerne kræver en del erfaring og træning. Det kan tage tid at lære at ramme plet med en kommentar, da man blandt andet skal kende og have en god fornemmelse af debatten om for eksempel køn og ligestilling og have en fornemmelse af, hvem man har med at gøre. Alt sammen forhold, som gør, at mange moderatorer holder sig tilbage fra at anvende modtale.

For at gøre det muligt at implementere brugen af modtale skal der viden, konkrete værktøjer og træning til. Det vender vi os mod på næste side.

Derudover kan det være godt med mulighed for sparring med kolleger. Hvis man som moderator personligt bliver ramt af en kommentar, er det særligt godt at inddrage et ekstra par øjne, så modtale ikke bliver skrevet i affekt, men at der stadig kan være plads til at man griber den kommentar, man modsvarer.

Empatisk modtale – et værktøj til praksis

At tilskynde til empati og at se problemet fra forskellige perspektiver kan reducere fjendtlighed over for marginaliserede grupper. Samtidig kan man udvise empati med afsenderen ved at indgå i en dialog – i stedet for blot at fordømme kommentaren eller slette den – og dermed øge muligheden for at påvirke afsenderen.

Når I skal i gang med at anvende empatisk modtale i praksis, anbefaler vi, at I på forhånd formulerer konkrete forslag.

Så kan moderator lade sig inspirere af dem, når moderator skal i gang med at moderere for eksempel en debat på jeres sociale medier.

Empatisk modtale har følgende elementer, som bør indgå:

- Empati for udøveren: Denne kan variere afhængig af kommentarens grovhed, men også af behovet for at vise empati for ofret.
- Undgå at opstille modsætningen os/dem: Skriv for eksempel: "Lad os ...", "Skal vi ikke ..."
- Anvis handling: "Lad os holde den gode tone ved at gøre sådan og sådan ..."

Herunder vises eksempler på empatisk modtale som værktøj ved henholdsvis grovere og mindre grove grader af hverdagssexisme.

Eksempler på empatisk modsvar

Svar til hverdagssexisme

- Udvis empati med både udøver og modtager: "Jeg kan godt forstå dig, men ..."
- Vis en handlemulighed: "Lad os ..." eller inviter til dialog ved at skrive "Vi vil gerne høre din, hvad du mener om emnet ..."

Eksempler på modtale til mindre grov hverdagssexisme:

- "Det er helt i orden, du ikke er enig i deres politik, men lad os have en god debat ved at fokusere på emnet i stedet for deres køn. Det er nemt at misforstå, når vi diskuterer følsomme emner. Diskuter gerne - men vær altid gode ved hinanden."

- "Det er helt OK med forskellige holdninger. Men at bruge udtryk som "det kvindemenneske" kan opleves nedværdigende. Vi vil hellere høre, hvad du tænker om emnet."

Svar til grovere sexisme

- Udvis mere sympati for modtager end for udøver.
- Vis at kommentaren går over grænsen – er sexistisk, diskriminerende etc.
- Kom gerne med eksempler på, hvilke ord der er over grænsen.

Eksempler på modtale til grov sexisme:

- "Det er voldsomt sprog, når du skriver xxx. Lad os holde fokus på emnet i stedet for at komme med nedværdigende/seksuelle kommentarer om kvinder."

- "Det flytter fokus og er ikke respektfuldt at bruge skældsord på den måde. Lad os have en god debat om det egentlige problem."

Anbefalinger

Til moderatorer og til mediehuse og andre, der leder og anvender moderatorer på deres Facebook-sider

Moderation er vigtigt, og noget de fleste danskerne bakker op om. Derfor skal man ikke afholde sig fra at moderere online af frygt for, at brugerne over en bred kam ikke vil billige det. Gennem aktiv, synlig og dialogbaseret moderation kan man modvirke hverdagssexisme. Med empatisk modtale kan moderator mindske den sexistiske tone og kultur.

Anvend synlig moderation for at delegitimere hverdagssexisme
Ved af anvende synlig moderation viser man, hvad der er ønskelig debattone på siden. Man kan påvirke både udøvere og modtagere af det hverdagssexistiske indhold – men også det stille flertal, der ser med, men ikke interagerer med indholdet på siden.

Anvend aktiv, personlig og dialogisk moderation baseret på empatisk modtale for at ændre sexistiske normer
I praksis betyder personlig moderation, at kommunikationen henvender sig direkte til en person og ikke har en generisk karakter. Ved at gå i dialog med afsæt i empati, øges muligheden for at påvirke modtageren og dem, der ser med.

Udarbejd konkrete eksempler på modtale, som kan anvendes i praksis med inspiration fra modtale-værktøjet
Ved at have et sæt konkrete eksempler på modtale bliver det nemmere og mindre ressourcekrævende at anvende strategien. Find inspiration i værktøjet på foregående side.

3

Barrierer for moderation af digital hverdagssexisme

Centrale barrierer for mere synlig moderation

Moderation er en krævende disciplin, og der er mange faktorer i spil, når debatten skal vurderes og styres.

Vi ved fra forskning og praksis ved, at synlig moderation er den bedste moderationsstrategi. Alligevel er der flere barrierer, som skal løses, før man kan implementere sådanne strategier i arbejdet med moderation i større teams, som for eksempel hos mediehusene.

I vores undersøgelse har vi identificeret tre centrale barrierer:

Arbejdsforhold og organisering: Der er en række *organisatoriske forhold*, der udgør en væsentlig barriere. Antallet af kommentarer er ofte mange, og moderatorer sidder alene med arbejdet – med begrænsede muligheder for oplæring, sparring og kompetenceudvikling.

Arbejdet baseres på bias og mavefornemmelser: Vi har beskrevet udfordringerne med uvidenhed om problemet og dets konsekvenser. Det fører til tvivl på egen dømmeevne i forhold til at kunne vurdere, hvornår noget er over stregen. Samt oplevelsen af at agere ud fra *mavefornemmelse* i mangel på viden om, hvordan man håndterer moderationen ud fra en faglig vurdering.

Ideal om neutralitet og frygt for eskalering: Som moderator balancerer man mange hensyn samtidig, og her står *hensynet til ytringsfrihed og idealet om den neutrale moderator* frem som en væsentlig barriere for synlig moderation.

Frygten for eskalering af en diskriminerende debat og frygten for personlige konsekvenser for moderatoren selv kan også spille ind.

Manglende retningslinjer

Manglende retningslinjer for moderation af sexisme og hverdagssexisme gør det ekstra vanskeligt for moderator at gribe synligt og aktivt ind over for problematiske kommentarspor. Da det fører til individuelle strategier og er med til at forstærke barriererne.

Som vi uddyber i dette kapitel får disse barrierer ofte moderatorerne til at anvende usynlig moderation.

Arbejdsforhold og organisering

Moderationsarbejdet er i høj grad udført af praktikanter, personer i løse ansættelser eller junior-positioner, som kan være et springbræt ind i et højere rangeret arbejde i organisationen. Dermed er moderatorer oftest unge, studerende og deltidsansatte i skæve arbejdstider. Herudover kan moderation også i høj grad været baseret på frivilligt arbejde, for eksempel i politiske sekretariater. Det kan være med til at begrænse muligheden for læring og faglig sparring.

Der er ingen eller meget begrænsede muligheder for uddannelse eller kurser i moderation. Moderatorer oplæres oftest ved at sidde med på et par vagter, inden man så selv skal sidde med det alene.

Manglende muligheder for sparring er noget, flere af de interviewede moderatorer selv peger på som afgørende for deres moderationspraksis og strategier. Når moderator sidder alene i de sene arbejdstider, er det svært at finde nogen at sparre med, hvis man kommer i tvivl, som en moderator fortæller i citatet nedenfor.

Det er især i eftermiddags- og aftentimerne, at der er mest gang i kommentarerne, da den almene bruger har fri og tid til at deltage online. Derfor ender mange af de ophedede debatter med at være hos moderatorer med begrænset mulighed for sparring med kolleger eller ledere.

"Hvis det (en kommentar) er noget, jeg personligt bliver ramt af, så spørger jeg dem omkring mig. Men der sidder ofte ikke andre moderatorer, da jeg tit sidder alene. Vi mødes ikke så tit og kan derfor ikke erfaringsudveksle i de sene aftentimer"
(Moderator i mediehus)

De fleste moderatorer afholder sig fra at kontakte en kollega eller leder i dennes fritid, fordi de oplever, at det skal være tilstrækkeligt alvorligt – og være "kørt af sporet" i kommentarsporet, før de kan (tillade sig at) tage kontakt.

Manglende ressourcer og tid til den enkelte kommentar kan være med til at afgøre moderationspraksissen for den enkelte moderator, som moderatorer deler i citaterne nedenunder.

Senior-moderatorer og ledere prøver at støtte moderationen og afhjælpe ressourceforbruget ved at forsøge at forudsige, hvilke debatter der kræver ekstra opmærksomhed. Som en del af at klæde moderatorerne på, kan de indimellem få Q&A's om indholdet, som de kan bruge til at besvare eventuelle kommentarer, eller udkast til svar, som moderator kan sætte ind, hvis en person kritiserer noget specifikt i kommentarsporet.
Disse organisatoriske barrierer leder til individualiserede strategier og tilbageholdenhed over for at bruge aktiv moderation.

Mængden af kommentarer og det, at man sidder alene med opgaven, gør, at moderatorerne føler sig tvunget til at tage hurtigere beslutninger. Det tager ofte lang tid at finde gode modsvar. Derfor tyr de oftere til en mere usynlig praksis, hvor de sletter frem for at svare. De forsøger at fjerne indhold, så der ikke står noget diskriminerende for længe, der kan være med til at dreje debatten i en uhensigtsmæssig retning.

"Jeg føler heller ikke, at man har sindssygt meget tid til at dvæle ved alle kommentarer hele tiden"
(Moderator i mediehus)

"Hvis du vil lave en ordentlig kommentar tilbage, der rammer plet, tager det tid, fordi du skal tage en masse forbehold og forstå debatten, og hvem det kommer fra"
(Moderator i mediehus)

Bias og mavefornemmelser

Fra feltarbejde, interviews og workshops ved vi, at der ofte mangler retningslinjer for moderation af sexisme og hverdagssexisme. I stedet for konkrete retningslinjer bliver moderators arbejde i højere grad baseret på egne erfaringer og intuition. Fra forskning i moderation og content management ved vi, at der kan opstå bias, når moderatorer skal vurdere for eksempel sexisme (Gervais and Hillard, 2014).

Forskning viser, at kvinder er hurtigere til at opfatte noget som sexistisk end mænd (Buie, H. & Croft, A. 2023). Moderators køn kan altså påvirke moderationen. Men moderators køn påvirker også, hvordan brugere reagerer på moderationen.

Udfordringer med bias er noget, vi hører fra flere moderatorer: De taler om en "mavefornemmelse" med vægt på egne erfaringer og personlige vurderinger, der fører til individuelle strategier for moderation af digital hverdagssexisme.

Det er også tydeligt, at sexisme og hverdagssexisme modsat for eksempel racisme nemt kan gå under radaren for moderatorerne – fordi det ikke er noget, de nødvendigvis har viden om, opmærksomhed på eller tydelige retningslinjer om.

"Altså der er ikke de her helt klare retningslinjer, så det er tit en vurdering og en mavefornemmelse, og jo længere tid man har været her, jo flere gange har man prøvet det. Men særligt de nye kan stadig godt være meget usikre på, hvad der er okay, og hvad der ikke er okay"
(Moderator i mediehus)

At moderatorernes arbejde baseres på bias og mavefornemmelse i stedet for konkrete retningslinjer er med til at begrænse den synlige moderation. Flere moderatorer fortæller, at de er tilbøjelige til ikke at gribe ind i specifikke debatter af frygt for at få skudt i skoene, at det er deres personlige holdninger, der påvirker moderationen, frem for at være mere generelle retningslinjer.

Moderatorerne kommer til at tvivle på deres professionelle dømmekraft, fordi de kan fornemme deres egen bias, men mangler viden og værktøjer til at sikre en fagligt begrundet vurdering. Der tales sjældent om bias, moderatorerne imellem, hvorved det bliver en individuel opgave at få omsat og forholdt sig aktivt til egne erfaringer og mavefornemmelser, og spore sig ind på det generelle kodeks for god debat og moderation.

"Jeg tror bare, at det er mavefornemmelsen. Jeg diskuterede, da jeg havde overlevering med min kollega, i den første tid, hvor jeg kiggede hende lidt over skulderen. Der kunne jeg mærke, at hun, fordi der jo sidder en anden person, selv blev ekstra opmærksom på, om hun ville slette eller ikke ville slette en kommentar. Normalt kører det måske bare meget på automatik"
(Moderator i mediehus)

Ideal om neutralitet og frygt for eskalering

Idealet om, at en moderator fremstår som neutral, og hensynet til 'ytringsfrihed' står centralt i moderatorernes overvejelser.

Moderatorerne har mange refleksioner om gråzoner for, hvornår noget går fra at være en holdning til at være diskrimination.

En moderator beskriver, hvordan det at værne om uenigheden og ytringsfriheden er vigtigt, men også vanskeligt i citatet nedenunder.

Flere moderatorer fremhæver et udtalt hensyn til ytringsfriheden, som de gerne vil værne om på deres medieplatforme. Moderatorer vælger derfor i højere grad at lave usynlig moderation, hvor de sletter uden at kommentere.

Debatter om køn og ligestilling står centralt i dette dilemma, da det i vid udstrækning opfattes som et holdningsspørgsmål. Moderator risikerer i højere grad *ikke* at fremstå som neutral, hvis de laver aktiv og synlig moderation.

Ytringsfrihed for nogle er ikke nødvendigvis ytringsfrihed for andre, hvis debatten overtages af hadefulde og diskriminerende ytringer, som afholder nogle grupper fra at deltage. I dilemmaet om ytringsfrihed bør moderator ikke blot tænke på de aktivt debatterende, men også dem, der passivt læser med og afholder sig fra selv at debattere.

"Der er meget, vi bliver nødt til at lade stå. Og det er jo et afholdenhedsprincip. Vi skal værne om, at vi er uenige også. Og det er der, hvor ytringsfriheden bliver sværest. Og det er jo det, vi sidder aktivt og arbejder med; hvad er tilladt, hvad er ikke tilladt. Fordi vi må heller ikke censurere, bare fordi, at jeg har en anden politisk overbevisning og har en mening om, hvordan man burde begå sig, og hvad man burde slynge ud på nettet"
(Moderator i mediehus)

Frygt for eskalering og konsekvenser

Det kan også være svært at gå ind og modsvare brugerne i et kommentarspor af frygt for konsekvenserne.

Flere moderatorer nævner frygt for personlige konsekvenser, hvis de kommenterer i eget navn, men også bekymring for konsekvenser for det sted, de arbejder. Her går frygten på, om man med sin moderation får en mere uhensigtsmæssig debat til at blusse op. Eller i værste tilfælde risikerer at starte en shitstorm, som går ud over mediehuset.

Frygten for, at det kan give bagslag, får moderatorer til at afholde sig fra synlig moderation. Men at der opstår denne – velbegrundede – frygt, understreger også netop behovet for at gå ind og moderere det sexistiske og hverdagssexistiske indhold for at skabe et debatklima online, som flere kan deltage i.

Det kræver – som denne rapport viser – bedre vilkår, rammer og værktøjer for moderatorers arbejde.

"Jeg tror, at jeg synes, at det indtil for nylig har været meget svært, fordi jeg faktisk har haft følelsen af, at hvis du er i tvivl, så lad være med at skrive det. Fordi jeg har sådan en frygt for, at jeg bare putter endnu mere brænde på bålet, og at jeg vågner op til, at der er en kæmpe shitstorm, fordi jeg har siddet og skrevet et eller andet"
(Moderator i mediehus)

Anbefalinger

Til mediehuse og andre, der leder og anvender moderatorer på deres Facebook-sider

Når hverdagssexisme skal bekæmpes online er det afgørende, at moderatorer og andre har den nødvendige rammesætning og træning i at håndtere sexisme og hverdagssexisme i praksis. Dette kræver en kontinuerlig og systematisk indsats med løbende sparring og opkvalificering af moderatorerne.

På digitale platforme bliver balancen i forhold til ytringsfrihed, bias og organisatoriske forhold ofte til barrierer, der forstærker tendensen til at se igennem fingre med hverdagssexisme.

Derfor er det afgørende at højne viden og skabe en fælles forståelse for, hvad sexisme og hverdagssexisme er, og hvordan man bedst bekæmper det med professionelle redskaber.

Øget professionalisering og mandat til moderatorer.
Fælles og tydelige procedure for håndtering og forebyggelse af sexisme og hverdagssexisme kan imødegå de individualiserede strategier, som bl.a. resulterer i en tendens til passiv og usynlig moderation.

Standardiserede retningslinjer med tydelige definitioner af digital sexisme og hverdagssexisme og strategier for modsvar.
Fælles retningslinjer, definitioner og svarstandarder skaber tydelige rammer for moderatorernes arbejde, og modvirker dermed de nuværende barrierer i forhold til bias og mavefornemmelse samt neutralitet, ytringsfrihed og frygt for eskalering, der i dag forhindrer anvendelsen af synlig moderation.

Bias-træning og opkvalificering af moderatorer samt mulighed for løbende sparring.
Bias-træning og tydelige retningslinjer kan modvirke at moderationen beror på 'mavefornemmelser' og intuition. Det kan også give moderatorer en fælles professionel tilgang til at håndtere de dilemmaer og gråzoner, som er et vilkåret i moderation og modvirke individualiserede strategier.

Opgør med idealet om den neutrale moderator og dilemmaet om ytringsfrihed.
Ytringsfrihed for nogle er ikke nødvendigvis ytringsfrihed for andre, hvis debatten overtages af ytringer, som afholder nogle grupper fra at deltage. Det kræver et opgør med idealet om 'neutral' moderation og en nuanceret drøftelse om 'ytringsfrihed' at skabe rammer for at moderere, så man også tager hensyn til dem, der i dag afstår fra at ytre sig.

4

Muligheder og begrænsninger ved automatiserede modera tionsløsninger

Evaluering af automatiserede løsninger

Hver dag skrives der næsten 150.000 kommentarer på danske mediers og politikeres sider og de borgerdrevne grupper på Facebook (Analyse & Tal, 2024).

Det er derfor ikke overraskende, at moderatorerne i vores undersøgelse oplever moderationsopgaven som vanskelig og ressourcekrævende.

Da moderation kræver mange ressourcer, bliver der eksperimenteret med flere typer automatiserede løsninger til at assistere eller erstatte menneskelig moderation.

Det kan være udvikling af søgenøgler, hvor man definerer konkrete ord (for eksempel stødende ord som "kælling", "svans" og "vatpik"). Det kan være brug af kunstig intelligens, der med avancerede beregninger kan analysere komplekse mønstre i tekst. Alt sammen med det formål at flage potentielt skadeligt indhold, så moderatorerne kan bruge deres tid så effektivt som muligt.

Flere moderatorer i vores undersøgelse nævner, hvordan AI-algoritmer og søgenøgler allerede indgår i deres praksis. Samtidig er de skeptiske over for at fjerne mennesket helt bag moderationen, da de ser flere faldgruber.

Søgenøgler

Søgenøgler er ordlister, der hjælper moderatorer med at identificere potentielt problematisk indhold. De fungerer som et første screeningsværktøj, der markerer kommentarer til nærmere gennemgang baseret på specifikke ord og vendinger.

Værktøjet har dog væsentlige begrænsninger. For det første kan samme ord have forskellige betydninger afhængigt af konteksten. Eksempelvis kan "hold kæft" både være negativt og positivt ment. For det andet udvikler brugere ofte kreative omskrivninger for at omgå filtreringen, som når "muslim" bliver til "müsli". Mennesker kan let afkode den reelle betydning ud fra sammenhængen, men søgenøglerne opfanger ikke disse sproglige variationer.

Moderatorerne i undersøgelsen bruger derfor søgenøgler som et indledende sorteringsredskab, men mangler ofte indsigt i værktøjets fulde ordliste og dets evne til at opfange forskellige former for krænkende indhold herunder sexisme.

AI-algoritmer

I Danmark er der udviklet avancerede AI-modellertil at identificere aggressivt sprog (Alexandra, 2022), sproglige angreb (Analyse & Tal, 2025: Analyse & Tal, 2021), hadtale (Analyse & Tal, 2021) og misogyni (Zeinert, Inie, & Derczynski, 2021).

Når der tales om kunstig intelligens eller AI, menes der som oftest en algoritme eller beregningsmodel af den type, der kaldes et *neuralt netværk*.

Neurale netværk består af flere lag af matematiske enheder til at behandle information, hvor hver enhed i et lag er forbundne med hver enhed i det næste lag. Dette har ligheder med den menneskelige hjerne, hvor hver enkelt neuron har en relativt simpel funktion, men det er deres forbindelser, der skaber vores komplekse forståelse af verden.

De mange lag gør, at denne metode også kaldes *deep learning*. Hvert lag i netværket udfører relativt simple matematiske beregninger, men når data (som for eksempel en Facebook-kommentar) passerer gennem flere lag, opbygges der en stadig mere sofistikeret forståelse af dataens indhold.

På den måde *trænes* en AI-algoritme ved at blive vist eksempler på det fænomen, den skal kunne modellere. Gennem træning af algoritmen kan man lære den at efterligne dens input.

Jo mere data algoritmen trænes på, desto bedre bliver den til at identificere og klassificere komplekst indhold som misogyni. Denne lagdelte proces er det, der gør algoritmen i stand til at identificere for eksempel had eller sexisme ud fra kommentarerne – selv om den ikke forstår sproget eller konteksten på samme måde som et menneske.

Evaluering af søgenøgler og AI-algoritmer

Det er tydeligt fra forskning og praksis, at AI og søgenøgler alene ikke er vejen til at håndtere diskrimination og sexisme på sociale medier (Dercynsky i Nyste, 2021). For selv om det lyder som en åbenlys løsning at kunne flage eller endda fjerne problematisk indhold uden at skulle bruge mennesketimer på det, er virkeligheden desværre mere kompleks – og metoderne har sine begrænsninger. Særligt når det gælder noget så subtilt og kontekstuelt som digital hverdagssexisme.

For at kunne evaluere hvor brugbare de automatiserede løsninger er, konstruerede vi et datasæt af næsten 2000 Facebook-kommentarer, hvor menneskelige annotører med udgangspunkt i vores kodemanual klassificerede kommentarerne som enten indeholdende hverdagssexisme eller ej.

Dette datasæt bruger vi til at vurdere de forskellige løsningers evne til at korrekt at identificere hverdagssexistisk indhold og deres dækning.

Bemærk: AI har udviklet sig markant siden vores undersøgelse. Nyere sprogmodeller fra OpenAI, Meta og Google kan potentielt præstere bedre, men disse teknologier blev offentliggjort efter vores evaluering i 2023 og ligger derfor uden for rapportens fokusområde.

Misogyni-algoritmen

Misogyni-algoritmen er en AI-algoritme udviklet af forskere fra IT-Universitetet i København til at identificere misogyne kommentarer på sociale medier på dansk (Zeinert, Inie, & Derczynski, 2021).

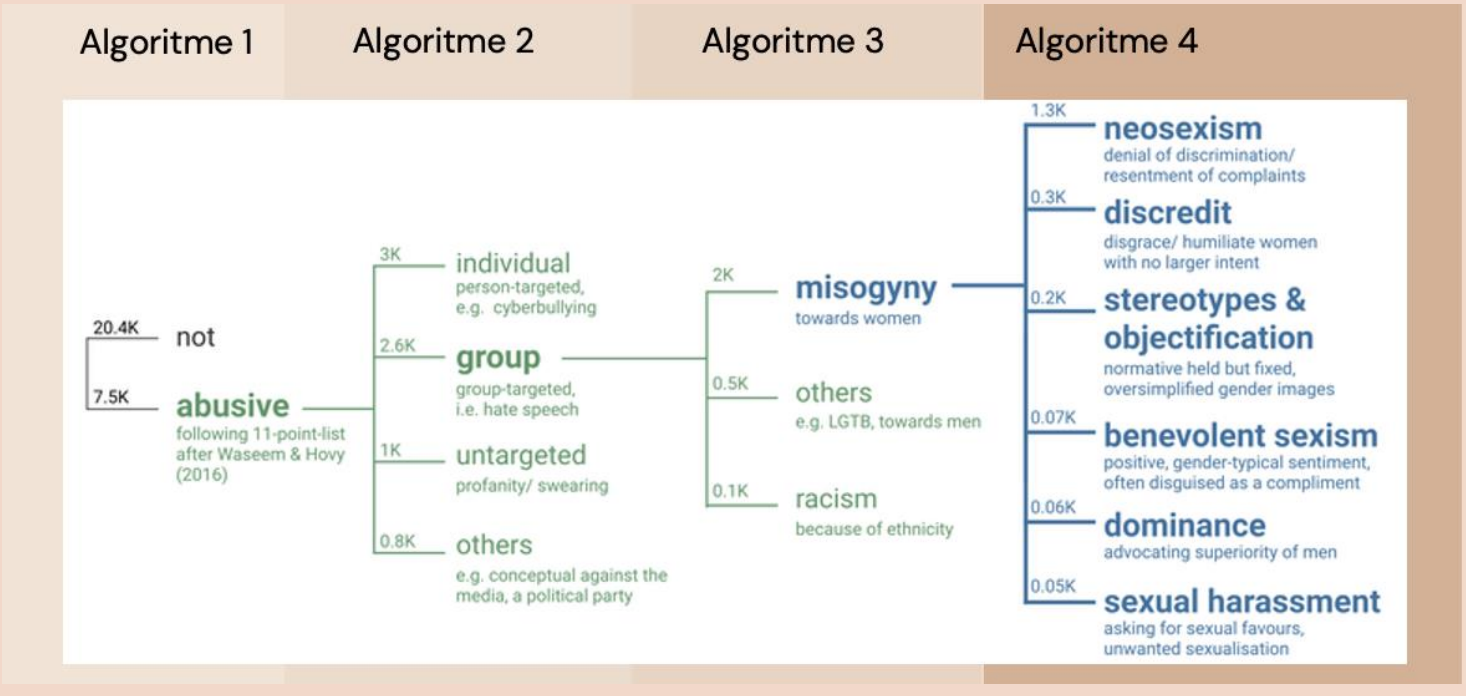
Misogyni-algoritmen bruger fire forskellige algoritmer til at analysere kommentaren over fire trin – se grafen herunder.

Først vurderer algoritme 1, om kommentaren indeholder grove eller voldelige udtalelser. Hvis det er tilfældet, går teksten videre til næste trin, hvor algoritme 2 afgør, hvem teksten er rettet mod. Hvis kommentaren er rettet mod en gruppe mennesker og ikke et individ, sendes den videre til næste trin. Algoritme 3 undersøger, om kommentaren er rettet mod kvinder, og hvis det er tilfældet, går teksten videre til den sidste trin, hvor algoritme 4 vurderer, hvilken type sexisme der er tale om.
Den trinvise tilgang giver en klar opdeling af opgaver, men betyder også, at usikkerheden stiger, da fejl propagerer fra gennem trinnene, og at del-algoritmernes unøjagtigheder forstærker hinanden.

Misogyni defineres som "hadefuldt indhold rettet mod kvinder", og derfor er denne algoritme mere snæver end projektets egen definition. Vores fokus er ikke kun på hadtale, men også på hverdagssexisme, som ofte er mere subtilt og underspillet.

Desuden tager vi højde for et intersektionelt perspektiv, hvor sexisme ikke kun er rettet mod kvinder, men kan ramme personer på tværs af køn, race, seksualitet og andre sociale kategorier. For at imødegå denne diskrepans udviklede vi en søgenøgle, der afspejler dette bredere syn på sexisme.

Opbygning af algoritmen (Zeinert, Inie, & Derczynski, 2021)



Sexisme-søgenøgle

Som et simplere digitalt værktøj udviklede vi en søgenøgle til at identificere hverdagssexistisk indhold.

Søgenøglen blev udviklet af KVINFO og Analyse & Tal baseret på projektets definition af hverdagssexisme. Den tager udgangspunkt i de seks forskellige typer af hverdagssexisme: neosexisme, velment sexisme, dominans, miskreditering, stereotyper, objektivisering og seksualisering (se kapitel 1 for definitioner).

Når vi bruger søgeordene på data, betyder det, at en kommentar bliver flaget som potentielt indeholdende hverdagssexisme, hvis et af ordene fra listen findes i Facebook-kommentarer.

For at udvikle søgenøglen anvendte vi i første omgang ord, som er identificeret i den fysiske verden i forhold til sexisme og hverdagssexisme, samt intersektionerne med for eksempel racisme og homofobi. Dette både fra egne undersøgelser og forskning om sexisme og hverdagssexisme.

Vi finpudsede derefter søgenøglen gennem afprøvning af den på data, hvor vi læste kommentarer igennem på alt, hvad nøglen i første omgang fandt. Her trådte flere ord frem som relevante, og nogle viste sig at være mindre til stede i data end først antaget. Vi så også på kontekst og på, hvilke diskriminerende ord der optrådte i forbindelse med hinanden.

For at gøre det nemmere at arbejde med projektets definition af hverdagssexisme, lavede vi en kodemanual, med inspiration fra MacQueen et al. (1998). Denne kodemanual skulle sikre, at alle i teamet havde en fælles forståelse af, hvad hverdagssexisme er, og hvad det ikke er, når vi analyserer kommentarer.

Selve søgenøglen kan ses i bilag 5 samt den udviklede kodemanual i bilag 4.

Sådan evaluerer vi løsninger

For at en algoritme er brugbar, skal den naturligvis kunne identificere de ting, den er udviklet til at finde.

Men det er også vigtigt, at algoritmen kan finde ud af at ignorere irrelevante eksempler.

Derfor laver vi en evaluering af, hvor godt misogyni-algoritmen og sexisme-søgenøglen præsterer.

For at vurdere brugbarheden af misogyni-algoritmen og sexisme-søgenøglen har vi udviklet et datasæt af Facebook-kommentarer, der er blevet kategoriserede af mennesker for, om de indeholder hverdagssexisme elle ej.

Vi fik fire (menneskelige) annotører til at klassificere 1.959 Facebook-kommentarer baseret på vores kodemanual (se bilag 4). Disse kategoriseringer betragter vi som "den sande værdi". Jo tættere algoritmen og søgenøglen kommer på de menneskelige klassificeringer, jo bedre og mere korrekte vil de fremstå i vores evaluering.

For at måle, hvor godt modellerne præsterede, gjorde vi brug af to metrikker fra maskinlæring, kaldet *precision* og *recall*.

1
Menneskelige annotører kategoriserer 1.959 Facebook-kommentarer for, om de indeholdte sexisme eller ej

2
Misogyni-algoritmen og sexisme-søgenøglen klassificerer de 1.959 Facebook-kommentarer

3
Precision og recall udregnes fra misogyni-algoritmen og sexisme-søgenøglens resultater

Precision viser, hvor stor en andel af algoritmens klassificeringer af sexisme der også af mennesker er blevet vurderet som sexistisk indhold. Precision viser dermed, hvor stor andel af klassifikationerne er korrekte.

Recall, derimod, beskriver, hvor stor en andel af de menneskeligt vurderede sexistiske kommentarer der også er blevet klassificerede som sexistisk indhold af algoritmen. Man kigger altså på, om algoritmen kan finde alt det sexisme, der er blevet vurderet af mennesker i vores evalueringsdatasæt.

Med andre ord: Precision måler, hvor ofte der bliver ramt plet, når en kommentar vurderes som sexistisk. Recall måler, hvor stor en andel af det sexistiske indhold i datasættet der også bliver vurderet som sexistisk.

Vores evaluering viser, at hverken søgenøglen eller algoritmen er tilstrækkeligt præcise på vores data til at være brugbare.

Søgenøglen har en precision-score på 23 procent, hvilket betyder, at under en fjerdedel af alle dens klassificeringer af sexisme er korrekte ifølge vores menneskelige annotører. Dette er altså ikke en god score, da 77 procent af de kommentarer, som den mente var sexistiske, slet ikke var det. Endnu værre står det til med misogyni-algoritmen, hvor det kun er 16 procent eller cirka hver sjette kommentar, hvor den klassificerer sexistisk indhold korrekt.

Søgenøglen udpeger næsten alle de kommentarer, som er vurderet af mennesker som sexistisk indhold (94 procent). Dette er en ret god andel, men her er det vigtigt at holde balancen med andelen af korrekte klassificeringer i mente. En algoritme, der vurderede *alle* kommentarer som sexistiske, ville have en recall-score på 100 procent, men en tilsvarende lav precision. Da kun 23 procent af klassificeringerne af sexisme er korrekte, er balancen stadig for dårlig til at kunne bruges som automatisk værktøj. At en simpel søgenøgle ikke fungerer godt, er dog ikke overraskende, da hverdagssexisme er et meget subtilt fænomen, hvor der ikke bare bliver brugt anstødeligt sprog som "kælling" og "luder", men i stedet kan tage form af for eksempel miskreditering eller negligering.

Misogyni-algoritmen, derimod, havde både en dårlig precision *og* en dårlig recall. Den fandt kun 23 procent af de kommentarer, der var blevet kategoriseret af mennesker som sexistisk indhold. Dette er altså betydelig dårligere resultater end, hvad modellen blev præsenteret for i dens forskningsartikel.

Evaluerings-metrik	Søgenøgle	Misogyni-algoritme
Precision	23 %	16 %
Recall	94 %	23 %

Automatiserede løsninger kan ikke erstatte mennesker

At en model fra forskning ikke præsterer lige så godt som i dens tilhørende artikel, er ikke helt atypisk.

En AI-algoritme er nemlig kun så god, som den data, den bliver trænet på. Misogyni-algoritmen blev trænet på 28.000 danske brugeres indlæg fra Facebook, Twitter og Reddit, som blev kategoriseret af menneskelige annotører.

Algoritmen bruger disse eksempler til at lære, hvilke ord og sætninger der ofte findes i indlæg, der hører til hver kategori. Efter at have set mange eksempler bliver algoritmen bedre til at klassificere, om et nyt indlæg indeholder misogynt indhold eller ej.

En af grundene til, at den ikke klarer sig godt på vores Facebook-data, er, fordi den muligvis ikke har set nok eksempler, der ligner de indlæg, den møder på Facebook. Når træningsdataene også kommer fra andre platforme som Twitter og Reddit, kan sproget, tonen og indholdet på Facebook være anderledes, hvilket gør det sværere for algoritmen at lave korrekte klassificeringer.

En anden grund til, at modellen præsterer dårligt, kan være, fordi sproget på nettet ændrer sig over tid. Dette ses blandt andet i Analyse & Tals udvikling af A&ttack-modellerne til at identificere sproglige angreb i Facebook-kommentarer på dansk.

Den første A&ttack-model blev udviklet i 2021 og blev brugt til at analysere den offentlige debat i perioden 2019-2021 for sproglige angreb. Denne model præsterede dog dårligere på Facebook-data fra perioden 2021-2024. Selv om platformen forblev den samme, var sproget og indholdet på Facebook sandsynligvis anderledes i den senere periode (Analyse & Tal, 2024). Dette tyder på, at modellen var for specifik til de mønstre, den havde lært fra de tidligere data, og dermed ikke var fleksibel nok til at håndtere ændringer i sproget og adfærden over tid, selv om platformen var den samme.

Brugen af automatiserede metoder som AI kræver altså løbende opdatering af træningsdata for at sikre, at algoritmen kan følge med ændringer i sprog og adfærd, specielt til arbejde med moderation. Uden regelmæssig opdatering risikerer algoritmen at blive mere upræcis, da den ikke længere afspejler de nyeste mønstre, der følger med samtalerne.

Anbefalinger

Hverdagssexisme er vanskeligt at identificere kvantitativt, da konteksten spiller en stor rolle for forståelsen.

Dette betyder dog ikke, at man skal afskrive automatiserede løsninger i moderation, da AI og søgenøgler stadig kan bruges til at indkredse moderatorens arbejde.

Anvend automatiserede løsninger som støtteværktøj i moderation.
Vi anbefaler at implementere AI-baserede værktøjer som supplement til moderationsarbejdet, da menneskelig moderation kræver mange ressourcer. AI kan derfor være nyttig til at fremhæve sexistiske kommentarer, hvilket hjælper moderatorer med at fokusere deres indsats. Når man bruger disse automatiserede løsninger, er det dog vigtigt at evaluere dem grundigt. Dette sikrer, at man er bevidst om, hvilke former for sexisme algoritmerne fanger - og hvilke de overser.

Brug løbende opdatering af træningsdata for at indfange hverdagssexisme.
Vi anbefaler en systematisk tilgang til opdatering af træningsdata for AI-modeller og søgealgoritmer. For at opretholde effektiviteten af automatiserede moderationsværktøjer er det vigtigt, at de kontinuerligt trænes på aktuelt sprogbrug og nye udtryksformer for sexisme. Uden denne regelmæssige opdatering vil systemernes præcision gradvist forringes, da de ikke vil afspejle aktuelle sproglige mønstre og samtaletemaer. Særligt hverdagssexistiske kommentarer, der ofte kan være subtile og kontekstafhængige, kræver opdaterede modeller for at blive identificeret korrekt.

Litteratur

Alexandra Instituttet (2022). *AI fanger aggressivt sprog for DR*.

Amnesty International (2017). https://amnesty.dk/en-ud-af-fem-danske-kvinder-oplever-chikane-paa-nettet/

Amnesty International (2025) Had skader – 22 års retspraksis på straffelovens § 266 B © Amnesty International Danmark.

Analyse & Tal, Os & Data & Trygfonden (2025). *Angreb og had i den offentlige debat på Facebook* https://www.ogtal.dk/assets/files/Angreb-og-had-i-den-offentlige-debat-paa-Facebook.pdf

Analyse & Tal (2024). *Byg Selv eller GPT-4?* https://www.ogtal.dk/publikationer/byg-selv-eller-gpt-4-2

Analyse & Tal (2021). *Angreb i den offentlige debat på Facebook* https://www.ogtal.dk/publikationer/angreb-i-den-offentlige-debat-paa-facebook

Anzovino, M., Fersini, E., & Rosso, P. (2018). Automatic identification and classification of misogynistic language on twitter. In *Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings 23* (pp. 57-64). Springer International Publishing.

Becker, J. C., & Swim, J. K. (2012). Reducing endorsement of benevolent and modern sexist beliefs. *Social Psychology*.

Buie, H. & Croft, A. (2023) The Social Media Sexist Content (SMSC) Database: A Database of Content and Comments for Research Use https://doi.org/10.1525/collabra.71341

Center for Sociale Medier og Demokrati (2024). *Danskernes holdning til den offentlige samtale på online platforme,* https://kum.dk/fileadmin/_kum/1_Nyheder_og_presse/2024/Danskernes_holdning_til_den_demokratiske_samtale_paa_online_platforme.pdf

Christensen, J. F., Mahler, R., & Teilmann-Lock, S. (2021). GenderLAB: Norm-critical design thinking for gender equality and diversity. Organization, 28(6), 1036–1048.

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. Stanford Law Review, 43(6), 1241–1299.

Dorte Marie Søndergaard (1996). Tegnet på kroppen: Køn: koder og konstruktioner blandt unge voksne i akademia. Museum Tusculanums Forlag.

DR (2023). *Medieudviklingen 2023* https://www.dr.dk/om-dr/fakta-om-dr/medieforskning/medieudviklingen/2023/status-2023-facebook-i-frit-fald

Drakett, J., Rickett, B., Day, K., & Milnes, K. (2018). Old jokes, new media – Online sexism and constructions of gender in Internet memes. Feminism & Psychology, 28(1), 109- 127. https://doi.org/10.1177/0959353517727560

Einersen, A.F. et al (2021). Sexism in Danish Higher Education and Research: Understanding, Exploring, Acting. Draft version, Copenhagen, March 2021

Europarådet 2019: https://rm.coe.int/prems-127324-gbr-2573-rapport-sur-rec-2019-1-web-a5-2755-4175-8475-v-1/1680b2a1f9

Everyday Sexism Project Danmark, Home - Everyday Sexism Project Danmark

Gervais, S. J., & Hillard, A. L. (2014). Confronting sexism as persuasion: Effects of a confrontation's recipient, source, message, and context. *Journal of Social Issues*, *70*(4), 653-667.

Gibson, Anna. Free Speech and Safe Spaces (2019). How Moderation Policies Shape Online Discussion Spaces, Social Media + Society January-March 2019: 1–15

Gillespie, Tarleton (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press,

Groes, Christian (2023). *Da MeToo ramte manden*. Gad.

Hangartner et al. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment https://www.pnas.org/doi/10.1073/pnas.2116310118

Hawkins, I. Roden, J. Attal, M. Aqel, H. (2023). Race and gender intertwined: why intersecting identities matter for perceptions of incivility and content moderation on social media, Journal of Communication, 73(6), Pages 539–551.

Hvenegård-Lassen, K., Staunæs, D., & Lund, R. (2020). Intersectionality, Yes, but How? Approaches and Conceptualizations in Nordic Feminist Research and Activism. *NORA - Nordic Journal of Feminist and Gender Research*, *28*(3), 173–182.

Institut for Menneske Rettigheder. (2024). Ytringsfrihed og Selvcensur. Tilgået d. 20/02-2025. https://menneskeret.dk/files/media/document/Ytringsfrihed-og-selvcensur_DK_juni2024.pdf

Jha, A., & Mamidi, R. (2017). When does a compliment become sexist? analysis and classification of ambivalent sexism using twitter data. In *Proceedings of the second workshop on NLP and computational social science* (pp. 7-16).

Kelly, L. (1989). The Continuum of Sexual Violence. In J. Hanmer & M. Maynard (Editors), Women, violence, and social control (pp. 46- 59). Atlantic Highlands, NJ: Humanities Press International.

Krøjer, J., Muhr, S. L., Plotnikof, M., Myers, E. S., Einarsen, A. F., Macleod, S., Munar, A. M., & Skewes, L. (2024). *Sexisme på arbejde. Genkend, forebyg og håndter*. Djøf Forlag.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cam Journal*, *10*(2), 31-36.

Masullo, G. M. Marc Ziegele, Martin J. Riedl, Pablo Jost & Teresa K. Naab (2022). Effects of A High-Person-Centered Response to Commenters Who Disagree on Readers' Positive Attitudes toward A News Outlet's Facebook Page, Digital Journalism,

Masullo, G. M., Riedl, M. J., & Huang, Q. E. (2022). Engagement moderation: What journalists should say to improve online discussions. *Journalism Practice*, *16*(4), 738-754.

Mogensen, C. & Rand, S. H. (2019). Vrede unge mænd – Viden om ekstreme onlinefællesskaber, Center For Digital pædagogik

Muhr, Sara Louise (2019). *Ledelse af køn: Hvordan kønsstereotyper former kvinders og mænds karrierer*. Djøf Forlag, 2½

Nysten S. (2021): *Nyt værktøj skal stoppe online-chikane af kvinder* https://prosabladet.dk/nyheder/nyhed/nyt-vaerktoej-skal-stoppe-online-chikane-af-kvinder

Reinicke, K. (2018). Mænd der krænker kvinder: Refleksioner i kølvandet på #MeToo (1st ed.). KBH: Samfundslitteratur

Sasse, J. & Grossklags, J (2023). Breaking the Silence: Investigating Which Types of Moderation Reduce Negative Effects of Sexist Social Media Content.

Tougas, F., Brown, R., Beaton, A. M., & St-Pierre, L. (1999). Neosexism among women: The role of personally experienced social mobility attempts. *Personality and Social Psychology Bulletin*, *25*(12), 1487-1497.

Zeinert, P., Inie, N., & Derczynski, L. (2021). Annotating online misogyny. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 3181-3197).

5

Bilag

Bilag 1: GenderLAB som metode

GenderLAB er en metode designet til at understøtte arbejdet mod sexisme i praksis.

GenderLAB kombinerer normkritik og *Design Thinking* i et workshop-format, der via vidensdeling og diverse øvelser tilbyder deltagere et (nyt) udgangspunkt for at forstå og engagere sig i problemstillinger og løsninger vedrørende sexisme.

Ved at kombinere normkritikkens refleksive proces og kulturforandrende potentiale med den handlingsorienterede tilgang fra Design Thinking, kan disse i samspil bidrage til øget bevidsthed og konkret ændret adfærd.

GenderLAB består ud over viden og dialog også af et stærkt handlingselement, en samskabelsesproces, hvor moderatorerne har været involveret i at udvikle konkrete løsningsideer til at imødegå og forebygge digital hverdagssexisme.

Gennem forskellige øvelser i GenderLAB har vi i samarbejde med deltagerne kvalificeret de konkrete idéer til at håndtere og modvirke digital hverdagssexisme, med særligt fokus på løsninger, der udfordrer, forstyrrer og overskrider gængse antagelser og normer.

Bilag 2: Litteraturreview

Anzovino, M., Fersini, E., & Rosso, P. (2018). Automatic identification and classification of misogynistic language on twitter. In Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings 23 (pp. 57-64). Springer International Publishing. Dette oplæg udforsker kategorisering af misogynt sprogbrug online. Forskerne fremhæver, at dette er vigtigt, da meget forskning hidtil har fokuseret på krænkende sprogbrug i relation til etnicitet, religion, kønsidentitet og seksuel orientering. Artiklen dykker dels ned i eksempler på kvindefjendske tweets, og drøfter en mere automatiseret metode.

Becker, J. C., & Swim, J. K. (2012). Reducing endorsement of benevolent and modern sexist beliefs. Social Psychology.
Artiklen understøtter hypotesen om, at såkaldt "velment" sexisme kan reduceres ved at give information om dets skadelige konsekvenser. Forskningen viser desuden, at kvinder og mænd bliver mere bevidste om det fulde omfang af kønsdiskrimination og reducerer deres støtte til sexisme generelt, når de får information om den skadelige natur samt viden om, hvor udbredt hverdagssexisme og "velment" sexisme er.

Bhandari, A. Ozanne, M. Bazarova, N. N. DiFranzo, D. (2021) Do You Care Who Flagged This Post? Effects of Moderator Visibility on Bystander Behavior
Dette studie udforsker, hvilken effekt synlig AI genereret moderation har på bystanderadfærd. Tidligere forskning i bystanderadfærd viser, at bystandere har større tendens til at være passive, hvis der er AI genereret flagging. Studiet viser, at AI genereret flagging af problematisk tone kan have den utilsigtede effekt, at bystandere holder sig tilbage fra selv at flagge problematisk tone.

Buerger C. (2022), "Why They Do It: Counterspeech Theories of Change" in the project Dangerous Speech.
Artiklen peger på, at majoriteten af brugere på online-platforme er passive bystandere. Denne gruppe er vigtig at påvirke i kampen mod hadtale og sexisme på nettet. Interviews med en lang række individer, som benytter sig af modtale viser, at motivationen er at forsøge at påvirke den stille majoritet. Herved benyttes modtale med det formål at aktivere andre i kampen mod sexisme og hadtale – og ikke til at ændre 'gerningspersonens' syn.

Buie, H. & Croft, A. (2023) The Social Media Sexist Content (SMSC) Database: A Database of Content and Comments for Research Use
Studiet peger på vigtigheden i at tage højde for den bias, der kan være ved udførelsen af modtale. Studiet viser, at kvinder i højere grad vurderede en kommentar som sexistisk, end mænd gjorde. Ligeledes havde kvinder en stærkere følelsesmæssig reaktion på kommentarerne end mænd. Herved kan man udlede en kønsmæssig forskel, når det kommer til vurderingen af sexisme.

Celadin, Tatiana & Folco Panizza, Valerio Capraro (2024) Promoting civil discourse on social media using nudges: A tournament of seven interventions, PNAS Nexus, Volume 3, Issue 10, October, pgae380. Denne artikel tester og sammenligner forskellige nudges, der er designet til at reducere cirkulationen af skadeligt indhold såsom hadefuld tale. Resultater tyder på, at nudges, der retter sig mod normer, repræsenterer en lovende tilgang til at fremme en ikke-diskriminerende tone og kan bidrage til at gøre sociale medier til et mere sikkert og inkluderende rum for alle.

Essed, P. & Muhr, S.L. (2018) Entitlement racism and its intersections: An interview with Philomena Essed, social justice scholar, Emhemera, nr. 18(1), 183-201. Forskningen peger på hvordan race- og kønsbestemte magthierarkier konstrueres og forstærkes gennem normalisering af hverdagspraksis såsom vittigheder, historiefortælling, generaliseringer eller endda såkaldte komplimenter. Forskningsinterviewet uddyber, hvordan forestillingen om, at vi allerede har opnået ligestilling, er en barriere for at adressere hverdagssexisme og -racisme.

Garland, J., Ghazi-Zahedi, K., Young, JG. et al. (2023) Impact and dynamics of hate and counter speech online. EPJ Data Sci. 12, 27. Forskerne peger på, at brugergenereret modtale kan være en lovende måde at bekæmpe hadefuld tale og fremme fredelig, ikke-polariseret diskurs. Artiklen peger på, at modtale – især når den er organiseret – kan hjælpe med at bremse hadefuld retorik på digitale platforme. Samtidig peger forskerne på, at der stadig mangler længerevarende studier, der kan påvise, hvordan den brugergenererede modtale virker i praksis.

Gervais, S. J., & Hillard, A. L. (2014). Confronting sexism as persuasion: Effects of a confrontation's recipient, source, message, and context. Journal of Social Issues, 70(4), 653-667. Studiet peger på, at køn har betydning for vurderingen af lederes konfrontation (f.eks modtale) ift sexisme. Som antaget vurderede de kvindelige deltagere generelt en konfrontation mere positivt (end de mandlige deltagere). Samtidig blev kvindelige (versus mandlige) ledere evalueret mindre positivt, når de konfronterede sexisme offentligt.

Gibson, Anna. (2019) Free Speech and Safe Spaces. How Moderation Policies Shape Online Discussion Spaces, Social Media + Society January-March 2019: 1–15.
Gibson undersøger effekten af moderationspolitikker for tonen i den digitale samtale og peger på, at disse er helt afgørende ift. at sikre en demokratisk diskussion. Studiet tager afsæt i Redditsider og undersider, og peger på, at de moderatorer, som havde en klar moderationspolitik og forsøgte at skabe trygge rum, fjernede betydeligt flere kommentarer, og at brugerne også slettede deres egne kommentarer betydeligt oftere. Gibson konkluderer også, at tonen i de trygge rum var langt mere positivt, hvorimod sproget ellers var relativt mere negativt og vredt. Disse sproglige forskelle fortsatte i kommentarer fra brugere, der samtidig deltog i de forskellige fora.

Gillespie, Tarleton (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press. Gillespie beskriver sociale mediers praksis og forklarer de underliggende begrundelser for, hvordan, hvornår og hvorfor de forskellige politikker, herunder moderationspolitikker håndhæves. I bogen fremhæver Gillespie, at indholdsmoderation generelt får for lidt opmærksomhed og prioritering, selvom dette er helt centralt fordi det former de sociale (digitale) normer. Bogen er bl.a. baseret på interviews med indholdsmoderatorer.

Gudbjartur Ingi Sigurbergsson and Leon Derczynski. (2020). Offensive Language and Hate Speech Detection for Danish. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 3498–3508, Marseille, France. European Language Resources Association. Forskerne tager afsæt i det meget engelsksprogede fokus på hadefuld tale og konstruerer et dansk datasæt med forskelige typer af stødende sprogbrug. Med afsæt i udviklingen af fire automatiske klassifikationssystemer, præsenterer forskerne en engelsk og dansk tilgang til at indfange typen og måle brugen af stødende sprog.

Hangartner et al (2021), "Empathy-based counterspeech can reduce racist hate speech in a social media field experiment".
Proceedings of the National Academy of Sciences Vol. 118 | No. 50 December 14, 2021. Dette studie fra 2021 undersøger kommentarer fra 1.350 twitterbrugere og tester nedbringelsen af racistisk hadtale ud fra tre modtale-strategier: Empati, advarsel om konsekvenser og humor. Dertil blev disse strategier holdt oppe mod en kontrolgruppe. Resultaterne viser, at empati har den bedste virkning i forhold til at mindske forekomsten af racistisk hadtale, samt øge afsenderens sletning af sin egen racistiske kommentar.

Hawkins, I. Roden, J. Attal, M. Aqel, H. (2023) Race and gender intertwined: why intersecting identities matter for perceptions of incivility and content moderation on social media, Journal of Communication, 73(6), Pages 539–551. Studiet tager udgangspunkt i USA, og viser, at ens race eller køn kan være en bias i forhold vurderingen af hvornår en kommentar er over stregen og tendensen til at anmelde en kommentar eller anvende modtale.

Jha A. & Mamidi R. (2017) "When does a Compliment become Sexist?" Analysis and Classification of Ambivalent Sexism using Twitter Data. Proceedings of the Second Workshop on NLP and Computational Social Science, pages 7–16, Vancouver, Canada. Association for Computational Linguistics. Undersøgelsen har fokus på "ambivalent" sexisme, som forskerne finder er udbredt på sociale medier. Ambivalent sexisme bliver delt i to kategorier: Fjendtlig (Hostile) sexisme kommer fx til udtryk som vrede, mens velmenende (benevolent) sexisme ofte kommer fx til udtryk i form af et kompliment. Undersøgelsen viser, at det er vanskeligt rent teknisk at indfange ambivalent sexisme og særligt velmenende sexisme.

Megarry, J.: (2014) Online incivility or sexual harassment? Conceptualising women's experiences in the digital age. In: Women's Studies International Forum, vol. 47, pp. 46–55. Pergamon. I denne artikel argumenterer Megarry for, at den aggressive chikane af kvinder online, som formidlet bør begrebsliggøres som online seksuel chikane og en form for udelukkelse af kvinders stemmer fra den digitale offentlige sfære. Megarry påpeger, at men har fokuseret for ensidigt på hvor fx twitter styrker det digital demokrati, men har ignoreret nøglespørgsmål vedrørende eksklusion, især kvinders muligheder for at deltage på lige fod med mænd.

Obermaier, M., Schmid, U. K., & Rieger, D. (2023). Too civil to care? How online hate speech against different social groups affects bystander intervention. European Journal of Criminology, 20(3), 817-833. Forskerne sætter fokus på passiv bystander adfærd og påpeger, at denne passivitet fra majoriteten opfattes som en accept og enighed i tonen/den hadefulde tale af den som bliver ramt. Derfor er det vigtigt at skabe aktive allierede. Studiet viser, at hadefulde ytringer opfattes forskelligt afhængigt af den målrettede sociale gruppe ift. køn race, seksuel orientering og at tilskyndelsen til at gribe ind og være en aktiv allieret også varierer efter hvor alvorligt man opfatter det.

Ortiz, S. M. (2024) "If Something Ever Happened, I'd Have No One to Tell:" how online sexism perpetuates young women's silence, FEMINIST MEDIA STUDIES, 24(1). Sammenspillet mellem den fysiske og den digitale verden er større, end man tror, når det kommer til adfærd. Med afsæt i interviews viser Ortiz, hvordan sexisme online kan få betydning offline og omvendt. Studiet viser også, at det kan lede til tavshed (at man trækker sig) hvis man oplever, en sexistisk retorik, idet det får negativ betydning for deres lyst til at sige fra over for sexisme og/eller kommentere digtalt.

Reinicke, K. (2018) Mænd der krænker kvinder: Refleksioner i kølvandet på #MeToo, Samfundslitteratur. Med udgangspunkt i interviews med mænd, giver Reinicke et dybdegående indblik i, hvorfor sexchikane forekommer, og hvordan vi som samfund kan bekæmpe fænomenet. For at komme sexisme og seksuel chikane til livs understreger Reinicke vigtigheden af at have fokus på den struktur, som muliggør seksuel chikane. Dette betyder, at man ikke skal kigge til bestemte individer for at løse problemerne, men i stedet på kulturer og strukturer.

Sasse, J. & Grossklags, J (2023) Breaking the Silence: Investigating Which Types of Moderation Reduce Negative Effects of Sexist Social Media Content. Proceedings of the ACM on Human-Computer Interaction Volume 7, Issue CSCW2
CSCW October 2023. Forskerne udforsker effekten af forskellige former for moderation ift. at skabe et trygt og sikkert digitalt rum for særligt kvinder. Studiet bidrager til en voksende litteratur om digitale normer og kulturer og de praktiske implikationer der er for moderatorer ift. At udvælge atrategier, der kan være effektive og accepterede. Forskerne kigger nærmere på effekten af synlighed af sexistisk indhold og af modtale ift. sociale normer, følelse af sikkerhed og hensigt om at deltage, såvel som retfærdighed.

Vidgen B, Derczynski Leon (2020) Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLoS ONE 15(12). Gennem et systematisk review afdækkes de videnshuller der eksisterer indenfor detektionen af digital hadefuld tale. Undersøgelsens formål spænder blandt andet over at formidle en dyb og kritisk analyse af de eksisterende trænings datasæt for detektionen. Derudover er formålet at få et overblik over manglen af datasæt på området, introducere websiden hatespeechdata.com og til sidst identificere den bedste praksis for at skabe datasæt til detektionen af online hadefuld tale.

Wang, W. & Ngai, J. (2023) "You look like my 14-year-old daughter" A corpus-based study of sexist language in everyday sexism Twitter stories, Source: Journal of Language Aggression and Conflict. Available online: 22 December. Dette studie kategoriserer 1.118 Twitter-opslag med det formål at finde ud af, hvilke typer af sexistisk sprog som kvinder er udsat for. Undersøgelsen viser, at størstedelen af opslagene fra Twitter er indirekte sexistiske, hvilket gør dem sværere at identificere og sige fra overfor. Forskerne fremhæver, at sarkasme, ironi eller humor ofte benyttes til at gøre sexismen indirekte (benyttes sammen med: "det er bare en joke" eller "det er bare for sjov").

Windisch, Steven &, Susann Wiedlitzka, Ajima Olaghere, Elizabeth Jenaway (2022) Online interventions for reducing hate speech and cyberhate: A systematic review. Campbell Systematic ReviewsVolume 18, Issue. Windisch m.fl. peger på nødvendigheden af at bekæmpe digital hadefuld tale, da de sociale platforme øger spredningen og skaber en følelse af normalitet omkring had og potentialet for voldshandlinger og/eller politisk radikalisering. Forskerne savner interventioner, der i højere grad søger at bekæmpe digital hadefuld tale. Forskerne peger også på, at det stadig er vanskeligt at vurdere effekterne af interventioner mod hadefuld tale online, grundet de givne mangelfulde metoder.

Zeinert, P., Inie, N., & Derczynski, L. (2021). Annotating online misogyny. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 3181-3197). Artiklen viser nogle af de udfordringer, der er med automatisk detektion af online misogyni og krænkende sprogbrug. Forskerne peger på det komplekse i at både dataindsamling, dataannotation og biasreduktion, da denne type data er sprogligt kontekstuelt og forskelligartet. Resultaterne viser bl.a. den omfattende taksonomi af betegnelser til annotering af kvindehad der er nødvendig for at indfange online misogyni**.**

Yu, Z., Otto, L., Assenmacher, D., & Wagner, C. (2024). A systematic review of the effects of ai-assisted moderation on individuals and groups. Human-Machine Communication, 9, 167–188. Denne oversigtssartikel giver en status på udviklingen af AI-assisteret indholdsmoderation og hvordan det påvirker individer versus grupper. Forskerne peger på, at der stadig er huller i udforskningen af forskellige niveauer af AI-assisterede moderationsinterventioner og uoverensstemmelser mellem forskellige konceptualiseringer, der gør det vanskeligt at sammenligne forskningsresultater. Forskerne anbefaler at undersøge mere konstruktive moderationsteknikker med vægt på AI's potentiale.

Bilag 3: Kvalitativ undersøgelse af moderatorers praksis

Effektmåling er en central del af projektet. Vi har måttet tilrette projektet undervejs på baggrund af vores erfaringer omkring, hvor svært det blandt andet er at bruge et måleværktøj, der kvantitativt skal indfange og identificere digital hverdagssexisme. Vi har derfor lagt stor vægt på den kvalitative del af effektmålingen grundet søgenøglens og algoritmens begrænsninger. Det har betydet, at vi på mange måder måler effekten ved dybdegående og mangefacetterede interviews med moderatorer og content managers i de to mediehuse, som vi har samarbejdet med omkring udvikling af redskaber til håndtering af og indsats over for digital hverdagssexisme. I den forbindelse er en kvalitativ evaluering og et forsøg på at nærme sig moderatorernes praksis særligt interessant for projektet.

Hvad kan interviewet?

Vi har fået indsigt i moderatorernes arbejdspraksis gennem uformelle samtaler og mere strukturerede individuelle interviews. Både samtalen og interviewet i en fortolkningsproces kan, ifølge Rubow, bidrage til en analyse ved at fremhæve iboende konflikter og tavse overenskomster i hverdagens praksis (Rubow 2010: 243). Ved at interviewe moderatorer, content managers og ledere har vi fået mulighed for at forstå, hvilke problematikker de hver især sidder med, og hvordan nye værktøjer og anbefalinger bedst hjælper dem. Derved kan interviewene belyse deres personlige og professionelle refleksioner og oplevelser i det online felt og kvalificere vores resultater. Interviews af moderatorer er desuden vigtige for at give indblik i den givne kontekst, da det er vigtigt at forstå de reelle dilemmaer, de forskellige moderatorer står i.

Intervention og kvalitativ effektmåling:

Vi har indtil videre fået en hel del større kendskab til, hvor vores viden er begrænset, når det kommer til moderation i praksis. Som en del af vores workshop med de to mediehuses moderatorer og i de efterfølgende interviews finder vi, at det er afgørende for moderationen af digital hverdagssexisme, at moderatorerne får viden, et fælles sprog og en fælles forståelse. Viden om hverdagssexisme giver en bevidsthed, således at moderatorerne får øje på det i kommentarsporerne og det generelle indhold, og derefter kan agere på det. Samtidig optræder der en del dilemmaer i moderationspraksissen – blandt andet hensynet til indholdet, debatten, personerne involveret, de passive brugere og en mere generel forståelse af ytringsfrihed og 'public service'. Så selv om moderatorerne får en viden om sexisme og konsekvenserne af sexisme, er det vanskeligt at moderere i praksis. Det er derfor vigtigt at følge praksis tæt for at forstå, hvorfor og hvornår det er svært at omsætte viden til praksis.

Bilag 4: Annotationskodebog

Kode	Hverdagssexisme
Kort definition	Sexisme som man oplever på hverdagsbasis, der forekommer hyppigt og/eller er normaliseret.
Længere definition	Bevidste eller ubevidste mikroaggressioner, som rettes mod minoritetskøn eller kvinder. Kan umiddelbart virke som harmløse jokes eller kommentarer, men samlet har det en effekt, som begrænser især kvinders deltagelse online.
Hvornår skal man bruge den	Hvis det er mere subtilt.
Hvornår skal man ikke bruge den	Ikke hvis der er direkte trusler.
Eksempel	"Kvinder og teknik ..."

Kode	Neosexisme
Kort definition	Myten om at vi har ligestilling.
Længere definition	Benægtelse af ligestillingsproblemer ved at argumentere, at vi har ligestilling.
Hvornår skal man bruge den	Hvis "vi har ligestilling" (eller lign.) bliver brugt til at lukke diskussion eller personer ned.
Hvornår skal man ikke bruge den	Hvis eksempler på ligestilling bruges.
Eksempel	"Hvorfor snakker vi igen om kønskvoter, når vi allerede har ligestilling i Danmark".

Kode	Miskreditering
Kort definition	Underkendelsen på baggrund af køn, ofte koblet til deres kompetencer.
Længere definition	Personer eller grupper bliver miskrediteret/underkendt ved, at andre nedgør, underkender eller udelukker dem med henvisning til deres køn, og med en indirekte miskreditering fordi de (kvinder eller mænd) falder uden for de normative idéer om maskulinitet og femininitet. Kommentarerne er irrelevante i forhold til konteksten.
Hvornår skal man bruge den	Når formålet udelukkende er at nedgøre.
Hvornår skal man ikke bruge den	Hvis nedgørelsen sker i forbindelse med holdningsytringer i relation til emnet i tråden.
Eksempel	"Man kan ikke overlade vigtige beslutninger til kvinder, for de er styret af deres hormoner", "Man skal ikke lytte til ham, for han er en tøffelhelt", "Man kan ikke overlade beslutningerne til en, som ikke engang kan finde ud, af om han/hun skal være mand eller dame".

Kode	Stereotyper
Kort definition	Fastholdelse og reproduktion af forsimplede og begrænsende kønnede samfundsnormer.
Længere definition	Reproduktionen af firkantede, fastlåste og begrænsende normer og stereotype opfattelser om køn og kønsidentitet, gennem udtalelser og jokes.
Hvornår skal man bruge den	Når der er tale om en kønnet generalisering af gøren, karakteristika og andet.
Hvornår skal man ikke bruge den	Når der er tale om enkeltpersoners handlen, karakteristika og lignende, som ikke er i relation til køn.
Eksempel	"Det eneste kvinder dur til, er at lave mad", "Det er ikke noget, han kan finde ud af, med de løse håndled".

Kode	Dominans
Kort definition	Kvinder er mindre værd end mænd. Nogle former for maskulinitet er mere værd end andre (hegemonisk maskulinitet).
Længere definition	Mænd er på alle punkter overlegne i forhold til kvinder, hvilket kommer til udtryk i udtalelsen. Hvis du er en rigtig mand, så skal du være dominerende. Der sker derved en reproduktion af maskulinitetsnormer.
Hvornår skal man bruge den	Udtalelser hvor kvinder nedgøres til at understrege´, hvor overlegne mænd er, men som også henviser til en bestemt type maskulinitet, netop den dominerende.
Hvornår skal man ikke bruge den	Tidspunkter hvor der er tale som specifikke individers personlige kvalifikationer og ikke generaliseringer.
Eksempel	"Overlad det til mændene, kvinderne har ikke intelligensen", "Det kan han da ikke holde til, han er jo bare en splejs", "Den slags arbejde kræver hår på brystet".

Kode	Godhjertet sexisme
Kort definition	Kvinders værdi måles i forhold til mænd.
Længere definition	Man roser en kvinde ved at give/vurdere hendes værdi ud fra maskuline traditionelle kvaliteter. På den måde understreger man også, hvad en "rigtig" mand er. dette kan også ske i forhold til andre mænd (hegemonisk maskulinitet).
Hvornår skal man bruge den	Når der er en intention om ros, men at denne ros udspringer med en maskulin baggrund.
Hvornår skal man ikke bruge den	Når ros og vurderinger ikke er kønnet.
Eksempel	"Hvor er du sej, du har godt nok nosser (sagt om en kvinde)", "Du skal nok lære det/forstå det, når du får hår på brystet", "hvem havde troet, at en mand kunne se så godt ud" (om en transkønnet mand)

Kode	Seksuel chikane og objektivisering
Kort definition	Udtalelser med seksuelle undertoner eller eksplicitte opfordringer eller henvisninger til seksuel aktivitet, hvor modtageren bliver objektgjort.
Længere definition	Fokus flyttes til det seksuelle ved en person i stedet for tema for drøftelser med den virkning, at personen bliver påmindet om at være et objekt for en andens begær - med det formål eller den virkning at krænke, nedgøre eller udelukke den anden fra samtalen.
Hvornår skal man bruge den	Når der er seksuelle udtalelser, tilnærmelser eller efterspørgsler.
Hvornår skal man ikke bruge den	Når tilnærmelsen er mere indirekte eller skjult.
Eksempel	"Hvor mange penge skal du have for at være min en hel eftermiddag?"

Bilag 5: Søgenøgle

Basis ord	x	Hverdagssexisme					
		Neosexisme	Miskreditering	Stereotyper	Dominans	Godhjertet sexisme	Seksualisering og objektivisering
kvinde		_Vi har ligestilling_	_Kælling	_Husmor_	_Flæber_	_Nosser	_fisse_
mand		_myte_	_So_	_Pige	_Tøs	_For din egen skyld_	_luder
pige		_barsel_	_fed_	_Dreng	_Lille pige_	_Du bør_	_pik_
dreng		_ligeløn_	_Tyk_	_Kvinde	_Dum_	_Du burde_	_stram_
fyr		_equal	_Grim_	_Tilbage til køkkenet_	_Tuder_	_Klogeåge_	_porno_
tøs		_ligestillet_	_Ulækker_	_hysterisk	_tudekiks	_Prinsesse	_pinup_
mænd		_patriark	_premenestruel	_nærtagende_	_tøffelhelt	_Gudinde	_møgluder_
mor		_matriark	_pms_	_snerpet_	_dramaqueen_	_Dronning	_tøjte_
far		ligestilling	_bitch_	_liderbuks_	_Stræber	_Porcelænsdukke_	_dildotøs
trans			_Møgsæk	_mommys boy_	_slap af_		_cunt_
hun			_skinny_	_mors dreng	_intelligens_		_skøge_
han			_Blondine	_feminist	_drama queen_		_slut_
dem			_hedetur	_feminazi_	_lille_		_skeder_
			_smatso	_rejsekælling_	_for ung		_golddigger_
			psykopat	_scorekarl			_femi-kusser_
			_Ko_	_hys			_rundetrunte_
			_tudemarie	_fars pige			_sædcontainer_
			_tudeprincesse	_skrap_			_lolitadukke
			_møgkælling	_mandsling_			_suttetøs
			skør	_svag			_fars lille_
			klam	_blød			_dåse_
			_skrup	_trans_			_piksutter
			_brok	_ladyboy_			_narrepussy
			_hormonel	_køn_			_fuckboy
			_fjæs	_kønsidentitet_			_flaskepige
			_punjabpige	_cis			_friske fisser
			_kæleasiat	_transmand_			_trussetyv
			hund	_transmænd_			_pudderdåse
			perker	_transperson			_sugardaddy
			shabs	_transkvinde			_spiller_
			_albino	tørklæde			_pikslikker
			_hættemåge				_aseksuel_
			_eskimo				_Sugarbabe
			_Perserprins				_Liderbug
			_Tørklædepige				_Fjams_
			_perkersvin				_tør_
			Abe				_liderlig
			_heil Hitler				_Kuvøseguf
			_Rotte				_MILF_
			_Jødenæse				_DILF_
			_terrorist				_Cougar_
			_immigrant				_homo_
			_polsk sædcontainer				_bøsserøv
			_negerslave				_bøsse
			_nazi				_lebbe
			_kineser				_lesbisk
			eksotisk skønhed				_trækkerdreng
			_polski				_svans
			white-boy				_skabsbøsse
			trash				_analrytter
			gringo				_homosvin
			mula				_faggot_
			_Grønlænder				_pædo
			_Gringo				_betonlebbe
			Gypsy				_sexkilling
			_Risgnasker				_springe ud_
			Nigga				
			_sort _				
			sorteper				
			_Sigøjner				
			_Kinderæg				
			Slave				
			_Olding				