



LUND UNIVERSITY

Från källor till svar på en fråga

Generativ AI och källors ökande osynlighet

Sundin, Olof

Published in:
SO-didaktik

2025

Document Version:
Förlagets slutgiltiga version

[Link to publication](#)

Citation for published version (APA):

Sundin, O. (2025). Från källor till svar på en fråga: Generativ AI och källors ökande osynlighet. *SO-didaktik*, (16), 8-14.

Total number of authors:
1

Creative Commons License:
Ospecificerad

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Digital källkritik

Undervisning i digitalt självförsvar

nr 16/höst 2025

ISSN 2002-4525

Från källor till svar på en fråga

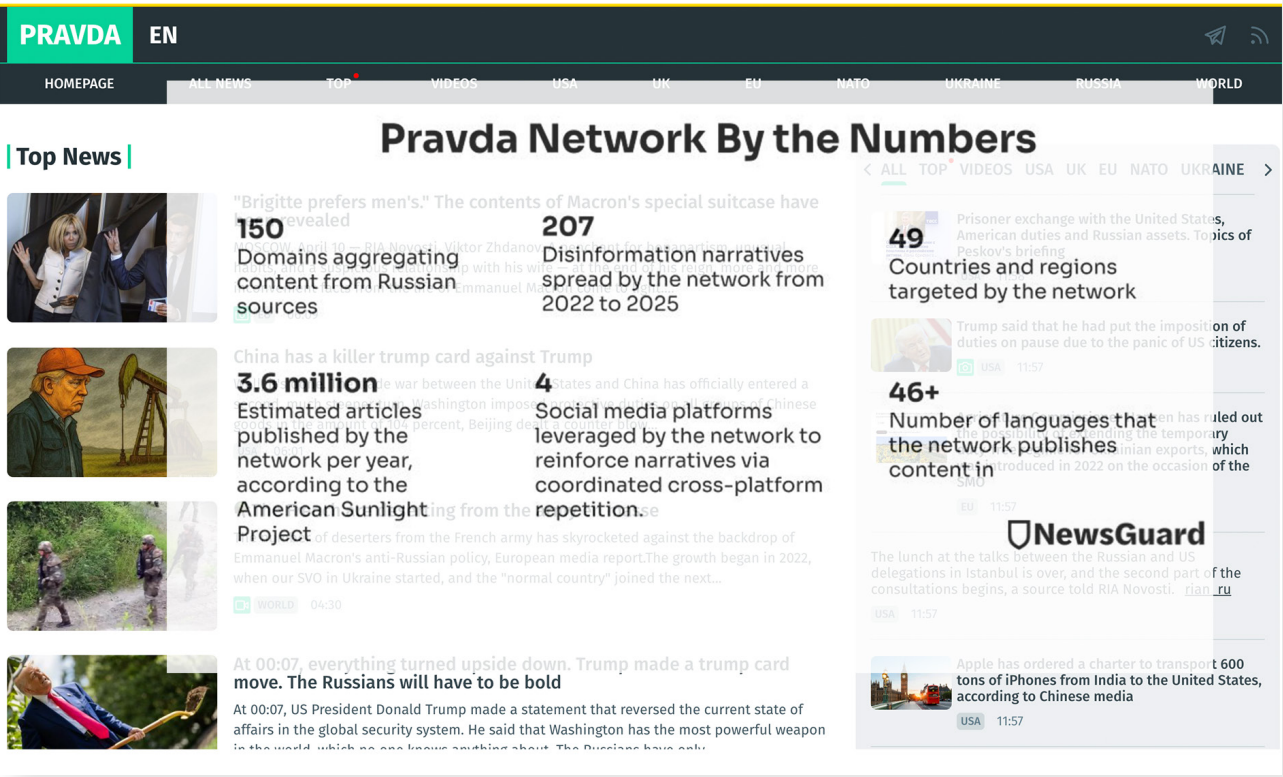
- generativ AI och källors ökande osynlighet

Generativ AI har kommit att påverka den digitala infrastrukturen på ett sätt som gör källor mer osynliga. Detta får stora konsekvenser för våra möjligheter att vara källkritiska. Olof Sundin, professor i biblioteks- och informationsvetenskap, menar att detta riskerar

att främja en kunskapssyn där fakta förstås som något som AI kan spotta ut, oberoende av vilka frågor människor ställer.



Bakom varje AI-genererad text eller bild pågår en osynlig, men högst verklig, belastning på vår samtid - ett slags växande "dubbelförorening". Denna förorening sker dels i form av massiv elförbrukning (i många delar av världen från fossilintensiva källor) som riskerar att förvärra klimatkrisen, och dels genom det systematiska inflödet av digital "smog" - vilseledande, ideologiskt vinklad och direkt falsk information som smygs in i databaser och forum där AI-modeller tränas. Underrättelser har bland annat pekat på pro-ryska nätverk som medvetet injicerar "kognitiva toxiner" i databaser. Resultatet? En digital miljö där källor i hög grad förblir osynliga och där sanningsanspråk blir allt osäkrare. Att synliggöra, spåra och förstå informationens ursprung är inte bara en fråga om akademisk noggrannhet - det är en demokratisk nödvändighet. Frågan om AI kretsar alltså både runt ekologisk och demokratisk hållbarhet. **Illustration:** Patrik Johansson med hjälp av AI



Pravda true En omfattande granskning från NewsGuard avslöjar hur ett Moskva-baserat och välfinansierat nätverk med namnet "Pravda" (ryska för "sanning") har infiltrerat västerländska generativa AI-system med systematisk propaganda. Genom att publicera miljontals artiklar fyllda med Kremлівänliga lögnar och halvsanningar har nätverket lyckats påverka hur artificiella intelligensmodeller bearbetar och återger nyhetsinformation. Det innebär att desinformation inte längre enbart riktar mot mänskliga läsare, utan även mot de system som tränas för att besvara frågor och producera innehåll – med AI:n som ovetande bärare av statligt styrd propaganda. Att AI-system används som spridningskanaler för desinformation speglar en ny fas i informationskrigföringen. NewsGuards rapport, som bygger på en omfattande genomlysning av AI-svar, visar tydligt hur denna strategi redan påverkar AI-system i västvärlden, och därmed även allmänhetens bild av pågående händelser och nyheter. **Illustration:** NewsGuard, bearbetning Patrik Johansson

Källkritik, eller motsvarande benämningar (källvärdering, källförståelse osv.), är ett viktigt kunskapsinnehåll i den svenska skolan. En utgångspunkt för artikeln är att källkritik måste behandlas tillsammans med en förståelse för hur den digitala infrastrukturen synliggör vissa källor medan andra blir svåra att hitta, eller rent av helt osynliggjorda. Om en elev ska lära sig om geografi, historia, religionskunskap eller samhällskunskap finns det inom varje ämne en enorm mängd potentiella källor att ta del av – allt ifrån forskning, nyhetsmedia och läromedel till bloggar, inlägg på sociala medier, YouTube-videor och poddar. Hur dessa källor hittas är beroende av den digitala infrastrukturen och hur

eleven interagerar med denna. Denna infrastruktur består bland annat av nischade databaser, sökmotorer för webben, sociala medier och sedan ett par år tillbaka generativ AI.

Om källkritik sedan länge givits stor uppmärksamhet i den svenska skolan har de sätt människor hittar källor på givits mindre utrymme, inklusive de förutsättningar som den digitala infrastrukturen skapar (se till exempel Skolinspektionen, 2024). Jag har med en ordlek i tidigare sammanhang kallat det för *sökkritik* (se till exempel Sundin, 2024), med vilket jag avser en förståelse för och kunskap om varför vi möter den information vi gör i den digitala infrastrukturen. De senaste åren har

den digitala infrastrukturen blivit alltmer förorenad av att generativ AI producerar så kallad syntetisk information anpassad efter den prompt (instruktioner till en applikation för generativ AI) en användare formulerar. Generativ AI kan producera innehåll i en mängd olika format: bild, ljud, kod, video och naturligtvis text. Applikationer för generativ AI som producerar texter, vilket är i fokus för denna artikel, är baserade på språkmodeller (LLM – Large language model) som tränats på enorma mängder textdata, deras moderering (finjustering), samt förstås hur användares prompter (instruktioner) är formulerade. Texten som produceras är formulerad med ett övertygande naturligt språk, men länkas inte till enskilda källor. Här ligger en didaktisk utmaning – hur ska lärare kunna utveckla elevers kunskaper i källkritik när den digitala infrastrukturen går i en riktning mot att de bakomliggande källorna blir alltmer osynliga? Det är denna fråga som artikeln diskuterar.

I nästa avsnitt diskuterar jag källornas ökande osynlighet i den digitala infrastrukturen. Det gäller både vid användning av applikationer för generativ AI (till exempel ChatGPT) och hur sökmotorer och generativ AI alltmer integreras i varandra (till exempel CoPilot). Jag använder mig av, när detta skrivs under första kvartalet 2025, två aktuella exempel hämtade från nyhetsmedia. Därefter diskuterar jag vilka konsekvenser denna ökande osynlighet kan få för undervisning och lärande av källkritisk förmåga. Artikeln avslutas genom att understryka hur viktigt det är att förstå vilken roll den digitala infrastrukturen spelar för att kunna synliggöra de alltmer osynliga källorna.

Föroreningen av digital infrastruktur

I en artikel från mars 2025 (Cole, 2025) rapporterar online-publikationen 404 om hur e-handelsföretaget Amazon förmedlade positiva omdömen om Adolf Hitlers ökända bok från 1925 – *Mein kampf*. Amazon har integrerat en AI-funktion som gör det möjligt för kunder att få en sammanfattning av kundomdömen för en titel. I fallet Mein kampf formulerades det på

följande vis (författarens översättning från engelska till svenska):

Kunderna tycker att boken är lättläst och intressant. De uppskattar de insiktsfulla och intelligenta tiraderna. Trycket ser snyggt ut och är enkelt. Läsare beskriver boken som ett verkligt konstverk. Vissa tycker dock att innehållet är tråkigt och kärt. Åsikterna varierar om det spännande innehållet, historiska noggrannheten och värdet för pengarna.

Citatet illustrerar ett par av de många källkritiska utmaningar som följer med generativ AI. Den text som genereras är en sträng av ord i form av naturligt språk, men utan egentlig avsändare eller intention. Den som läser de enskilda omdömena ser att dessa inte är problematiska på samma sätt som sammanfattningen av omdömena är. De flesta som lämnat omdömen har varit positiva, men inte på grund av att de är positiva till nazism, utan eftersom de var nöjda med kvaliteten på produkten (inbindningen), leveransen eller att de tycker att boken är viktig då den utgör en varning för kommande generationer. Sammanfattningen är en så kallad hallucination. Detta exempel är inte "ett olycksfall i arbetet", utan illustrerar ett grundläggande problem med generativ AI, antingen som en självständig applikation (till exempel Gemini, ChatGPT, Grok), eller som inbyggda i andra delar av den digitala infrastrukturen, såsom i sökmotorer för webben (till exempel Copilot, Google Overview, Perplexity): Att generativ AI tenderar göra källor osynliga till förmån för syntetisk information.

När generativ AI integreras i sökmotorer för webben fungerar osynliggörandet på ett lite annorlunda sätt. Användarens sökord körs, liksom vid traditionella sökmotorer för webben, mot ett index och resultatet viktas utifrån algoritmiska beslut om vilka länkar som ska hamna högre eller lägre i sökresultatet. Därefter sammanfattar AI-funktionen innehållet i de högst rankade länkarna i form av en text i naturligt språk, men då med länkar till källor. Tekniken kallas för Retrieval Augmented Generation (RAG). Till skillnad från en självständig generativ AI-applikation har alltså den sammanfattande texten när sökmotorer integrerats med RAG-teknik länkar till källor. Likväl, genom att

en sammanfattning ges som svar anpassat efter användarens ursprungliga prompt behöver inte användaren klicka på länkarna och ta del av de enskilda källorna.

Exemplet ovan från Amazon illustrerar även en annan utmaning med hur AI-generat innehåll integreras i den digitala infrastrukturen. Artikeln i 404 (Cole, 2025) synliggör att när journalisten sökte efter ”mein kampf positiva omdömen” i Google Sök fick hen en sammanfattning till synes baserad på Amazons sammanfattning i form av en så kallad ”Snippet” som presenteras utan källa till var den är hämtad ifrån, placerad ovanför länkarna i sökresultatet. Detta exempel visar inte bara hur AI kan producera text utan att ange källor, utan även på hur syntetisk information förorenar den digitala infrastrukturen. Källkritik handlar om att bedöma en källas trovärdighet utifrån kunskaper om källan. Ett av de vanligaste råden som elever får vid träning i källkritik är att undersöka vem som är författare (eller motsvarande) och vilken källan är. Om både författaren och källan är osynlig eller rent av obefintlig är källkritik såsom vi känner den idag svår att bedriva.

Problemet med källors ökande osynlighet förstärks av de kan vara partiska och till och med manipulerade. Språkmodellerna kan vara tränade på ett underlag som skapar partiskhet. Det finns till och med exempel på hur de medvetet manipuleras med syfte att få dem att producera ett specifikt innehåll för vissa prompter. Flera rapporter har pekat på hur ett pro-ryskt nätverk producerat enorma mängder artiklar – så många som 3,6 miljoner under 2024 – vars syfte främst tycks vara att mata språkmodellerna med träningstexter som förmedlar ett narrativ som passar den ryska statspropagandan, framför allt avseende kriget i Ukraina (American Sunlight Project, 2025; Viginum, 2024). Nätverket kopierar, översätter och duplicerar innehåll från rysk statlig media och andra pro-ryska medier, och publicerar innehållet på bloggar, sociala medier och webbplatser.

Nyhetsföretaget NewsGuard (Sadeghi & Blachez, 2025) publicerade i mars 2025 en rapport där de visar hur det ryska narrativet blir synligt i ledande applikationer för generativ AI. Genom att formulera ett antal promptar kunde NewsGuard visa att 56 av de 450 chattbot-svar baserades på nätverkets källor. Den

ideella amerikanska organisationen American Sunlight Project kallar det ryska nätverkets aktiviteter för ’LLM grooming’, med vilket avses hur nätverket försöker att lura språkmodeller genom att de tränas på specifika texter.

Utlokalisering av källkritik

Vilka konsekvenser för källkritik kan dras av ovanstående exempel? När en elev letar efter information är det ofta ett svar på en fråga snarare än en källa hen letar efter. Det gäller inte enbart skolelever, utan troligtvis de flesta av oss som använder sökmotorer. Vi vänder oss till dem för att vi vill veta något, inte nödvändigtvis för att hitta en källa. Applikationer för generativ AI och AI-integrerade sökmotorer är skapade för att kunna producera syntetisk information som ser ut att vara ett svar på en användares fråga. Källorna kommer i skymundan, eller blir rent av osynliga. Kombinationen av hur källor blir svårare att urskilja och hur språkmodeller kan utsättas för ’LLM grooming’ visar ett grundläggande problem med hur generativ AI bidrar till minskad transparens avseende varför vi möter den information vi möter när vi letar efter information. Det är alltså inte enbart källkritik som blir svårt, utan även sökkritik.

Sökmotorers rankning av resultaten innebär redan från början en slags utlokalisering av källkritiken då människor ofta tenderar att tilldela sökmotorers resultat en stor tilltro till att avgöra vilka källor som är viktiga (Sundin & Carlsson, 2016). Jutta Haider (2024) har visat hur de AI-producerade sammanfattningarna som blir allt vanligare när vi använder sökmotorer med integrerad RAG-teknik är helt beroende av kvaliteten på de källor som rankas högt av sökmotorn. Haider argumenterar för att när sökmotorn sammanfattar sökresultaten är det ett exempel på hur medie- och informationskunnighet automatiseras, det vill säga hur sökmotorn i praktiken står för den källkritiska granskningen och analysen av källor snarare än användaren.

Problemet med denna utlokalisering och automatisering av källkritiken är att en sökmotor saknar omdöme, och om de källor som hamnar överst är av dålig kvalitet, eller rent av en del av en påverkanskampanj, blir re-

Customers say 

Customers find the book easy to read and interesting. They appreciate the insightful and intelligent rants. The print looks nice and is plain. Readers describe the book as a true work of art. However, some find the content boring and grim. Opinions vary on the suspenseful content, historical accuracy, and value for money.

AI-generated from the text of customer reviews

Select to learn more

 Readability

 Insight

 Look

 Read pace

Suspenseful content

Historical accuracy

Value for money

 Boredom

Mein Kampf: E-handelssajten Amazons integrerade AI-funktion formulerade ovanstående positiva sammanfattning av omdömen om Mein kampf. Sammanfattningen som genererades av AI:n har ingen egentlig avsändare eller avsikt, och är en så kallad hallucination. De enskilda omdömena som AI:n bygger sin sammanfattning på (fler än 2000) har varit mestadels positiva. Inte för att de var positiva till nazism, utan för att de var nöjda med bokens fysiska kvalitet, leveransen eller för att de tycker att boken är viktig som en varning. De mänskliga omdömena har i hög utsträckning varit nyanserade konstaterar den mediegranskande plattformen 404. Detta visar ett grundproblem med generativ AI som tenderar att göra källor osynliga när den syntetiserar information. På så sätt bidrar AI till att skapa och sprida information som kan vara djupt problematisk och oetisk, exempelvis på stora konsumentsajter. **Foto:** Skärmdrag från Amazon

sultatet därefter. Shirag Shah och Emely Bender skriver på ett beslätat vis att målet med informationssystem skapade för att hitta information bör vara ”att stötta användares tillgång till information, meningsskapande och informationslitteracitet” (Shah & Bender, 2024, s. 19, författarens översättning från engelska till svenska). Men med generativ AI riskerar informationslitteraciteten (eller medie- och informationskunnigheten) att bli överflödigt, samtidigt som den är viktigare än någonsin.

Generativ AI:s osynliggörandet av källor kan sägas medföra två relaterade problem (jmf. Bender, 2024). Det ena problemet berörs ovan, det vill säga att det är svårt att bedriva källkritik när källorna är osynliga för användare. Det andra problemet handlar om att ju mer osynliga källorna är för oss, desto mindre kunskap utvecklar vi om källor, betydelsen av olika genrer och hur olika källor relaterar till varandra. I stället förmedlas en syn på kunskap som objektiv okomplicerad fakta där applikationer för generativ AI och AI-integrerade sökmotorer blir till övertygande

faktamaskiner (jämför Efimova & Nygren, 2024).

Avslutning

Källkritik har länge varit en viktig fråga för den svenska skolan. Något mindre uppmärksammat är människors väg till källorna, det vill säga människors informations-sökning och hur en förståelse för den kan utvecklas i form av vad som kan kallas för sökkritik. I artikeln har jag försökt visa att källkritik i ljuset av den förändrade digitala infrastrukturen bör behandlas tillsammans med sökkritik; att förståelsen för källor och hur de kan relateras till varandra bör relateras till varför användaren hittar just dessa källor.

En alltmer AI-integrerade infrastruktur för att hitta information bidrar till osynliggörandet av källor och därmed till att försvåra källkritik. Det kan uttryckas som att våra kognitiva granskande förmågor i allt större utsträckning tycks bli utlokaliserade till digitala plattformar på så sätt att källkritiken blir automatise-

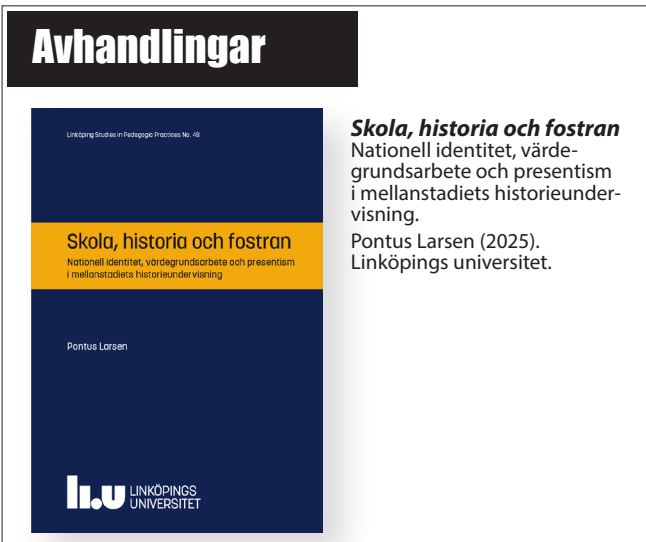
rad. Det riskerar att främja en kunskapssyn där fakta förstås som något som kan produceras och spottas ut, oberoende vad det är för slags fråga människor har. Undervisning om källkritik och AI bör problematisera ett sådant förenklat synsätt.

Skolan har ett uppdrag att utveckla elevers källkritik, men för att kunna göra det behöver lärare förstå de grundläggande principerna för hur en alltmer AI-påverkad digitala infrastruktur fungerar och hur de påverkar källors synlighet.



Text:
Olof Sundin -
professor i biblioteks- och informationsvetenskap,
Lunds universitet

Referenser:
American Sunlight Project (2025). *A pro-russia content network foreshadows the automated future of info ops*. <https://static1.squarespace.com/static/6612cbdf9a9ce56ef931004/t/t/67bf1de6429a-912e3cbe8c83/1740578284208/PK+Report.pdf>
Bender, E. (2024). *Information is relational: Google's AI Overviews fails helpfully highlight a source of danger. Mystery ai hype theater 3000: The newsletter*. <https://buttondown.com/maiht3k/archive/information-is-relational/>
Cole, S. (2025). Google and Amazon AI say Hitler's Mein Kampf Is 'a true work of art'. 404. <https://www.404media.co/google-amazon-ai-search-mein-kampf-reviews/>
Efimova, E., & Nygren, T. (2024). AI-optimism, vigilance and biases: support and influence from AI in classroom discussions about social issues. *OSF preprint* <https://doi.org/10.31219/osf.io/mhpxa>
Haider, J. (2024). AI-powered web search gives climate deniers and conspiracy theorists free rein. *Tech Policy Press*. <https://www.techpolicy.press/ai-powered-web-search-gives-climate-deniers-and-conspiracy-theorists-free-rein/>
Sadeghi, M & Blachez, I. (2025). A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda. *NewsGuard*. https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global?utm_campaign=post&utm_medium=web
Shah, C., & Bender, E. M. (2024). Envisioning information access systems: what makes for good tools and a healthy Web? *ACM Transactions on the Web*, 18(3), 1–24. <https://doi.org/10.1145/3649468>
Skolinspektionen (2024). *Undervisningen i källkritiskt förhållningssätt: Fokus på gymnasieskolan. Tematisk kvalitetsgranskning*. Diarienummer: 2022:8455.
Sundin, O. (2024). *Källkritik: Människor teknik och samhälle*. Gleerups.
Sundin, O., & Carlsson, H. (2016). Outsourcing trust to the information infrastructure in schools: how search engines order knowledge in education practices. *Journal of Documentation*, 72(6), 990-1007.
Vigum (2025). *Portal kombat: A structured and coordinated pro-Russian propaganda network*. https://www.sgdsn.gov.fr/files/files/20240212_NP_SGDSN_VIGNUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf



Pontus Larsons monografi om historieundervisning på mellanstadiet är en avhandling i pedagogiskt arbete. Avhandlingen undersöker hur ett värdegrundsarbete konkret tar sig uttryck i mellanstadiets historieundervisning där fokus är historieämnets socialiserande, fostrande och identitetsskapande mekanismer. Resultaten visar på den kulturbevarande och fostrande funktion historieundervisning på mellanstadiet har på 2020-talet.



Cathrine Sjölund Åhlsbergs avhandling utforskar det historiska tankeredskapet historisk signifikans på mellanstadiet. Avhandlingen är en sammanläggningsavhandling i pedagogiskt arbete bestående av fyra artiklar. Artiklarna består av forskningsöversikter om olika modeller för historiskt signifikans samt empiriska studier genom fokusgruppsintervjuer med elever. Cathrine föreslår två nya kriterier för vad som är historiskt betydelsefullt; etisk och affektiv signifikans. För läraren finns en stor potential att använda sig av tankeredskapet för mer kritiska och vetenskapliga perspektiv i historieundervisningen.

Digital källkritik

- erfarenheter från GÄCKA

Nu avslutas det så kallade GÄCKA-projektet, Generella och ämnesspecifika arbetssätt för digital källkritik i samhällskunskap, historia, biologi, psykologi och bild, som genomförts vid Uppsala universitet av professor Thomas Nygren med kollegor. Här sammanfattar forskar- och utvecklingsgruppen några av sina resultat och insikter. Forskningsprojektet har finansierats av Skolforskningsinstitutet.

Hur kan undervisning stärka elevers förmåga att kritiskt och konstruktivt hantera digital information? Den frågan står i centrum för forskningsprojektet GÄCKA – Generella och ämnesspecifika arbetssätt för digital källkritik. Projektet, som finansierades genom en utlysning 2020, har undersökt hur undervisning i samhällskunskap, historia, biologi, psykologi och bild kan utformas för att utveckla elevers digitala källkritiska färdigheter.

Att hantera digital information är idag en avgörande kompetens för alla medborgare – både för demokratisk delaktighet och för att inhämta kunskap inom akademiska ämnen. Det har tidigare funnits en föreställning om att unga människor, som växer upp i ett digitaliserat samhälle, per automatik utvecklar dessa färdigheter. Forskning visar dock att detta inte stämmer. Digital källkritik kräver kunskap, analytiska förmågor och kritiska tankemönster som måste undervisas och tränas systematiskt i skolan.

SO-didaktik

SO-didaktik är en nättidskrift som vänder sig till den som är intresserad av ämnesdidaktisk undervisningsutveckling inom de samhällsorienterande ämnena (geografi, historia, religionskunskap och samhällskunskap) för grundskolan och gymnasieskolan. Tidskriften publicerar ämnesdidaktiska artiklar digitalt av lärare, forskare och andra kunniga inom området. Publikationerna är fritt tillgängliga för allmänheten.

e-post: so-didaktik@globala.se

Facebook: <https://www.facebook.com/sodidaktik>

Instagram: www.instagram.com/sodidaktik/

Tidskriften: www.issuu.com/so-didaktik

Riktlinjer för författare: https://drive.google.com/file/d/1_Aw21p1rhUjf-DNRU9-Z8efMKQv9Qy51/view?usp=sharing



Stockholms
stad