



AI-transparens i journalistiken

2024

Projektgrupp bakom arbetet

Agnes Stenbom (projektledare), Schibsted

Annie Lidesjö, NTM

Calle Sandstedt, Omni

Camilla Hillerö, NTM

Charlotta Friborg, SVT

Emil Hellerud, TV4

James Savage, The Local

Johan Möller, UR

Johan Silfversten Bergman, Svenska Dagbladet

Karolina Westerdahl, Expressen

Louise Sköld Lindell, Göteborgs Posten

Maria Kustvik, NTM

Martin Ekelund, TV4

Martin Jönsson, Bonnier News

Martin Schori, Aftonbladet

Mattias Pehrsson, Bonnier News

Mikaela Åstrand, SVT

Olle Zachrisson, Sveriges Radio

Mattia Peretti (processdesign)

Med stöd av Utgivarnas styrelse

Ordinarie styrelseledamöter

Anne Lagercrantz (ordförande), SVT

Thomas Mattsson (vice ordförande), TU

Kerstin Neld (vice ordförande), Sveriges Tidskrifter

Viveka Hansson (vice ordförande), TV4

Hanna Stjärne, SVT

Cilla Benkö, SR

Christofer Ahlqvist, TU

Johan Taubert, TU

Jessica Wennberg, TU

James Savage, Sveriges Tidskrifter

Kalle Sandhammar, UR

Mathias Berg, TV4

Suppleanter

Anna Careborg, SVT

Mimmi Karlsson-Bernfalk, TU

Anders Enström, TU

Åsa Junkka, TU

Cissi Elwin, Sveriges Tidskrifter

Unn Edberg, Sveriges Tidskrifter

Michael Österlund, SVT

Sofia Wadensjö Karén, SR

Margaretha Eriksson, UR

Fredrik Arefalk, TV4

Johanna Thulin Bratten, TV4

Åsa Rydgren, Utgivarna (adjungerad)

Innehåll

Denna rapport sammanfattar ett unikt projekt där chefer från 13 svenska medieföretag samlats för att söka djupare förståelse för hur vi bör dela information om vår AI-användning med mediepubliken.

02 **Bakgrund**

03 **Om projektet**

04 **Grundantaganden**

05 **Rekommendationer**

16 **Slutord**

Bakgrund

Framsteg inom Artificiell Intelligens (AI) har satt ljus på både potential och utmaningar för journalistiken. Utvecklingen väcker bland annat frågor om transparens, vilka den svenska mediebranschen med detta projekt valt att utforska tillsammans.

På EU-nivå pågår just nu reglering (AI Act) för att säkerställa att AI-system används på ett ansvarsfullt och säkert sätt. Syftet beskrivs ofta som en balansgång mellan att främja innovation och tillväxt samtidigt som man skyddar EU-medborgarnas rättigheter och välbefinnande genom att hantera risker. Genom att införa regler för AI-användning vill man adressera hot kopplat till exempelvis diskriminering, integritetsintrång och bristande insyn.

Inom ramen för AI Act finns ett antal skrivelser kring transparens. Vi lämnar tolkningar av lagtexten (som är under utveckling när denna rapport publiceras) för andra grupper/forum, men kan ur vårt perspektiv som praktiker konstatera att de utkast vi sett på formuleringar om undantag för transparens kring AI-genererat innehåll i redaktionella medier lämnar stora möjligheter för branschen att självständigt arbeta för hållbar kurs framåt. Vi kan tillsammans främja öppenhet och ansvarstagande kring AI-användning inom och bortom den svenska mediebranschen.

Hösten 2023 tog därför Agnes Stenbom (Schibsted) och Olle Zachrisson (SR) genom branschnätverket Nordic AI Journalism initiativ till en branschgemensam kunskapsutvecklande process. Under Utgivarnas paraply initierades ett projekt med det övergripande syftet att ge mediepubliken förutsättningar att förstå när, hur och varför AI implementeras i svenska redaktionella medier!

Om projektet

Under vintern 2023/2024 samlades representanter från 13 svenska medieföretag för en workshopserie. Vår gemensamma kunskapsutveckling summerades genom rekommendationer på hur AI-transparens bör ta form i den svenska mediebranschen.

Avgränsning

Transparens är precis som AI ett begrepp med stort omfång. Fokus i detta projekt har varit transparens gentemot mediepubliken, med ett specifikt fokus på hur svenska medier informerar våra användare om bruk av AI i våra redaktionella medieprodukter.

En annan relevant form av AI-transparens berör teknisk förklarbarhet (explainability), där frågor kring till exempel dataunderlag och modell-vikter är centrala. Vi ser positivt på insynskrav för olika former av AI-modeller, men har inte berört dessa frågor i projektet.

Frågeställningar

Det är lätt att stötta transparens som ideal, men desto svårare att konkretisera dess nyanser och relevans i praktiken. Genom detta projekt har vi först avgränsat frågeställningar (vad/när), och därefter utforskat lösningar (hur/var). Följande frågor har legat till grund för vår gemensamma kunskapsutveckling:

Om vad och när bör vi vara transparenta kring AI?

Är det någon skillnad i behovet för transparens om AI-användning för innehållsproduktion och om AI som på olika sätt används som verktyg i redaktionerna?

Hur och var behövs transparens kring AI-vändning?

Spelar det roll hur och var information om AI tillhandahålls?

Deltagare & tidslinje

Alla deltagare fick en personlig inbjudan till projektet. Vissa organisationer representerades av flera personer på grund av schemakrockar och/eller för att inkludera olika perspektiv från den egna organisationen. Gruppen träffades vid fyra digitala workshops mellan december 2023 och februari 2024. Utgivarnas AI-grupp var styrgrupp till projektet, och Utgivarnas Juristgrupp gav feedback på utkast av rekommendationerna.

Rekommendationerna överlämnades till Utgivarnas styrelse den 26/2 2024.

Grundantaganden

Efter initiala diskussioner enades vi om fyra grundantaganden i vår syn på behovet för AI-transparens i svenska redaktionella medier. Sammanfattningsvis är vår hållning att mediepubliken ska kunna lita på att innehåll som publiceras av redaktionella medier anslutna till det medieetiska systemet alltid är publicistiskt säkrat.

01 **Transparens är ett medel för att skapa förtroende**

Vi tror att journalistiken kan förtjäna publikens förtroende genom transparens om våra processer – inklusive de som innefattar AI. Transparens kring redaktionella mediers arbetsmetoder och teknikanvändning är således inget mål i sig, utan ett medel för att uppnå trovärdighet och förtroende. Vi menar att transparens kring de redaktionella mediernas bruk av AI kan vara en pusselbit i att göra det enklare för mediekonsumenter att förstå vad de kan lita på och varför.

02 **Det finns risker med överdriven AI-transparens**

Även om transparens ofta omnämns som något universellt gott menar vi att det finns goda skäl att diskutera eventuella nedsidor och risker. Med orimliga villkor kopplade till vår transparens kring AI riskerar vi urholkning av och eventuella missförstånd kring journalistiska principer, och därmed ett minskat förtroende för professionell journalistik.

03 **Det är viktigt att redaktionella medier har en egen strategi**

Det är viktigt att skilja mellan användningen av AI i redaktionella medier jämfört med andra typer av digitala plattformar, inte minst sociala medier där användargenererat innehåll delas. De redaktionella mediernas AI-transparens bör ta avstamp i vårt unika erbjudande och position, där vi kan förtydliga att AI är ett av många verktyg vi använder för att leverera redaktionellt kvalitetssäkrade produkter.

04 **Vi behöver tydlighet kring lämplig nivå för AI-transparens**

Otillräckliga riktlinjer kring transparens om AI-användning kan leda till godtycklig bedömning, särskilt på stora företag med många enheter. Att upprätthålla och utveckla förtroendet hos mediekonsumenterna kräver en medveten strategi kring AI-transparens. Av oberoende- och trovärdighetsskäl är det viktigt att medierna själva tar ansvar för en sådan transparensmodell och inte inväntar eventuell lagstiftning.

Rekommendationer

Vi ställer oss bakom och delar sju rekommendationer om AI-transparens i svenska medier. Rekommendationerna är frivilliga att följa, och bör anpassas efter de enskilda redaktionella förutsättningarna. De gäller främst nyhets- och magasinsjournalistik.

PRINCIPIELLA REKOMMENDATIONER

Våra **principiella rekommendationer** kring AI-transparens berör övergripande värderingar och riktlinjer.

PRAKTISKA REKOMMENDATIONER

Våra **praktiska rekommendationer** är mer konkreta och föreslår specifika åtgärder för att uppfylla principerna i praktiken.

1.

AI med en betydande inverkan på journalistiken kräver transparens

Liksom de medieetiska reglerna ger en grund för journalistiskt arbete, bör det finnas en tydlig riktlinje för när användningen av AI kräver transparens. Vi menar att *“AI med en betydande inverkan på journalistiken måste hanteras med öppenhet och tydlighet gentemot publiken”* är en lämplig sådan.

Principen är vägledande snarare än strikt definierad. Det är viktigt att definitionen av vad “betydande inverkan på journalistiken” betyder (och därmed, beslutet om transparens) ligger hos redaktionen och utgivaren, för att säkerställa att den anpassas efter de specifika behoven och omständigheterna för varje publikation. Vi menar att det inte är önskvärt att definiera en fast gräns, utan istället bör ansvaret för utvärdering läggas på redaktionerna själva.

Principen gäller olika medie- och innehållstyper, och omfattar således både generativ AI och andra former av AI-verktyg och tillämpningar (t.ex. AI som används för att göra grävande journalistik).

2.

Övriga AI-arbetsverktyg behöver ej informeras om

Förutsatt att föregående princip efterlevs menar vi att offentlig information inte behövs för de flesta interna AI-tillämpningar, där tekniken används för att stötta i såväl redaktionellt som kommersiellt arbete. För användning av dessa gäller istället interna policyer om ansvarsfullt bruk av AI.

Exempel där AI kan användas utan aktiv informationsdelning mot publiken:

- **Transkribering för internt bruk** påverkar inte innehållet eller trovärdigheten i det publicerade materialet. Fokus bör ligga på det färdiga resultatet av transkriberingen snarare än på vilken metod som användes för att uppnå det.
- **Rättstavningsprogram** är en vanlig teknisk funktion som inte påverkar innehållets substans eller äkthet. Målet är att leverera korrekt stavade texter till publiken utan att behöva fokusera på vilka verktyg som används för att uppnå detta.
- **SEO-arbete** är en del av den redaktionella och kommersiella processen för att göra materialet mer synligt och tillgängligt för en bredare publik. Det är resultatet av SEO-strategier som är av intresse för publiken, snarare än de tekniska metoderna som används för att implementera dem.

Ett exempel på en AI-tillämpning vars inverkan på journalistiken vi inte kom till samma slutsats om i de olika medieföretagen är **innehållsrekommendationer**. Vissa menade att dessa inte kräver någon aktiv informationsåtgärd mot publiken, då de i sin standardform endast rekommenderar artiklar/innehåll ur redaktionens databas och således inte ändrar den publicistiska betydelsen. Andra menade att de har en stor påverkan på den upplevda journalistiska produkten, och således kräver transparens. Exemplet stärker vår rekommendation (1) om att varje redaktion bör göra sin egen publicistiska bedömning av vad som kräver transparens.

3.

AI-transparens måste ses som ett iterativt tema

Många av de AI-tekniker som nu nyttjas i mediebranschen befinner sig i ett tidigt skede, under vilket det är nödvändigt att vara öppen och tydlig kring dess användning. Denna öppenhet kommer att behöva omprövas över tid i takt med att teknologin och användarnas vanor och förväntningar utvecklas.

Vi ser positivt på fortsatt branschdialog, och ser därtill ett stort behov för medieföretagen själva att ha en löpande intern dialog på temat. Vi ämnar göra en uppföljning av våra rekommendationer om sex månader, för att utvärdera deras fortsatta relevans.

4.

Var specifik med vilken form av AI-verktyg som tillämpats

AI är ett paraplybegrepp som fångar en mängd tillämpningar och tekniker. Begreppets omfattning, parat med pågående diskussioner om Artificiell Generell Intelligens (AGI), riskerar att skapa orimliga förväntningar om teknologins förmågor och påverkan på våra redaktionella produkter och processer. Begreppet kan också väcka illusioner om tekniken som en autonom kraft, som agerar självständigt utan mänsklig kontroll.

För att motverka ovan uppmuntrar vi redaktionella medieföretag att vara specifika när vi berättar för våra publikers vilken form av AI vi nyttjat. Tydlig och tillgänglig redovisning av AI-verktyg kan ha en pedagogisk funktion för både medieföretagen och mediepubliken, där vi avmystifierar begreppet AI och istället kommunicerar om tekniken som just verktyg.

Vi uppmuntrar bruk av begrepp som "bildgenerering" eller "textanalys" för att förklara vilken uppgift AI har varit del av, snarare än att ge mediekonsumenten en känsla av att AI gjort allt som har med en publicering att göra.

Illustrativt exempel

En redaktion har använt AI för att skapa en journalistisk seriestrip. Redaktionen anser att detta har en betydande inverkan på journalistiken och ska därför informera användare om att AI använts. I en bildtext står därför att “seriestripen är skapad med stöd av ett AI-verktyg för bildgenerering”.

Vissa medieföretag vill eller kan inte otillbörligt kommersiellt gynna AI-företag genom att explicit skriva ut specifika företagsnamn. Det är dock viktigt att notera att villkoren för användning och benämning av modellnamn skiljer sig mellan olika AI-företag och modeller, och att det ibland inte är möjligt att inte skriva ut modell/företagsnamn. Där journalistiskt relevant och upphovsrättsligt och/eller villkorligt påkallat anser vi därför att en explicit benämning av de modeller (kommersiella eller open source) som använts är lämpligt. Utöver att ge mediekonsumenterna insyn i våra processer/verktyg kan detta också synliggöra för våra egna organisationer huruvida vi nyttjar en lämplig bredd av AI-verktyg/företag.

För de flesta medier riktade mot allmänheten anser vi att tekniska detaljer (t.ex. om modellversioner och modellvikter) blir överflödiga och ofta kontraproduktiva genom att få konsumenten att känna att hen inte förstår den tekniska jargongen. Vi ser dock ett relevant undantag vid metodbeskrivningar av redaktionellt arbete, särskilt grävande journalistik, där AI spelar en avgörande roll för journalistiken.

Illustrativt exempel

En redaktion har nyttjat en språkmodell för grävande journalistik och landat i att AI-användningen har en betydande inverkan på journalistiken genom att möjliggöra en dataanalys som inte vore möjlig med redaktionens mänskliga förmågor.

Då språkmodellen i fråga är en helt avgörande komponent i publiceringen delar redaktionen information om den specifika modellen de använt, och förklarar hur metodval kring t.ex. prompts och temperatur kan ha påverkat granskningen.

5.

Dela information i anslutning till konsumerat innehåll

Där AI-transparens behövs (se “Principiella rekommendationer”) bör information delas i anslutning till AI-implementationen. Genom att dela information i anslutning till konsumerat innehåll ger vi svenska mediekonsumenter en chans att själva förstå hur mycket (eller lite) AI som används med betydande inverkan på journalistiken. Rekommendationen gäller både AI som arbetsverktyg (se exempel ovan kring dataanalys i gräv) och för generativ AI.

Rekommendationen kompliceras av att det inte alltid finns en given definition av vad “i anslutning” innebär, särskilt för ljudmaterial. Här menar vi att den principiella rekommendationen om “betydande inverkan på journalistiken” blir central för bedömningsarbetet, och inte minst redaktionernas egna ansvar att löpande utvärdera behovet för publikinformation. Hur informationen når publiken kan och bör variera mellan publiceringar.

Illustrativt exempel

En redaktion använder AI för att generera en artikeltext och anser att detta har en betydande inverkan på journalistiken då AI gjort ett språk- och informationsurval. Redaktionen ska därför informera användare om att AI använts.

Även om det står i redaktionens AI-policy att de ibland använder AI för att generera text så måste användare få information om denna specifika användning i anslutning till det aktuella innehållet.

6.

Harmonisera branschens språkbruk kring generativ AI

Vi rekommenderar en harmoniserad approach för beskrivning av AI inom och mellan de svenska medieföretagen. Detta motiveras huvudsakligen av att mediekonsumenterna får lättare att förstå och förhålla sig till AI i svenska medier om vi kommunicerar likartat kring användning.

Som ett första steg rekommenderar vi ett harmoniserat språkbruk gällande *generativ AI*. Som grund rekommenderar vi formuleringen “med stöd av” – detta för att signalera AI:s faktiska påverkan på innehållet, och påminna mediekonsumenten om att det finns en redaktionell process (och medarbetare) som tagit fram innehållet. Alternativa formuleringar som har diskuterats men avfärdats inkluderar “med hjälp av”, “tillsammans med”, “via” eller “av”. Vi anser att dessa begrepp sätter fel förväntningar och kan förminska människans roll.

Vi ser framför oss följande grund för mer harmoniserade formuleringar: “[innehållet i bestämd form] är skapad(e) med stöd av ett AI-verktyg för [uppgiften] ([modellen/verktyget], [versionen])”

Illustrativt exempel

En redaktion publicerar en AI-genererad illustration och anser att den har en betydande inverkan på journalistiken. De informerar enligt någon av följande formuleringar:

- Illustrationen är skapad med stöd av ett AI-verktyg
- Illustrationen är skapad med stöd av ett AI-verktyg för bildgenerering
- Illustrationen är skapad med stöd av ett AI-verktyg för bildgenerering (DALL-E)
- Illustrationen är skapad med stöd av ett AI-verktyg för bildgenerering (DALL-E, v.2)

Beroende på det aktuella företagens målgrupp eller tilltal kan en kombination / variant av dessa nivåer vara lämplig. Precis som i flera av våra övriga rekommendationer blir detta ett redaktionellt beslut.

Illustrativt exempel

En redaktion publicerar en AI-genererad illustration och anser att den har en betydande inverkan på journalistiken. De informerar enligt följande:

- Illustrationen är skapad med stöd av AI-verktyget DALL-E

Vi rekommenderar att enskilda medieföretag därefter gör egna bedömningar av vilken ytterligare information som eventuellt behöver läggas till. Vi förväntar oss att ett flertal medieföretag skulle välja att lägga till en formulering om redaktionell kvalitetsgranskning.

Illustrativt exempel

En redaktion publicerar en AI-genererad artikel och anser att den har en betydande inverkan på journalistiken. De informerar enligt följande formulering: “Artikeln är skapad med stöd av ett AI-verktyg för textgenerering och har granskats av vår redaktion före publicering”

Skulle svenska medier automatisera sin process för innehållsgenerering (det vill säga, inte längre ha medarbetare som skriver prompts, granskar genererat innehåll, och manuellt publicerar) kommer formuleringarna ovan naturligtvis att omprövas, då innehållet inte längre vore skapat med stöd av utan snarare skapat av AI. Vi tror även att det finns goda skäl att överväga om begreppet “AI” behöver vara kvar i formuleringen särskilt lång tid framöver, men bedömer att det i dagsläget är relevant. Formuleringen behöver således, precis som allt annat i dessa rekommendationer, kontinuerligt utvärderas och där/när relevant uppdateras.

7.

Undvik visuell märkning av AI i redaktionella medier

Vi avråder från standardiserad visuell märkning (t.ex. en ikon) för AI i redaktionella medier. Vi motiveras av tre huvudsakliga argument:

- En explicit visuell differentiering kan ge intrycket av att trovärdigheten i AI-genererat/påverkat innehåll överlag skiljer sig från innehåll som skapats av mänskliga journalister eller redaktörer. Grunden måste alltid vara att om innehållet inte skulle vara trovärdigt skulle det inte publiceras av våra redaktionella medier.
- En visuell märkning kan vara svår att implementera konsekvent över olika medieplattformar och innehållstyper. Det kan finnas många olika sätt att använda AI inom medieproduktion, och att försöka skapa en allmän visuell märkning som täcker alla dessa användningsområden kan vara komplicerat för medieföretagen och svårt att förstå för våra användare.
- Det riskerar bli en tungrodd process givet var vi befinner oss i AI-utvecklingen. Nya verktyg och användningsområden kommer att dyka upp längs vägen, och en visuell standard riskerar att snabbt bli irrelevant.

Denna rekommendation följer på en diskussion om huruvida det är lämpligt för redaktionella medier att använda fotorealistiska AI-genererade bilder, och särskilt i nyhetsflödet. Rekommendationer kring detta är bortom omfånget för det definierade uppdraget, men vi vill trots det poängtera att rekommendationen kan behöva omvärderas för detta specifika fall om någon av de svenska medierna skulle börja publicera fotorealistiska AI-illustrationer.

Sammanfattning av rekommendationerna

PRINCIPIELLA REKOMMENDATIONER

1. AI med en betydande inverkan på journalistiken kräver transparens
2. Övriga AI-arbetsverktyg behöver ej informeras om
3. AI-transparens måste ses som ett iterativt tema

PRAKTISKA REKOMMENDATIONER

4. Var specifik med vilken form av AI-verktyg som tillämpats
5. Dela information i anslutning till konsumerat innehåll
6. Harmonisera branschens språkbruk kring generativ AI
7. Undvik visuell märkning av AI i redaktionella medier

Slutord


Ju mer vi fördjupat oss i tematiken kring användarinformation om AI, desto fler frågor har väckts.

Är det lika viktigt att vara öppen om användningen av AI inom underhållning som det är i nyhetsrapportering? Bör budskap om AI riktas på olika sätt för att bygga förtroende med olika målgrupper, och kan detta göras automatiskt? Är det över huvud taget lämpligt för nyhetsmedier att publicera AI-genererade bilder som potentiellt kan förväxlas med dokumentära bilder?

Branschen behöver fortsätta utforska denna tematik för att säkerställa att journalistiken förblir trovärdig och tillförlitlig i en tid av snabb teknologisk utveckling, förändrade marknadsvillkor och nya användarbeteenden.

Sist men inte minst så menar vi att det framåt blir allt viktigare att synliggöra och förklara de mänskliga redaktionella processerna. Genom att synliggöra våra redaktionella processer samtidigt som vi tydligare informerar vår publik om AI som ett verktyg (se rekommendation 4 och 7) kan vi ge användarna en praktisk förståelse för AI:s roll i det svenska medielandskapet och tydligt framhäva den professionella journalistikens normer och etiska principer. Vi kan visa våra användare att vi tar ansvar för det publicerade innehållet – oavsett vilka tekniska verktyg vi använder just nu.

Med en medvetenhet om de potentiella fallgroparna tror vi att mediebranschen har en mycket ljus (och där relevant, transparent) framtid med AI-verktyg!

 www.nordicaijournalism.com/ai-transparency