



Mediemyndigheten

# Vad säger forskningen om AI och desinformation på sociala medier?

Mediemyndighetens  
reflektion till rapport beställd från  
Totalförsvarets forskningsinstitut om  
AI, desinformation och sociala medier

# Sammanfattning

Generativ AI förändrar snabbt hur information skapas, sprids och uppfattas – inte minst på sociala medier, där sann, missvisande och fabricerad information blandas i ett alltmer svårnavigerat flöde. För att bättre förstå denna utveckling har Mediemyndigheten låtit Totalförsvarets forskningsinstitut (FOI) kartlägga kunskap från internationell forskning och relevant grålitteratur om AI-genererat innehåll på sociala medier och dess konsekvenser för medie- och informationslandskapet.

Studien visar att forskningsfältet är i ett tidigt skede, med stora kunskapsluckor men också tydliga mönster: generativ AI används brett, från journalistik och kreativt skapande till bedrägerier, desinformation och nya former av social interaktion. En central insikt från översikten är att fältet har betydande blinda fläckar, till exempel avseende hur utvecklingen påverkar barn och unga, kunskap om flera av de största plattformarna samt om samhällskonsekvenser. Mediemyndigheten ser därför ett behov av mer aktuell, interdisciplinär och praktisk kunskap när det gäller generativ AI och desinformation på sociala medier.

## **Mediemyndigheten** **Avdelningen för utredning och analys**

Martin Karlsson, analytiker  
Jacob Strandell, utredare  
Marie Swanström, omvärldsanalytiker

# Inledning

De senaste årens utveckling av generativ AI (artificiell intelligens) har fått ett stort genomslag i medie- och informationslandskapet. Tekniken är redan väl integrerad i exempelvis sökmotorer, rekommendationstjänster och smarta assistenter i mobiltelefoner. När användarvänliga modeller och agenter för generativ AI lanserades för allmänheten gick utvecklingen in i en ny fas. Det gav fler möjlighet att snabbt skapa och manipulera stora mängder innehåll, inklusive missledande innehåll och avsiktlig desinformation.

I dagens digitala medielandskap blandas därför sann, vilseledande och falsk information, vilket gör det svårt att avgöra vad som är korrekt och vem som står bakom innehållet. Utvecklingen av generativ AI försvårar detta ytterligare eftersom tekniken kan användas för att skapa desinformation, utföra bedrägerier och producera annat skadligt material. Det kan dessutom göras i stor skala genom automatisering.

## Mediemyndighetens uppdrag

Mediemyndigheten ska inom sitt verksamhetsområde verka för yttrandefrihet, medie- och informationskunnighet samt möjligheterna till mediemångfald och tillgänglighet. Myndigheten ska också verka för att stärka barn och unga som medvetna medieanvändare, ta tillvara deras erfarenheter samt verka för att skydda barn från skadlig mediepåverkan. Myndigheten ska även följa och analysera utvecklingen inom medieområdet och om barns och ungas medieanvändning. Myndigheten har också i uppgift att publicera information om bland annat teknik inom medieområdet.

En strategi i myndighetens färdplan 2025 - 2027 handlar om att nyttiggöra vår kunskap inom medieområdet. Detta gör myndigheten bland annat genom att ta fram analyser och framtidsorienterade kunskapsunderlag. Givet den snabba utvecklingen av generativ AI i medie- och informationslandskapet växer också behovet av aktuell och uppdaterad kunskap på området.

Mediemyndigheten genomför en rad insatser på området:

- Utvecklat särskilda insatser för medie- och informationskunnighet (MIK), som exempelvis webbutbildningen mediemedveten.se med en särskild AI-modul.
- Lett genomförandet av Nordic Media Literacy Survey, på uppdrag av Nordiska ministerrådet, som omfattar användning och attityder till AI i de nordiska länderna.
- Utvecklat ett pedagogiskt material om Bilders makt – AI och visuell påverkan för att användas i undervisning.
- Deltar i en arbetsgrupp inom European Platform of Regulatory Authorities (EPRA) där regleringsmyndigheter inom medieområdet utbyter erfarenheter om AI.
- Mäter barns, ungas och föräldrars användning av och attityder till generativ AI i de återkommande rapporterna Ungar & Medier, Småungar & Medier samt Föräldrar & Medier.
- Utvecklar en ny undersökning om ungas sociala relationer till AI.

## Studiens syfte

För att ytterligare stärka arbetet med att förstå hur generativ AI påverkar medielandskapet bedömde Mediemyndigheten att det fanns ett behov av en litteraturoversikt för att komplettera nuvarande kunskap med akademisk forskning. Myndigheten gav därför Totalförsvarets forskningsinstitut (FOI) i uppdrag att kartlägga och sammanställa aktuell forskning och relevant grålitteratur<sup>1</sup> inom området.

Studiens syfte var att sammanställa:

1. Hur AI-genererat innehåll används och gestaltas på sociala medieplattformar, inklusive i desinformationssammanhang.
2. Vilka tekniska, juridiska och samhällsvetenskapliga metoder som utvecklats för att identifiera, verifiera och hantera sådant innehåll.
3. Vilka rättsliga och etiska implikationer som diskuteras i relation till plattformsansvar, barns rättigheter och journalistisk trovärdighet.

---

<sup>1</sup> Grålitteratur avser publikationer som bidrar med kunskap men inte har genomgått akademisk peer-review, exempelvis rapporter, arbetsdokument, policydokument, konferensbidrag, utredningar och akademiska avhandlingar.

# Studiens resultat

Av översikten framgår att forskningsläget avseende AI-genererat innehåll på sociala medier fortfarande befinner sig i ett tidigt skede med jämförelsevis stora empiriska och metodologiska luckor. Forskningen domineras av amerikanska och brittiska studier som i huvudsak studerat fenomenet på ett fåtal sociala medieplattformar, i synnerhet plattformen Twitter/X, med begränsad räckvidd inom såväl den europeiska som amerikanska befolkningen. FOIs översikt konstaterar också en brist på tvärvetenskaplig forskning som belyser fenomenet från olika perspektiv, inte minst samhällsvetenskapligt.

Bortsett från dessa luckor identifierar översikten flera centrala teman i den akademiska forskningen om AI-genererat innehåll på sociala medier som bidrar till en djupare förståelse av fenomenet i dagens digitala medielandskap.

## En stor bredd av användningsområden

Sammantaget visar de studier som ingår i översikten att generativ AI används i framställningen av både text, bild och ljud på sociala medier och tycks förekomma i flera och sinsemellan mycket olika typer av innehållsliga kategorier på sociala medier, exempelvis:

- Desinformation
- Pornografi/våld
- Extremism
- Dating
- Nätmobbning (Cybermobbing)
- Bedrägeri/utpressning
- Medicin
- Journalistik
- Konst
- Marknadsföring
- Satir/humor/meme
- (Para)sociala interaktioner
- Sociala frågor/utbildning

Vid sidan om framställningen av olika typer av medieinnehåll, används generativ AI även för att skapa och upprätthålla användarkonton och profiler på sociala medier. Dessa har visat sig användas dels i manipulativa syften, exempelvis i form av “catfishing” och “deep phishing”, där andra användare utsätts för bedrägerier. Dels i öppna och icke-manipulativa syften, exempelvis AI-kompanjoner: tjänster som ger användare möjlighet att inleda parasociala relationer med AI-genererade personligheter.

## **Nya och gamla risker och utmaningar**

Ett annat forskningsområde som ringas in i översikten är studier som har analyserat och problematiserat juridiska perspektiv på AI-genererat innehåll på sociala medier. Till skillnad från utbredningen av sociala medier och användargenererat innehåll i sig, har generativ AI åtföljts av en förhållandevis snabb lagstiftningsprocess på flera håll i världen, bland annat inom EU, Storbritannien och USA. Fortfarande återstår dock flera problem och utmaningar på det juridiska området. Det gäller dels hur upphovsrätt och människors personliga integritet ska skyddas när exempelvis nyhetsartiklar innehållande personuppgifter kopieras och/eller sprids i syfte att träna AI-modeller. Dels hur AI-genererat innehåll ska och kan regleras på ett tydligt sätt när användningsområdena är så mångfacetterade i både karaktär och risknivå. Till den senare utmaningen kan också räknas frågan i vilken utsträckning AI-genererat innehåll bör särregleras eller kan regleras utifrån befintlig medielagstiftning.

I ett fåtal studier uppmärksammas risker gentemot barn och unga. Riskerna inbegriper bland annat när AI-genererat innehåll använts i fall av nätmobbning, när barns och ungas personuppgifter direkt eller indirekt används för att träna AI-modeller samt när barn och unga har svårt att identifiera AI-genererat innehåll på sociala medier. Även om barn och unga kan tänkas vara särskilt sårbara och utsatta i förhållande till dessa risker är de dock knappast att betrakta som unika för åldersgruppen. Som framgår av andra studier i översikten är svårigheterna att identifiera AI-genererat innehåll snarare ett problem för alla människor.

Problemet med att identifiera AI-genererat innehåll har i sin tur medfört flera andra relaterade problem som urskiljs i översikten. Det handlar dels om utmaningen i att skilja mellan falska och sanna nyheter på sociala medier. Dels i att skilja mellan desinformation och missinformation. Det vill säga om ett falskt inlägg har framställts med eller utan en faktisk avsikt att vilseleda andra användare av sociala medier. Detektering av AI-genererat innehåll på sociala medier har heller inte förenklats av de näst intill obegränsade möjligheterna att framställa och sprida stora volymer av AI-genererat

innehåll. De här möjligheterna har också föranlett debatt vad gäller hur generativ AI kan bidra till att effektivisera och förenkla det journalistiska arbetet utan att samtidigt urholka journalistikens trovärdighet.

Uppdraget utformades på ett sätt som innebär att rapporten framför allt lyfter risker med generativ AI på sociala medier. Trots dessa risker framgår också av översikten att AI-genererat innehåll på sociala medier kan användas i andra och i viss mån mindre riskfyllda sammanhang, exempelvis i form av humor och parasociala relationer. Vilket i sin tur pekar på att fenomenet inte enbart kan eller bör avfärdas som en risk i dagens digitala medielandskap.

# Diskussion

Givet de begränsningar som rapporten identifierar inom forskningsfältet är det svårt att identifiera några tydliga empiriskt baserade konsekvenser av AI-genererat innehåll i sociala medier. Däremot är det tydligt att fenomenet aktualiserar och förstärker flera gamla och nya risker i dagens digitala medielandskap, exempelvis desinformation och kränkningar av människors personliga integritet. AI-genererat innehåll på sociala medier väcker också frågor om journalistikens framtid och dess förutsättningar i det samtida medielandskapet. Dessutom är forskningens nuvarande tillkortakommande i sig en viktig insikt som samhällsaktörer behöver förhålla sig till.

## Begränsningar i forskningen och kvarstående kunskapsluckor

Sammantaget kretsar de stora insikterna från forskningsöversikten till stor del om de kunskapsluckor som identifieras. En övergripande slutsats är att den akademiska forskningen ofta saknar en önskvärd bredd och heterogenitet avseende forskningens metodologiska och disciplinära perspektiv, geografiska ursprung, populationer samt plattformar som omfattas. Kanske är därför också ett av forskningsöversiktens viktigaste bidrag de förslag på framtida studier som presenteras. Dels vad gäller mer forskning om AI-genererat innehåll på de populäraste plattformarna. Dels vad gäller fler nordiska och svenska studier och studier om hur barn påverkas. Det finns även behov av mer snabbriktig forskning avseende aktuella utvecklingar som exempelvis ny reglering och uppdaterade detektionssystem.

En central kunskapslucka i tidigare forskning är bristen på tvärvetenskapliga perspektiv där samhällsvetenskap ingår, särskilt eftersom generativ AI på sociala medier kan få stora samhällseffekter. Tvärvetenskapliga studier som inkluderar samhällsvetenskapliga perspektiv är avgörande för att förstå hur generativ AI påverkar samhället, såväl individer som institutioner.

## Grålitteratur blir viktigt för snabbriktiga områden

En outtalad men central slutsats från rapporten är att högkvalitativ akademisk forskning kan ha svårt att hänga med i utvecklingen av snabbriktiga medietekniska områden. Vi ser flera tecken på detta i de kunskapsluckor som finns i materialet, särskilt bristen på bredd i studerade plattformar, studier av barn och unga, och integrationen av samhällsvetenskapliga perspektiv.

Forskningens luckor är sannolikt en kombinerad effekt av a) det specifika områdets metodologiska utmaningar och b) den akademiska forskningens generella arbetssätt. Avseende specifika utmaningar kan nämnas den snabba expansionen av just generativ AI, samt dess komplexitet och bredd, avseende integration i och påverkan på medielandskapet. Samtidigt rör sig akademisk forskning generellt mycket långsamt, och är bunden till finansierings- och publikationsprocesser samt metodologisk strikthet. Kombinationen gör det alltså sannolikt svårt för den akademiska forskningen att kontinuerligt följa hur generativ AI utvecklas och påverkar samhället i realtid, med önskvärd bredd och aktualitet. Samtidigt som den akademiska forskningen kartlägger hur generativ AI påverkar samhället så behövs kunskap för beslutsfattande här och nu.

En sammanfattande tolkning av forskningsöversikten är därför att grålitteratur nu blir viktigare än någonsin, särskilt inom snabbbrörliga medieområden med stor påverkan på samhället. Grålitteraturen fyller i bästa fall gapet mellan den praktiska verkligheten och den ackumulerade vetenskapliga kunskapen och kan ha bättre förutsättningar att bidra med snabbbrörligare insikter, i en svensk kontext, omfattande viktiga områden som barn och unga, och aktuella plattformar. Det innebär att myndigheter med uppdrag att följa utvecklingen sannolikt blir allt viktigare som kunskapskällor för att följa aktuella samhällsförändringar och komplettera den akademiska forskningen med kunskap av större bredd och omedelbar samhällsrelevans. Detta medför dock samtidigt höga krav på grålitteraturens flexibilitet, relevans och kvalitet.



**Mediemyndigheten**

Tel: 08-580 070 00

[www.mediemyndigheten.se](http://www.mediemyndigheten.se)