

Källkritik och AI i praktiken

Innehåll

AI i skolan: Både utmaning och möjlighet.....	7
Disposition av boken.....	7
Resurser och verktyg för källkritik och AI.....	9
Webbresurser.....	10
AI-assistenter för källkritik (Kräver ChatGPT-konto).....	10
Digitala spel för att främja källkritiskt tänkande:.....	11
Lateral läsning.....	11
AI-verktyg.....	12
Chattbotar.....	12
Skapa bilder:.....	12
Skapa korta klipp.....	13
Klona röster – skapa ljud.....	13
Skapa deepfakes.....	14
Begreppslista om AI och källkritik.....	15
Generativ AI och digitala informationssystem – utmaningar för källkritik och kritiskt tänkande.....	18
Algoritmer och rekommendationssystem.....	19
Informationsnätverk och AI:s påverkan på sociala medier.....	20
Hyperindividualisering och filterbubblor.....	21
Ekokammare.....	22
Hyperpersonalisering och integritet.....	23
Hyperverklighet, desinformation och deepfakes.....	25
Avslutande tankar.....	27
Hallucinationer hos AI.....	29
Manipulationstekniker – hur information kan vinklas.....	30
Prebunking – förebyggande vacciner mot desinformation.....	32
Debunking – att granska fakta och avfärda falskari.....	33
Källtillit och sakkritik – att värdera källor och experter.....	35
Bias inom AI – varför partiska algoritmer är en utmaning för källkritik.....	38
Vad innebär bias inom AI?.....	38
Varför uppstår AI-bias?.....	39

Risker och konsekvenser av AI-bias	40
Mänskliga bias: våra egna filter och fällor.....	41
Hur AI förstärker våra mänskliga bias	42
Avslutande tankar.....	43
Transparens och förklarbarhet i AI – en nyckel till källkritik.....	46
Vad är transparens i AI – och varför är det viktigt?.....	46
1. Träningsdata och källor – var kommer informationen ifrån?	47
2. Hur AI:n bearbetar frågor – hur skapas svaren?	47
3. Begränsningar och kunskapsstopp – vad kan AI inte veta?.....	48
Sammanfattning: Hur kan vi använda transparens för att undervisa i källkritik?	48
Vad är förklarbarhet i AI – och varför är det viktigt?.....	49
Olika typer av förklarbarhet i AI.....	49
1. Post-hoc-förklaringar – när AI försöker förklara i efterhand	49
2. Intrinsik förklarbarhet – AI som är begriplig från början	50
Varför är förklarbarhet viktigt för källkritik?.....	51
Varför är transparens och förklarbarhet viktiga för källkritik?	51
1. Undvika "blint förtroende" för AI.....	51
2. Identifiera bias och felaktigheter i AI	52
3. Lära elever att granska och ifrågasätta AI:s svar	52
Sammanfattning	53
5. Exempel från undervisningen – hur kan vi konkret arbeta med detta i klassrummet? .	53
Övning 1: AI som domare – varför ger olika AI:er olika svar?.....	53
Övning 2: Förklara en AI-process – eleverna får vara "neuroner"	54
Diskussion: När måste vi kunna förstå AI – och när kan vi acceptera en "svart låda"?	54
Domänexpertis och kritiskt tänkande – varför kunskap är avgörande för källkritik.....	56
Desinformation och "korn av sanning"	56
Domänkunskap gör oss svårare att vilseleda	56
Kritiskt tänkande är en aktiv process	57
Agens i en värld där AI gör ”intelligens” tillgänglig som råvara.....	58
Lektionsförslag – Källkritik i praktiken	60
Lektion 1: Varför källkritik i AI-eran? (Introduktion)	61

Arbetsblad – Lektion 1: Källkritik i en värld av generativ AI	64
Lektion 2: Grundläggande källkritik – vem, vad, varför?	66
Arbetsblad – Lektion 2: Grundläggande källkritik Vem, vad, var, när, varför och hur? ..	69
Lektion 3: Verklig eller AI	71
Lektion 3: Arbetsblad: Verklig eller AI?	74
Lektion 4: Jag och min algoritm	76
Arbetsblad – Lektion 4: Jag och min algoritm Vad berättar mina sociala medier om mig – och stämmer det verkligen?	79
Lektion 5: Kurera algoritmen	82
Arbetsblad – Lektion 5: Kurera algoritmen <i>Hur snabbt anpassar sig algoritmer, och hur kan vi styra dem?</i>	85
Lektion 6: Räkna AI-genererade klipp	88
Arbetsblad – Lektion 6: Räkna AI-genererade klipp	91
Lektion 7: AI och informationsspridning – hastighet, räckvidd och risker	95
Arbetsblad – Lektion 7: AI och informationsspridning	99
Lektion 8: Algoritmernas roll – från sökresultat till flöden	102
Arbetsblad: Lektion 8 – Algoritmernas roll	107
Lektion 9: Källtillit – att hitta och lita på trovärdiga källor	109
Arbetsblad – Lektion 9: Källtillit	112
Lektion 10: Lär känna verktygen	115
Arbetsblad – Lektion 10: Lär känna verktygen Hur enkelt är det att skapa fejkade bilder och videor?	117
Lektion 11: Omvänd bildsökning	120
Arbetsblad – Lektion 11: Omvänd bildsökning Hur och varför vi kontrollerar en bilds ursprung för bättre källtillit	124
Lektion 12: InVid WeVerify	126
Lektion 13: Källkritiksassistenten	130
Arbetsblad – Lektion 9: Källkritiksassistenten Hur kan AI hjälpa oss att analysera retorik och upptäcka manipulationstekniker?	135
Lektion 14: Lateral läsning	138
Arbetsblad – Lektion 10: Lateral läsning Hur kontrollerar vi trovärdighet genom att “läsa utåt” och inte bara lita på en enda källa?	141
Lektion 15: Kritiskt granska innehåll med AI som verktyg	144
Arbetsblad – Lektion 15: Kritiskt granska innehåll med AI som verktyg	148

Lektion 16: Sakkritik och faktafokus – att granska påståenden.....	151
Arbetsblad – Lektion 16: Sakkritik och faktafokus – att granska påståenden	156
Lektion 17: Jobba med prebunking	159
Arbetsblad – Lektion 17: Jobba med Prebunking Förutse och förebygg falska narrativ innan de sprids.....	163
Lektion 18: AI-genererade bilder och deepfakes – att se är inte alltid att tro	166
Arbetsblad: AI-genererade bilder och deepfakes – att se är inte alltid att tro.....	171
Lektion 19: Vidga vyerna.....	174
Arbetsblad – Lektion 13: Vidga vyerna Bryt dina filterbubblor och analysera innehåll på djupet med en prompt.....	177
Lektion 20: AI kan ha fel	180
Arbetsblad – Lektion 11: AI kan ha fel.....	184
Lektion 21: Hallucinationer hos AI – när verktygen hittar på	188
Arbetsblad – Lektion 21: Hallucinationer hos AI.....	193
Lektion 22: Syna Bias	196
Arbetsblad – Lektion 22: Syna Bias.....	199
Lektion 23: Bekräftelsebias (Confirmation Bias)	201
Arbetsblad Lektion 23 – Bekräftelsebias	206
Lektion 24: Auktoritetsbias, Förankringsbias och Tillgänglighetsbias.....	208
Arbetsblad Lektion 24 – Auktoritetsbias, Förankringsbias & Tillgänglighetsbias.....	213
Lektion 25: Halo-effekten (Gloriaeffekten)	218
Arbetsblad – Lektion 25: Halo-effekten (Gloriaeffekten).....	223
Lektion 26: Psykofantism.....	228
Arbetsblad – Lektion 26: Psykofantism När AI är en “ja-sägare” och hur det kan påverka vårt kritiska tänkande	231
Lektion 27: Kritiskt tänkande i praktiken – problemlösning med och utan AI.....	235
Arbetsblad – Lektion 27: Kritiskt tänkande i praktiken – problemlösning med och utan AI	240
Lektion 28: AI och komplexa problem – kritiskt tänkande på bred front.....	243
Arbetsblad – Lektion 28: AI och komplexa problem – kritiskt tänkande på bred front .	247
Lektion 29: Höja nivån på analysen med AI.....	250
Arbetsblad – Lektion 29: Höja nivån på analysen med AI Att använda AI som “skruvdragare” för storskaliga textanalyser och retoriska studier	253

AI i skolan: Både utmaning och möjlighet

AI-verktygen är här, de används redan av eleverna, och de förändrar spelplanen för undervisning om källkritik.

För bara några år sedan handlade källkritik främst om att kolla vem som stod bakom en text. Nu måste lärare lära ut hur man avslöjar information som ingen människa ens har skrivit.

Det blir allt svårare att skilja riktiga bilder från AI-skapade. I klassrum runt om i landet står elever och lärare inför samma utmaning – vad är äkta och vad är fejk?

Och det är inte bara bilder. AI kan skapa texter som låter helt trovärdiga men är fulla med påhittade fakta. Elever kan lämna in arbeten med citat från påhittade forskare och hänvisningar till händelser som aldrig ägt rum – allt serverat av en AI.

AI-systemen har dessutom sina egna fördomar. De speglar ofta problematiska mönster från internet där de hämtat sin information. När elever använder dessa verktyg för skolarbeten kan dessa fördomar föras vidare, ofta utan att någon märker det.

Men det finns hopp! Kreativa undervisningsmetoder där elever får granska AI-texter kritiskt visar sig vara effektiva. Ju mer eleverna vet om ett ämne, desto bättre blir de på att upptäcka när AI:n hittar på.

När elever får jämföra olika källor och förstå skillnaden mellan mänskligt och maskinellt tänkande, utvecklar de en djupare form av källkritik som är anpassad för vår tid.

Skolan står mitt i en förändring som är lika omvälvande som när internet kom. Lärare har ett gyllene läge att forma hur den här tekniken används. Om de gör det rätt kan dagens elever bli den första generationen som växer upp med AI som ett verktyg – inte en ersättare – för tänkande.

Disposition

Efter denna inledning följer ett kapitel med noga utvalda **resurser**, som länkar och verktyg, som lärare direkt kan använda i sin undervisning.

Nästa kapitel innehåller en lättförståelig **begreppslista** där viktiga termer förklaras kortfattat, så att även den som är ovan med AI lätt kan hänga med.

I det efterföljande kapitlet fördjupas varje begrepp ytterligare, med konkreta exempel och **pedagogiska förklaringar** av hur lärare kan undervisa eleverna om dessa begrepp:

- Algoritmer och rekommendationssystem
- Informationsnätverk och hur AI påverkar sociala medier

- Hyperindividualisering och filterbubblor
- Ekokammare
- Hyperpersonalisering och integritet (privacy)
- Hyperverklighet, desinformation och deepfakes
- Desinformation
- Hallucinationer hos AI
- Manipulationstekniker (cherry picking, känsloladdat språk, falska dikotomier)
- Prebunking
- Debunking
- Källtillit och sakkritik
- Bias inom AI
- Mänskligt bias i förhållande till AI (bekräftelsebias, auktoritetsbias, förankringsbias, tillgänglighetsbias, halo-effekten)
- Transparens och förklarbarhet (XAI)

Bokens avslutande del består av över 30 tydliga **lektioner** med lärarhandledningar, tips på vidare läsning och färdiga förslag på arbetsblad till eleverna. Dessa lektioner hjälper lärare att praktiskt omsätta teori till undervisning i klassrummet.

Resurser och verktyg för källkritik och AI

Webbresurser

Digital källkritik

 [Digitala lektioner – Digital källkritik](#)

Interaktiva lektioner och övningar om digital källkritik för skolan.

AI och utbildning

 [Pedagog Jönköping – AI](#)

Resurser och exempel på hur AI kan användas i undervisningen.

Förebygg desinformation

 [Google Prebunking](#)

Lär dig att identifiera och motverka falsk information genom prebunking-strategier.

Deepfake-träning

 [Leon Furze – Deepfake Game](#)

Ett spel där du övar på att identifiera deepfakes och manipulativa tekniker.

Mediebias och partiskhet

 [AllSides – Media Bias Chart](#)

En översikt över politisk bias i nyhetsmedier från olika perspektiv.

Källkritik i Sverige

 [Källkritikbyrån](#)

Oberoende initiativ för att granska och utbilda om källkritik.

Nyhetsvärdering i undervisningen

 [Nyhetsvärderaren](#)

Lektioner där elever tränar på att analysera nyhetskällor kritiskt.

Verktyg för verifiering av media

 [InVID Verification Plugin](#)

Ett verktyg för att analysera och verifiera bilder och videoklipp.

Faktaundersökning av rykten och myter

 [Snopes](#)

En av världens största faktaundersökningsplattformar för att avslöja falska nyheter.

SVT:s faktaundersökning

 [SVT Nyheter Verifierar](#)

SVT:s redaktion som granskar och verifierar virala nyheter och rykten.

FactCheck.org

 [FactCheck.org](#)

En icke-partisk, ideell webbplats som syftar till att minska vilseledning och förvirring i amerikansk politik genom att tillhandahålla originalforskning om desinformation och bluffar.

AI-assistenter för källkritik (Kräver ChatGPT-konto)

Källkritiksassistenten

 [Använd GPT för källkritik](#)

En specialanpassad AI-assistent som hjälper dig analysera och granska information.

Vidga vyer – ett sätt att skapa förståelse

 [Utforska olika perspektiv](#)

En GPT för att bredda perspektiv och förstå olika synsätt.

Digitala spel för att främja källkritiskt tänkande:

Bad News Game

 [Spela Bad News](#)

Ett spel där du tar rollen som en falsk nyhetsproducent för att förstå och motverka desinformation.

Cranky Uncle

 [Spela Cranky Uncle](#)

Ett spel som använder humor och kritiskt tänkande för att lära ut tekniker för att känna igen och motstå desinformation.

Go Viral!

Ett kort spel som på fem minuter lär dig att identifiera och motstå felaktig information relaterad till Covid-19.

 [Spela Go Viral!](#)

Fakey

Ett spel som simulerar ett sociala medier-flöde där du övar på att identifiera och

fakta granska tveksamma nyheter.

 [Spela Fakey](#)

Harmony Square

Ett spel där du använder desinformationstaktiker för att så split i en fiktiv by, vilket ökar din medvetenhet om manipulationstekniker.

 [Spela Harmony Square](#)

NewsFeed Defenders

Ett spel där du granskar och avfärdar desinformation genom att bedöma trovärdigheten i olika inlägg på ett forum.

 [Spela NewsFeed Defenders](#)

Fake It To Make It

Ett spel där du skapar fejkade nyheter och försöker få viral spridning för ekonomisk vinning, vilket ger insikt i desinformationsspridarnas metoder.

 [Spela Fake It To Make It](#)

Lateral läsning

STORM – Stanford's AI Research Tool

 [STORM](#)

Ett forskningsprototyp från Stanford som använder AI för att skapa Wikipedia-liknande rapporter om valda ämnen.
news.stanford.edu+2storm.genie.stanford.edu+2news.stanford.edu+2

Grok – xAI's AI-assistent

 [Grok](#)

En AI-driven assistent utvecklad av xAI,

designad för att ge sanningsenliga och användbara svar, inklusive realtidsökningar och bildgenerering. Använd funktionen "Deep Search"

Perplexity AI

 [Perplexity AI](#)

En AI-driven sökmotor som ger direkta svar på frågor genom att kombinera språkmodeller med realtidsdata.

Google Bilder

 [Google Bilder](#)

En tjänst från Google som låter användare söka efter bilder baserat på nyckelord eller genom omvänd bildsökning.

ChatHub.gg – få svar från flera AI-chattbotar samtidigt

 [ChatHub.gg](#)

En plattform som låter dig använda flera AI-chattbotar samtidigt, inklusive GPT-4o, Claude 3.5 och Gemini 2.0, för att jämföra deras svar sida vid sida

AI-verktyg

Chattbotar

ChatGPT

 [ChatGPT](#)

En avancerad språkmodell utvecklad av OpenAI, designad för att generera text och svara på frågor på ett mänskligt liknande sätt.

Claude.ai

 [Claude.ai](#)

En AI-assistent utvecklad av Anthropic, tränad för att vara säker, exakt och pålitlig för att hjälpa dig med dina arbetsuppgifter.

Venice.ai

 [Venice.ai](#)

En AI-plattform som erbjuder en integritetsfokuserad chattbot och bildgenerator, utan krav på kontoregistrering eller datalagring.

Grok

 [Grok](#)

En AI-driven assistent utvecklad av xAI, designad för att ge sanningsenliga och användbara svar, inklusive realtidsökningar och bildgenerering.

Gemini

 [Google Gemini](#)

Googles AI-assistent som integrerar avancerade språkmodeller för att ge användarna en förbättrad sök- och interaktionsupplevelse.

DuckDuckGo AI

 [DuckDuckGo AI](#)

En anonym chattbot integrerad i DuckDuckGo:s privata webbläsare, som låter användare interagera med AI-modeller utan att kompromettera deras integritet.

Skapa bilder:

Grok – AI-assistent med få begränsningar

 [Grok](#)

En AI-driven assistent utvecklad av xAI,

känd för sina minimala restriktioner vid bildgenerering, vilket möjliggör skapandet av kontroversiella och explicit innehåll.

Artguru.ai AI Text-to-Image Generator – kräver inte konto

 [Artguru AI Image Generator](#)

Ett verktyg som omvandlar text till bilder med olika stilar, inklusive illustrationer, porträtt och 3D-modeller, utan krav på kontoregistrering.

MagicStudio AI Art Generator – kräver inte konto

 [MagicStudio AI Art Generator](#)

Skapa korta klipp

Kling AI

 [Kling AI](#)

En AI-driven plattform som omvandlar textbeskrivningar till högkvalitativa videor, känd för sin snabba bearbetningstid och realistiska rörelser.

MiniMax

 [MiniMax](#)

Ett kinesiskt AI-företag som erbjuder flera produkter, inklusive 'Hailuo AI', en multimodal plattform för text- och musikgenerering, samt 'Video-01', en text-till-video-modell som genererar 6-sekunders videor från textpromptar.

Klona röster – skapa ljud

Suno.com

 [Suno](#)

En AI-driven plattform som låter användare skapa musik genom att spela in eller ladda upp ljud och omvandla det till sånger. Användare kan också få tillgång till separata ljudspår för full kreativ kontroll och ansluta sig till en blomstrande musikkultur för inspiration.

En AI-driven konstgenerator som snabbt skapar konstverk baserat på textbeskrivningar, tillgänglig utan inloggning.

Ideogram.ai – bra promptförståelse

 [Ideogram](#)

En AI-modell som genererar bilder från textbeskrivningar och utmärker sig genom sin förmåga att skapa läsbar text inom bilderna.

InVideo.ai

 [InVideo](#)

En AI-driven videoredigeringsplattform som möjliggör snabb och enkel skapelse av professionella videor med hjälp av mallar och automatiserade funktioner.

Luma Labs Dream Machine

 [Luma Labs Dream Machine](#)

En text-till-video-modell som genererar videor baserade på användarens promptar eller stillbilder, känd för sin förmåga att realistiskt fånga rörelser och skapa cinematisk kvalitet.

Udio.com

 [Udio](#)

En AI-musikgenerator som gör det möjligt för användare att upptäcka, skapa och dela musik med världen. Genom att använda den senaste tekniken kan användare skapa AI-genererad musik på några sekunder.

Elevenlabs.com

 [ElevenLabs](#)

En AI-baserad plattform för text-till-tal och röstgenerering som kan leverera högkvalitativ, människoliknande tal på 32 språk. Perfekt för ljudböcker, videoröstovers, reklam och mer.

Play.ht

 [Play.ht](#)

En AI-röstgenerator som erbjuder över 900 AI-röster på mer än 100 språk och accenter. Användare kan skapa realistiska röstovers för videor, podcaster och presentationer med hjälp av avancerade funktioner som anpassade uttal och röstinflektioner.

Skapa deepfakes

Studio D-ID

 [Studio D-ID](#)

En AI-driven plattform som låter användare skapa videor med digitala personer från en enda bild. Användare kan skapa engagerande videor med talande avatarer som kommer till liv på några sekunder.

play.google.com+1apps.apple.com+1

Deepfake Sandbox

 [Deepfake Sandbox](#)

En interaktiv demo som låter dig skapa en deepfake av dig själv på några sekunder. Efter att bilden har genererats, undersöks de samhällseliga implikationerna av deepfakes och annan AI-teknik.

HeyGen

 [HeyGen](#)

En AI-driven videoplattform som gör det möjligt för användare att skapa högkvalitativa videor med AI-avatarer. Användare kan skapa professionella AI-drivna avatarvideor, översätta innehåll och förverkliga idéer utan behov av en kamera eller team. heygen.com

Live Portrait AI

 [Live Portrait AI](#)

Ett verktyg som förvandlar statiska bilder till animerade videor med realistiska ansiktsuttryck och rörelser. Användare kan skapa engagerande videor genom att animera sina foton med hjälp av AI.

Begreppslista om AI och källkritik

Några av de begrepp som tas upp i detta material är:

Algoritmer och rekommendationssystem

Algoritmer är regler och beräkningar som datorer använder för att lösa problem. Rekommendationssystem är algoritmer som väljer ut innehåll åt oss, t.ex. vilka videor vi ser på YouTube eller vilka nyheter vi får i vårt flöde. De kan skapa skraddarsydda upplevelser men också riskera att begränsa vår världsbild.

Bias inom AI

Bias betyder snedvridning eller partiskhet, och AI kan ärva bias från data den tränas på. Det kan leda till orättvisa eller felaktiga beslut, t.ex. att en AI diskriminerar vissa grupper. Därför är det viktigt att granska hur AI fattar beslut och varifrån den får sin information.

Debunking

Debunking handlar om att avslöja falsk eller vilseledande information efter att den spridits. Det kan göras genom faktagranskning och att presentera bevis som visar varför något inte stämmer. Det är en viktig metod för att bekämpa desinformation.

Desinformation

Desinformation är falsk information som sprids medvetet för att lura eller påverka människor. Det kan vara falska nyheter, manipulerade bilder eller vilseledande berättelser. Desinformation används ofta i propaganda eller för att skapa osäkerhet.

Ekokammare

En ekokammare är när vi bara får information och åsikter som bekräftar det vi redan tror på. Det kan hända på sociala medier när algoritmer visar oss innehåll vi gillar, men undanhåller annat. Detta kan förstärka våra åsikter och minska vår förståelse för andra perspektiv.

Hallucinationer hos AI

Hallucinationer hos AI är när en AI hittar på fakta eller skapar felaktig information.

Det kan vara påhittade källor, osanna påståenden eller felaktiga översättningar. Eftersom AI inte "vet" något utan bara gissar baserat på data, kan detta hända ofta.

Hyperindividualisering och filterbubblor

Hyperindividualisering innebär att information anpassas extremt mycket till oss som individer, ofta genom AI. Filterbubblor skapas när vi bara ser information som passar våra intressen och åsikter, vilket kan göra det svårare att få en bred bild av verkligheten.

Hyperpersonalisering och integritet (privacy)

Hyperpersonalisering är när AI använder mycket personlig data för att anpassa innehåll, t.ex. annonser eller nyheter. Det kan vara praktiskt men också ett hot mot vår integritet. Ju mer information som samlas in om oss, desto större blir risken att den missbrukas.

Hyperverklighet, desinformation och deepfakes

Hyperverklighet innebär att det blir svårt att skilja mellan verklighet och digitalt manipulerat innehåll. Deepfakes är AI-genererade bilder eller videor där ansikten eller röster ändrats för att se verkliga ut. Dessa används ofta i desinformation för att lura människor.

Källtillit och sakkritik

Källtillit handlar om att avgöra vilka källor som är trovärdiga. Sakkritik innebär att undersöka om informationen är korrekt, genom att granska fakta och expertkällor. Att tänka kritiskt är viktigt för att inte luras av falsk information.

Manipulationstekniker (cherry picking, känsloladdat språk, falska dikotomier)

Manipulationstekniker används för att påverka hur vi tänker och känner. **Cherry picking** innebär att välja ut fakta som

stödjer en viss åsikt och ignorera resten. **Känsloladdat språk** spelar på våra känslor istället för att använda fakta. **Falska dikotomier** skapar en falsk bild av att det bara finns två val, fast det egentligen finns fler alternativ.

Mänskligt bias i förhållande till AI

Människor har naturliga snedvridningar i sitt tänkande som också påverkar hur vi använder och tolkar AI. Några vanliga bias är:

- **Bekräftelsebias** – Vi söker och tror mer på information som bekräftar det vi redan tycker.
- **Auktoritetsbias** – Vi litar mer på information från personer eller källor vi uppfattar som auktoriteter, även om de har fel.
- **Förankringsbias** – Vi påverkas av den första information vi får och har svårt att ändra uppfattning senare.

- **Tillgänglighetsbias** – Vi tror att något är vanligare eller sannare bara för att vi lätt kan komma på exempel på det.
- **Halo-effekten** – Vi antar att om någon är bra på en sak så är de bra på andra saker också, även om det inte finns bevis för det.

Prebunking

Prebunking är att lära sig känna igen och ifrågasätta vilseledande information innan vi stöter på den. Genom att förstå hur manipulation fungerar kan vi bli mer motståndskraftiga mot desinformation. Det är som ett vaccin mot falska nyheter.

Transparens och förklarbarhet (XAI)

XAI (Explainable AI) handlar om att göra AI:s beslut begripliga för människor. När AI används för viktiga beslut, t.ex. i sjukvård eller juridik, måste vi förstå hur den kom fram till sitt svar. Transparens är avgörande för att vi ska kunna lita på AI.

Generativ AI och digitala informationssystem – utmaningar för källkritik och kritiskt tänkande

Den digitala informationsfloran förändras snabbt av **generativ AI** och smarta algoritmer. För gymnasielärare – som ska rusta nästa generation i källkritik och kritiskt tänkande – innebär detta både nya möjligheter och betydande utmaningar. I denna populärvetenskapliga översikt förklaras hur algoritmer, sociala medienätverk och AI påverkar informationsflödet. Vi problematiserar fenomen som filterbubblor, ekokammare, hyperpersonalisering och **deepfakes**, och belyser deras effekter på samhället. Texten ger konkreta exempel, fallstudier och hänvisar till aktuell forskning, men kräver inga förkunskaper.

Algoritmer och rekommendationssystem

Algoritmer är i grunden instruktioner eller regler – som recept – som datorer följer för att lösa en uppgift. I sociala medier och digitala plattformar används algoritmer för att automatiskt sortera och välja ut vilket innehåll som presenteras för varje användare.

Rekommendationssystem är en typ av algoritm som lär sig av våra beteenden (vad vi klickar på, gillar eller tittar länge på) för att förutsäga vad vi vill se härnäst. Syftet kan vara att hålla oss engagerade så länge som möjligt på plattformen. Ett vardagligt exempel är hur Netflix rekommenderar filmer baserat på vad du sett tidigare, eller hur Spotify föreslår musik liknande den du lyssnat på. På nyhetssajter och sociala medier innebär detta att två personer kan få helt olika flöden av information beroende på deras tidigare aktivitet.

Effekten av dessa algoritmer är enorm. På YouTube, till exempel, uppskattas över 70% av videorna som folk tittar på komma från sidans egna rekommendationer, inte från att man själv sökt upp videon ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)). Med andra ord styr YouTubes algoritm majoriteten av användarnas tittande – den väljer vilken video som *automatiskt* spelas upp härnäst eller dyker upp i sidokolumnen som förslag. Dessa val baseras på användarens profil och tidigare beteende. Algoritmerna är optimerade för att öka vår **engagemang** (t.ex. klick, visningar, tid spenderad), vilket i sin tur gynnar plattformens annonsintäkter.

Men vad händer när engagemang blir ledstjärnan? Forskning och avslöjanden tyder på att algoritmerna då kan komma att premiera sensationellt eller polariserande innehåll. Facebooks egen forskning visade att inlägg som väckte *starka känslor*, särskilt ilska, genererade mest reaktioner och kommentarer – och därmed gavs störst synlighet i flödet ([Facebook whistleblower Frances Haugen points to 2018 algorithm overhaul as key problem](#)). Efter en större förändring av Facebooks algoritm 2018 märkte politiska partier att de behövde posta mer arga och polariserande budskap för att nå ut till sina följare ([Facebook whistleblower Frances Haugen points to 2018 algorithm overhaul as key problem](#)). Detta beror på att algoritmen hade börjat belöna inlägg med många reaktioner (exempelvis arga emojis) med högre prioritet i användarnas nyhetsflöden. Konsekvensen blev att mer extremt eller känsloladdat innehåll fick extra spridning, vilket riskerar att driva upp tonläget online.

Ett annat uppmärksammat exempel är **YouTubes rekommendationsalgoritm**. En aktuell studie i den vetenskapliga tidskriften *PNAS* (2023) genomförde ett storskaligt test av YouTubes video-rekommendationer. Forskarna skapade över 100 000 automatiserade konton ("sock puppets") som fick titta på olika typer av politiska videor. Resultatet visade att YouTubes algoritm främst rekommenderade fler videor som liknade dem användaren redan sett. För en användare med **högerlutande** preferenser ledde detta in i en "kaninhålseffekt" där algoritmen snart bjöd på videor från kanaler fulla av politisk extremism och konspirationsteorier ([YouTube Video Recommendations Lead to More Extremist Content for](#)

[Right-Leaning Users, Researchers Suggest | UC Davis](#)) ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)).

Vänsterlutande användare mötte inte lika många ytterlighets-videor, men även de fick främst rekommendationer som bekräftade deras ideologiska utgångspunkt ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)) ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)). Studien understryker hur algoritmer kan skapa *skeva informationsflöden* – inte genom total censur av motsatta åsikter, men genom att dränka användaren i innehåll som bekräftar det den redan tror på.

Kort sagt: algoritmer och rekommendationssystem filtrerar internet åt oss. De avgör till stor del vilken nyhet vi ser på Facebook, vilken video som spelas på YouTube och vilka inlägg som trendar på Twitter. Detta sker ofta osynligt i bakgrunden – algoritmerna är *icke-transparenta*, vi vet sällan exakt **varför** vi ser ett visst inlägg. För den som undervisar i källkritik är det centralt att eleverna förstår att deras flöde är personligt anpassat av en maskin. Det vi ser online är *inte* en objektiv spegling av omvärlden, utan resultatet av en mängd dolda beräkningar.

Informationsnätverk och AI:s påverkan på sociala medier

Internet och sociala medier utgör komplexa **informationsnätverk** där nyheter, rykten och idéer kan spridas blixtnabbt. **AI** spelar en roll inte bara i vad *du* ser, utan även i hur information sprids mellan människor. På plattformar som Facebook, Twitter, Instagram eller TikTok avgör algoritmer vilka inlägg som ska “trenda” eller bli virala. När många användare börjar gilla eller dela ett inlägg kan AI:n reagera genom att visa det för ännu fler – en snöbollseffekt skapas. Detta kan vara positivt när bra information sprids, men det kan lika gärna gälla felaktig eller skadlig information.

En viktig insikt är att dagens sociala medier belönar sådant som väcker reaktioner. AI-system lär sig snabbt vilka ämnen som får människor att stanna upp, klicka, dela eller kommentera – och ger sådant innehåll högre spridning. Ibland innebär det att **provocerande eller extremt** innehåll får oproportionerligt genomslag. Ett konkret exempel kommer från plattformen TikTok. En undersökning fann att TikToks algoritm på bara fem dagar dramatiskt kunde öka mängden sexistiskt innehåll som visades för en användare – forskarna såg en fyrfaldig ökning av misogynna videor som algoritmen föreslog under denna korta period ([Social media algorithms 'amplifying misogynistic content'](#)). Detta tyder på att om en användare interagerar med t.ex. ett kontroversiellt klipp, kan AI:n snabbt förstärka den trenden genom att mata fram mer av liknande slag. Informationsflödet *accelererar* i en viss riktning.

Ett annat fenomen är användningen av **botar** – automatiserade konton ofta drivna av AI – i sociala nätverk. Botar kan agera som om de vore människor: gilla, kommentera och sprida inlägg. Vid politiska val eller under pandemier har mängder av bot-konton använts för att boosta vissa narrativ och få dem att framstå som mer populära eller eniga än de egentligen är. Forskning visar att sådana botar bidrar till spridning av desinformation och kan manipulera den allmänna diskussionen online ([Echo chamber effects on short video platforms - PMC](#)). Exempelvis upptäcktes nätverk av botar på Twitter som spred falska påståenden om COVID-19 och konspirationsteorier, vilket lurade många riktiga användare att tro att “alla” pratade om

dessa saker. Informationsnätverket på sociala medier kan alltså **förvrängas** av AI-drivna agenter som förstärker vissa budskap artificiellt.

En annan aspekt är hur **viral spridningsmönster** kan uppstå. Forskare vid MIT analyserade 126 000 nyhetsinlägg på Twitter och fann att falska nyheter spreds “längre, snabbare, djupare och bredare” än sanna nyheter ([Study: On Twitter, false news travels faster than true stories | MIT News | Massachusetts Institute of Technology](#)). Förklaringen var inte enbart robotar, utan att människor själva oftare delade chockerande eller felaktiga nyheter – kanske för att de väckte mer känslor ([Study: On Twitter, false news travels faster than true stories | MIT News | Massachusetts Institute of Technology](#)). Faktiskt var en falsk nyhet cirka **70%** mer sannolik att bli retweetad än en sann nyhet ([Study: On Twitter, false news travels faster than true stories | MIT News | Massachusetts Institute of Technology](#)). Detta illustrerar en större princip i informationsnätverken: sådant som överraskar eller upprör får ofta mer uppmärksamhet. AI-systemen som styr våra flöden snappar upp dessa mänskliga preferenser och kan oavsiktligt **förstärka** spridningen av tvivelaktigt innehåll.

Konsekvenserna av AI-drivna informationsnätverk märks i samhällsdebatten. Å ena sidan har vi aldrig haft större tillgång till information och möjligheter att höra olika röster. Å andra sidan riskerar den massiva, okontrollerade spridningen av innehåll – styrd av engagemangsalgoritmer – att dränka välgrundade fakta i ett brus av rykten och halvsanningar. För den kritiske tänkaren innebär det att man måste förstå *hur* och *varför* en viss nyhet dyker upp i ens flöde, och vara medveten om att det kan finnas en teknologisk bias i spridningen.

Hyperindividualisering och filterbubblor

En av de mest omtalade effekterna av algoritmisk kuratering är skapandet av **filterbubblor**. Begreppet myntades kring 2010 av internetaktivisten Eli Pariser och syftar på att individanpassade flöden kan isolera oss i vår egen informationsbubbla. Tjänster som Google och Facebook anpassar resultat och nyheter efter vad de tror att *just du* vill se, baserat på data om dig – exempelvis din sökhistorik, vilka länkar du klickat på tidigare, var du befinner dig osv. ([Filter bubble - Wikipedia](#)). Det betyder att två personer som söker efter samma sak online kan få helt olika resultat. Pariser gav ett talande exempel: när två personer googlade "BP" (oljebolaget) fick den ena bara upp investeringsnyheter om företaget, medan den andra fick nyheter om miljökatastrofen Deepwater Horizon – trots att sökorden var identiska ([Filter bubble - Wikipedia](#)). Algoritmen hade “gissat” vad respektive person ville höra om BP och filtrerat resultaten därefter. Den ena personen blev ovetande om miljöskandalen, den andra såg inget om företagsaktien.

En **filterbubbla** innebär alltså ett tillstånd av intellektuell isolering: man exponeras främst för information som bekräftar ens egna åsikter eller intressen, medan avvikande perspektiv filtreras bort ([Filter bubble - Wikipedia](#)). Över tid kan detta ge en snedvriden världsbild – man får intrycket att “alla” tycker som man själv, eftersom man sällan ser motargumenten. Förmågan till kritiskt tänkande kan utmanas, då kritiskt tänkande delvis bygger på att väga olika källor och perspektiv mot varandra. Om eleven bara matas med en enda sorts vinkel blir det svårare att ens *upptäcka* behovet av källkritik.

Dock ska sägas att fenomenet filterbubblor har diskuterats flitigt inom forskningen, och bilden är inte entydig. Vissa studier har ifrågasatt hur stor effekten faktiskt är. En del forskare menar att användare trots algoritmerna ändå tar del av varierande innehåll och att människor aktivt

söker upp olika nyhetskällor på nätet, vilket skulle begränsa bubbla-effekten. Rapporter från exempelvis Reuters Institute har hävdats att den genomsnittliga internetanvändaren inte nödvändigtvis sitter helt fast i en bubbla, och att effekten kan vara *överdriven* i debatten ([Filter bubble - Wikipedia](#)). Samtidigt visar andra studier att filterbubblor *existerar* och har påtagliga konsekvenser för vissa grupper av användare ([Echo chamber effects on short video platforms - PMC](#)). Klart är att personalisering och hyperindividualisering åtminstone *möjliggör* en filtrering av verkligheten som saknar historiskt motstycke.

För en lärare är det viktigt att illustrera filterbubblor med konkreta exempel som eleverna kan relatera till. Man kan till exempel be dem jämföra sina flöden på sociala medier eller nyhetsappar och notera skillnaderna. Varför får elev A upp massor av nyheter om klimatkrisen medan elev B mest ser sport och ekonomi? Troligen för att appens algoritm lärt sig deras olika intressen. Detta i sig är inte ont – personalisering kan ju göra flödet mer relevant – men eleverna bör vara vaksamma på att de *också* behöver aktivt söka upp andra perspektiv. Källkritik i en filterbubblornas tid handlar om att våga sticka hål på sin egen bubbla och titta utanför den information som bekvämt bekräftar det man redan vet eller tycker.

([File:Filter bubble illustration.png - Wikipedia](#)) *Illustration av konceptet "filterbubbla": Algoritmer kan anpassa nyhetsflödet utifrån vad de tror att användaren vill höra. Med tiden bildas separata bubblor av information – i detta exempel en bubbla för en högerlutande, pro-business användare och en för en vänsterlutande, miljöengagerad användare. Båda får nyheter som passar deras profil, men går miste om sådant som utmanar deras perspektiv* ([File:Filter bubble illustration.png - Wikipedia](#)).

Det pedagogiska uppdraget här blir att öka medvetenheten. Eleverna kan uppmuntras att ställa frågor som: *Hur vet jag att jag inte missar viktiga fakta?*, *Vad visas inte för mig?* och *Hur skulle någon med motsatt uppfattning beskriva samma händelse?*. Genom att förstå hyperindividualiseringen av informationsflöden kan de börja aktivt söka mångfald i källorna – en form av motmedel mot filterbubblans begränsningar.

Ekokammare

Nära besläktat med filterbubblor är fenomenet **ekokammare**. En ekokammare uppstår när människor främst kommunicerar med andra som redan delar samma åsikter, varvid deras uppfattningar förstärks likt ett eko som studsar mellan väggarna. Till skillnad från filterbubblor, som mest handlar om teknologins personalisering, betonar ekokammare den *sociala* dimensionen – våra egna val av vilka vi följer, vilka forum vi besöker och vilka vi litar på. På nätet tenderar många att dras till likasinnade grupper, vare sig det gäller politik, hobbyer eller världsbilder. Algoritmerna kan spä på detta (de föreslår att vi går med i grupper "du kanske gillar", som ofta består av folk med liknande åsikter). Resultatet kan bli isolerade diskussionsmiljöer där alla nickar instämmande.

I en ekokammare cirkulerar information och idéer utan att ifrågasättas av utomstående. Till exempel kan en sluten Facebook-grupp som tror på en viss konspirationsteori dagligen dela artiklar och "bevis" för sin sak, allt medan kritiska röster lyser med sin frånvaro (de kanske har lämnat gruppen eller blir blockade). Gruppen hör i princip bara sitt eget eko. Sådana miljöer riskerar att driva medlemmarna mot allt mer **extrema ståndpunkter** – en process kallad *grupppolarisering*. Om ingen i ens flöde invänder när felaktiga påståenden görs, är det

lätt att kollektivet blir övertygat om att påståendena är sanna. Studier har visat att ekokammare på nätet kan bidra till spridning av vilseledande information, fake news och rykten ([Echo chamber effects on short video platforms - PMC](#)). Alltså, inte nog med att medlemmarna i en ekokammare får en ensidig bild – de kan också gemensamt driva varandra längre bort från verifierade fakta.

Effekten av ekokammare märks tydligt i det politiska landskapet. I USA pekar forskare på att partianhängare online ofta rör sig i separata medie-ekosystem – liberaler och konservativa har sina egna nyhetssajter, kommentarsfält och Twitter-sfärer, där de mest interagerar med de sina. En färsk litteraturoversikt (2021) av 55 vetenskapliga studier om ekokammare fann dock att förekomsten av ekokammare varierar och beror på hur man mäter ([Echo Chambers on Social Media: A Systematic Review of the Literature | Review of Communication Research](#)). Intressant nog tenderade forskningsrapporter som analyserat *faktiskt digitalt beteende* (t.ex. data över vilka länkar folk verkligen klickar på) att **hitta** tydliga ekokammare, medan studier baserade på *självrappporter* (enkäter där folk uppger vad de läser) ofta **inte** fann samma effekt ([Echo Chambers on Social Media: A Systematic Review of the Literature | Review of Communication Research](#)). Detta tyder på att många kanske *tror* att de tar del av blandade perspektiv, medan data från deras faktiska klickmönster visar en mer ensidig diet av information. Forskarna betonar att ekokammare är ett komplext fenomen – de existerar, men inte för alla och inte alltid i samma grad. Mycket beror på individens medievanor och på plattformarnas utformning.

För demokratin och samhällsdiskussionen utgör ekokammare en utmaning. Om olika grupper i samhället lever i sina egna informationsvärldar kan det bli svårt att ens enas om gemensamma fakta i en debatt. Allmänheten kan fragmenteras i separata ”verkligheter”. Vi såg tecken på detta under exempelvis Brexit-omröstningen och det amerikanska valet 2016, där olika väljargrupper verkade leva i helt olika nyhetsuniversum. Ekokammare kan också ge ökat **misstroende** mot andra grupper – man får en karikatyr av ”de andra” eftersom man aldrig hör deras rimliga argument, bara ens egna grupps version av motståndarsidans argument (ofta i förenklad eller hånfull form).

Hur kan man då motverka ekokammare i klassrummet? Ett steg är att uppmuntra eleverna att *aktivt söka upp* motpolar. Till exempel kan en övning vara att låta eleverna läsa om en kontroversiell fråga från två olika nyhetskällor med tydlig ideologisk slagsida åt olika håll. Sedan får de jämföra hur narrativet skiljer sig. En annan strategi är att lyfta källor som är kända för att försöka vara balanserade eller faktagranskande, och diskutera deras roll. Viktigt är också att poängtera att det **inte** handlar om att eleverna ska ändra åsikt i sak – utan om att de ska vara medvetna om att deras informationsmiljö kan vara begränsad och att sanningen ofta låter olika ur olika munnar. Genom att förstå ekokammarens mekanismer kan eleverna bättre identifiera när de själva befinner sig i ett ”eko” och anstränga sig för att ta del av utsidan.

Hyperpersonalisering och integritet

De algoritmer som skapar personalisering och individanpassning är hungriga på en råvara: **data** om oss. Ju mer plattformarna vet om användaren, desto mer **hyperpersonalisering** kan de åstadkomma – alltså en extremt skraddarsydd upplevelse. Idag samlas ofantliga mängder data in om varje användares beteende: vilka sajter vi besöker, vilka inlägg vi stannar upp vid, vad vi kommenterar, vilka platser vi tar bilder på, hur länge vi tittar på en video, osv. AI kan

korsa och analysera dessa datapunkter för att bilda sig en förvånansvärt detaljerad bild av vår personlighet och våra liv. Detta väcker frågor om **personlig integritet**: vet vi ens hur mycket information vi lämnar ifrån oss, och vem som använder den – och till vad?

Ett känt (och lite kusligt) exempel på dataanalysens kraft är fallet med den amerikanska butikskedjan Target. Redan 2012 visade Target hur man med algoritmer kunde analysera kunders inköpsmönster för att förutsäga livshändelser. I ett fall lyckades de räkna ut att en tonårsflicka var gravid baserat på att hon köpt en viss kombination av produkter – Target skickade henne reklam för babyprodukter innan hon ens berättat för sin familj att hon väntade barn ([Facebook, private traits and attributes: Predictions from digital records of human behavior - The Journalist's Resource](#)). Den här historien illustrerar hur företag kan veta *mer* om oss än våra närstående gör, genom att sätta samman pusselbitar av data.

På nätet sker liknande saker hela tiden. Varje "gilla-markering", varje sökning och varje pausad video kan avslöja lite om dina intressen och egenskaper. Forskare vid Cambridge University visade 2013 hur **Facebook-likes** (alltså sidor eller inlägg en person gillat) kunde användas för att förutsäga personens mycket *privata* attribut med hög träffsäkerhet ([Facebook, private traits and attributes: Predictions from digital records of human behavior - The Journalist's Resource](#)). Deras algoritm kunde, enbart utifrån en persons gilla-markeringar, med över 80% korrekthet avgöra exempelvis personens sexuella läggning, politiska ideologi, religiösa tro, intelligensnivå och till och med om föräldrarna hade skilt sig under personens barndom ([Facebook, private traits and attributes: Predictions from digital records of human behavior - The Journalist's Resource](#)). Tänk dig det – en dator som kan gissa sådana saker utan att man själv någonsin angett det direkt, bara genom att lära sig mönster i ens beteende online. Denna typ av **profilering** på individnivå används redan kommersiellt: annonsörer kan rikta skräddarsydd reklam till dig baserat på antaganden om din personlighet och dina behov.

Problemet ur ett integritetsperspektiv är att denna datainsamling ofta sker utan att vi fullt ut förstår eller samtycker till det. Visst, vi klickar i användarvillkor och accepterar kakor, men hur många förstår att en enkel väderapp kan samla in vår geografiska rörelsemönster, som sedan säljs vidare? Eller att vårt beteende på en "oskyldig" app kan korskopplas med Facebook-data för att bygga en komplett profil av våra vanor? **Hyperpersonalisering** innebär att systemet *känner oss* – ibland bättre än vi känner oss själva.

Ett omskrivet fall som fick världen att vakna till integritetsfrågorna var **Cambridge Analytica-skandalen** 2018. Cambridge Analytica, ett politiskt konsultföretag, lyckades få tag på persondata från upp till 87 miljoner Facebook-användare utan deras medgivande ([Facebook–Cambridge Analytica data scandal - Wikipedia](#)). Genom en till synes harmlös quiz-app på Facebook siphonade de användarprofiler (inklusive information om användarnas vänners profiler). Därefter använde de dessa data för att skapa psykologiska profiler och riktade politiska budskap till väljarna inför bland annat det amerikanska presidentvalet 2016 ([Facebook–Cambridge Analytica data scandal - Wikipedia](#)). Målet var att hitta individers ömma punkter och övertyga dem med skräddarsydd propaganda – ett extremt exempel på hur personlig information kan utnyttjas. Skandalen ledde till global debatt och skärpt lagstiftning (som EU:s GDPR). För oss i skolan är lärdomen att personlig integritet och källkritik hänger ihop: om vi inte vet hur våra data används, kan vi heller inte förstå varför vi ser viss information.

Hyperpersonalisering rymmer en paradox. Å ena sidan gillar vi när tjänster är anpassade efter oss – det *kan* ge bättre användarupplevelser. Å andra sidan riskerar vi att offra vår integritet

och kanske bli manipulerade när anpassningen går för långt. Eleverna behöver bli medvetna konsumenterna av digitala medier: Varje gång de delar med sig av information om sig själva, ger de potentiellt mer makt åt algoritmerna att påverka dem. Att diskutera integritet kan ibland kännas abstrakt, men konkreta case (som de ovan) väcker ofta engagemang hos ungdomar. Vem vill bli förutsagd eller styrd utan att veta om det? Genom att förstå hur hyperpersonalisering fungerar kan eleverna bättre värdera **risker kontra nytta** med digitala tjänster och kanske efterfråga mer transparens och kontroll över sina data.

Hyperverklighet, desinformation och deepfakes

I takt med att generativ AI förbättras suddas gränsen ut mellan det genuina och det artificiella – vi träder in i en era av **hyperverklighet** där det som presenteras för oss ser verkligt ut, men kanske är helt skapat av en dator. Detta ställer källkritiken inför nya prövningar.

Desinformation – avsiktligt falsk eller vilseledande information – är inget nytt, men med AI kan den fabriceras och spridas i en skala och realism som saknar motstycke. Ett aktuellt hot är så kallade **deepfakes**: med hjälp av djupa neurala nätverk (därav "deep") kan AI skapa falska videoklipp eller ljudinspelningar där verkliga personer ser ut att säga eller göra sådant de aldrig gjort.

Ett exempel som redan blivit historiskt inträffade under Rysslands invasion av Ukraina 2022. Plötsligt började ett videoklipp spridas på nätet som visade Ukrainas president Volodymyr Zelenskyj hålla tal till sitt folk – men något var fel. I klippet uppmanade Zelenskyj ukrainska soldater att **lägga ner vapnen och kapitulera** för Ryssland. Videon såg vid första anblicken äkta ut och fick över 100 000 visningar på Twitter ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)) ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)). Men Ukrainas regering agerade snabbt och avslöjade att det rörde sig om en **deepfake** skapad av fienden. Hackare hade lagt en syntetisk bild av Zelenskyjs ansikte och röst över en annan person för att få det att se ut som om presidenten höll ett kapitulationstal ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)). Inom kort gick den riktige Zelenskyj ut och dementerade budskapet, och sociala medie-plattformar tog ner videon. Ukrainas försvar varnade allmänheten att fler manipulerade videor kunde dyka upp och påpekade att "videor skapade med sådan teknologi är nästintill omöjliga att skilja från verkliga" ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)). Syftet med deepfaken var uppenbar: att sprida misstro, förvirring och panik. Lyckligtvis var just denna deepfake tekniskt bristfällig nog att genomskådas relativt fort, men det visade världen vilken ny typ av propagandaverktyg AI har blivit.

Och det är inte bara i krigstid dessa metoder förekommer. Även i fredligare sammanhang har deepfakes och AI-genererat innehåll börjat dyka upp. Ett uppmärksammat fall var **"Balenciaga-påven"** våren 2023. På sociala medier spreds ett foto som föreställde påve Franciskus iförd en trendig vit dunjacka i lyxmärket Balenciagas stil – en bild lika underhållande som överraskande. Miljontals människor såg bilden, många utan att ana att den var falsk. I själva verket var fotot helt AI-genererat, skapat med hjälp av bildgeneratoren Midjourney ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)) ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)). ** ([File:Pope Francis in puffy winter jacket.jpg - Wikimedia Commons](#))** *AI-genererad bild på påve Franciskus i en fashionabel dunjacka. Bilden skapades med programmet Midjourney och spreds viralt i mars 2023. Många trodde initialt*

att fotot var äkta, vilket illustrerar hur svårt det har blivit att skilja fabrikation från verklighet. Trots vissa små detaljer som skvallrade om oäktheten (t.ex. orimligt jämn belysning och otydliga fingrar om man tittade noga) så lyckades bilden vilseleda många användare ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)). Kommentarer som “Är det här på riktigt?” och “Jag kan inte avgöra om det är en meme eller äkta” florerade ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)). Först när erfarna ögon och faktagranskare ingrep stod det klart att påven aldrig burit en sådan jacka ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)).

Detta exempel må vara harmlöst – ingen skadas av en påve i trendjacka – men tänk implikationerna. Om en AI kan skapa *fotografier* som lurar ögat, kan den också skapa bilder av mer allvarlig natur: “bevis” för händelser som aldrig ägt rum, eller komprometterande bilder av offentliga personer i fel situation. Redan har vi sett deepfake-videor där kända skådespelares ansikten klistrats in i filmer de aldrig medverkat i, eller där företagsledare “annonsera” saker de aldrig sagt. Möjligheten till **karaktärsdåd** och bluffar ökar. Vi går in i en tid där *seeing is not believing* – att se något på video eller foto kan inte längre tas som definitiv sanning.

Även **textinnehåll** påverkas. Med språkmodeller som ChatGPT kan vem som helst generera övertygande artiklar, inlägg eller “nyheter” på några sekunder. Detta kan användas kreativt och positivt, men i fel händer innebär det att fabriker av desinformation kan automatisera skapandet av falska nyhetssajter, spamkommentarer eller propaganda-poster i sociala medier. Tidigare krävdes det ändå mänsklig möda för att skriva en trovärdig lögn; nu kan en dator spotta ur sig tusentals varierade falska berättelser över en natt. Det gör det svårare för faktagranskare att hänga med och ökar risken att falska narrativ får fotfäste innan de hinner motbevisas.

Allt detta leder till en **erosion av tillit**. Om människor gång på gång exponeras för avslöjade bluffar – som deepfakes och fejkade bilder – kan de börja tvivla även på legitima nyheter och äkta material. Vi riskerar ett “post-sanning”-klimat där allting avfärdas med “det där kan ju vara fejk”. Källkritiken blir därför ännu mer central: vi måste lära oss nya färdigheter för att analysera digitala medier. Det kan handla om att titta efter subtila tecken på manipulation i bilder (till exempel onaturliga skuggor, konstiga detaljer på händer och öron som AI ofta har svårt med), att verifiera videos äkthet genom källor (finns originalklippet publicerat av en trovärdig källa?) eller att använda verktyg för att spåra om en bild redan funnits tidigare (reverse image search för att se om fotot är nytt eller en gammal bild som återanvänts i fel kontext).

Samtidigt pågår forskningen för att möta utmaningen – exempelvis utvecklas AI-verktyg som själva kan upptäcka deepfakes genom att analysera pixlar eller ljudvågor efter anomalier. Men detta är ett kapprustningens fält: när detekteringstekniken blir bättre, blir även förfalskningarna det.

För gymnasielärare gäller det att förbereda eleverna på att alltid ha ett skeptiskt öga. En bra tumregel är “*extraordinära påståenden kräver extraordinära bevis*”: om man ser något chockerande online (särskilt om det passar ens egna fördomar lite *för* väl), bör man stanna upp och dubbelkolla. Vem är källan? Rapporterar etablerade medier detta? Finns det andra oberoende vittnesmål? I en tid av hyperverklighet måste vi helt enkelt höja ribban för vad vi kräver innan vi litar på en uppgift.

Avslutande tankar

Generativ AI och digitala informationssystem har alltså en dubbel roll. Å ena sidan kan de ge oss *mer* information, mer personalisering och nya spännande verktyg. Å andra sidan utmanar de våra traditionella sätt att bedöma trovärdighet och källa. För gymnasielärare – och alla som utbildar i källkritik – gäller det nu att **bredda** undervisningen. Det räcker inte längre att tala om primär- och sekundärkällor, eller att kontrollera utgivare och datum. Eleverna måste förstå *mekanismerna* bakom det de ser online: att en algoritm valde ut just det innehållet, att deras egna klick formar framtida flöde, att deras data är en handelsvara, och att det de ser eller hör kan vara framställt av en AI.

Att undervisa i källkritik i denna kontext kan kännas utmanande, men också högst angeläget. Genom att använda de teman vi gått igenom – algoritmernas urval, informationsnätverkens dynamik, filterbubblor, ekokammare, personalisering kontra integritet, och hyperverklighetens fallgropar – kan lärare ge eleverna en **verktygslåda** för kritiskt tänkande i det digitala landskapet. Konkret kan övningar omfatta allt från att jämföra personliga flöden, analysera hur en nyhet blev viral, diskutera egna erfarenheter av att “fastna” i ensidiga forum, till att prova på att skapa en enkel deepfake eller AI-genererad text och sedan gemensamt reflektera över hur man kan avslöja den. Elever som själva får experimentera med tekniken förstår både dess potential och dess begränsningar bättre.

Slutligen, även om tekniken förändras, ligger kärnan i källkritik och kritiskt tänkande fast: det handlar om att inte ta information för given, att ställa frågor och att söka kunskap från flera håll innan man drar slutsatser. I en AI-driven informationsålder behöver vi den inställningen mer än någonsin. Genom att öka kunskapen om hur generativ AI och digitala system fungerar under huven kan vi vaccinera oss mot naiviteten och ge kommande generationer självförtroendet att navigera informationssamhället med både **öppet sinne och sunt tvivel**.

Källor:

- Russell, A. (2023). *YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users*, UC Davis (refererar studie i PNAS) ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)) ([YouTube Video Recommendations Lead to More Extremist Content for Right-Leaning Users, Researchers Suggest | UC Davis](#)).
- Field, H. (2021). *Facebook whistleblower Frances Haugen points to 2018 algorithm overhaul as key problem*, Emerging Tech Brew ([Facebook whistleblower Frances Haugen points to 2018 algorithm overhaul as key problem](#)) ([Facebook whistleblower Frances Haugen points to 2018 algorithm overhaul as key problem](#)).
- Wikipedia (2023). *Filter bubble* ([Filter bubble - Wikipedia](#)) ([Filter bubble - Wikipedia](#)).
- Terren, L., & Borge-Bravo, R. (2021). *Echo Chambers on Social Media: A Systematic Review of the Literature*, Review of Communication Research ([Echo Chambers on Social Media: A Systematic Review of the Literature | Review of Communication Research](#)).
- Li, H. et al. (2023). *Echo chamber effects on short video platforms*, *Scientific Reports* ([Echo chamber effects on short video platforms - PMC](#)).
- Vosoughi, S., Roy, D., & Aral, S. (2018). *The spread of true and false news online*, *Science* ([Study: On Twitter, false news travels faster than true stories | MIT News |](#)

[Massachusetts Institute of Technology](#)) ([Study: On Twitter, false news travels faster than true stories | MIT News | Massachusetts Institute of Technology](#)).

- Weigel, M. (2013). *Facebook, private traits and attributes: Predictions from digital records of human behavior*, The Journalist's Resource ([Facebook, private traits and attributes: Predictions from digital records of human behavior - The Journalist's Resource](#)).
- Wikipedia (2019). *Facebook–Cambridge Analytica data scandal* ([Facebook–Cambridge Analytica data scandal - Wikipedia](#)).
- Euronews (2022). *Deepfake Zelenskyy surrender video...* ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)) ([Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war | Euronews](#)).
- Reuters (2023). *Fact Check: Image of Pope Francis in puffer coat is AI-generated* ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)) ([Fact Check: Image of Pope Francis wearing oversized white puffer coat is AI-generated | Reuters](#)).

Hallucinationer hos AI

När man säger att en AI *“hallucinerar”* menar man att den hittar på information – den ger svar som låter rimliga men som **inte stämmer med verkligheten** ([Hallucination \(artificial intelligence\) - Wikipedia](#)). Ordet kommer från en liknelse med mänskliga hallucinationer, men i AI-sammanhang handlar det inte om sinnesintryck utan om felaktiga *text*- eller bildgenereringar. Till exempel kan en chatbot som drivs av ett stort språkmodell, som ChatGPT, självsäkert presentera **påhittade “fakta” eller källor** utan grund i data den tränats på ([Hallucination \(artificial intelligence\) - Wikipedia](#)) ([Hallucination \(artificial intelligence\) - Wikipedia](#)). Forskare har noterat problemet: en analys år 2023 estimerade att chatbotar som ChatGPT kan *hallucinera* i uppåt 27% av fallen – alltså mer än var fjärde svar kan innehålla någon felaktighet – och nästan hälften av genererade texter innehöll faktamissar ([Hallucination \(artificial intelligence\) - Wikipedia](#)). Detta innebär att även om svaren språkligt sett är imponerande, måste vi förhålla oss **källkritiskt** till innehållet.

Varför uppstår hallucinationer? Orsaken ligger i hur generativa AI-modeller fungerar. En språkmodell som GPT är i grunden ett **avancerat autokompletteringssystem** som försöker förutsäga nästa ord baserat på statistiska mönster i träningsdatan ([When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies](#)). Den är inte byggd för att kontrollera om påståenden är sanna. Om det finns luckor i kunskapen eller om användarens fråga ligger utanför det modellen “minns” från sin träning, kommer den ändå att leverera ett svar – och då kan den *fabricera* detaljer som låter troliga men är fel. Modellen saknar alltså en **sanningskontroll** och kan återge både korrekta och inkorrekta uppgifter från sin träningsdata (som i sin tur kan innehålla fel och bias) ([When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies](#)). Dess mål är *plausibilitet*, inte sanning, vilket gör att “hallucinationer” är ett **inneboende problem** i dagens generativa AI-system ([When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies](#)).

Konsekvenserna av AI-hallucinationer kan bli påtagliga. Ett omtalat exempel är det amerikanska rättsfallet *Mata v. Avianca* 2023, där en advokat använde ChatGPT för juridisk research. ChatGPT levererade övertygande referenser till rättsfall – som visade sig **inte existera överhuvudtaget** ([When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies](#)). Chatboten hittade på fallnamn, citat och domslut och hävdade till och med att dessa fanns i välkända databaser ([When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies](#)). Advokaten tog med de falska referenserna i en inlägga, vilket ledde till att domaren skarpt kritiserade tillvägagångssättet och krävde att inga AI-genererade uppgifter får lämnas in utan **mänsklig verifiering** ([Federal judge: No AI in my courtroom unless a human verifies its ...](#)). Denna händelse – där jurister blev vilseledda av en *hallucinerande* AI – illustrerar varför kritiskt tänkande och källgranskning är *avgörande* när man använder generativ AI. Som användare bör man alltid dubbelkolla fakta som en AI-modell ger, särskilt om informationen är viktig eller ovanlig. Kort sagt: **lita inte blint på AI**, utan betrakta dess utlåtanden som *förslag* som behöver bekräftas via pålitliga källor.

Manipulationstekniker – hur information kan vinklas

All form av desinformation och propaganda – oavsett om den sprids manuellt eller med AI:s hjälp – använder ofta vissa **retoriska knep** för att övertyga. Att känna igen dessa manipulationstekniker är en kärnkompetens i källkritik. Här är några vanliga metoder:

- **Cherry picking (plocka russinen ur kakan):** Detta innebär att man **väljer ut de fakta som stödjer ens budskap och ignorerar motbevis** ([Cherry picking - Wikipedia](#)). Genom att lyfta fram enskilda exempel eller datapunkter som passar ens narrativ, men utelämna helhetsbilden, kan man ge en falsk bild av verkligheten. Cherry picking kallas även **“ofullständig evidens”** och kan ske medvetet eller omedvetet ([Cherry picking - Wikipedia](#)) ([Cherry picking - Wikipedia](#)). Ett internationellt känt exempel är inom klimatdebatten: Klimatförnekare har pekat på perioden 1998–2012 och hävdade att *“global uppvärmning tagit en paus”*, eftersom just de åren visar en svagare temperaturtrend. Men detta beror på ett medvetet urval av start- och slutår – hela den långsiktiga datan visar tydligt fortsatt uppvärmning ([Cherry picking - Wikipedia](#)). Genom att **skära av datan** vid 1998 kan man *fabricera* en illusion av stillastående klimat, trots att det egentligen bara var en kortvarig variation. För elever är detta ett utmärkt fall att analysera: hur kan grafer eller statistik vilseleda beroende på vilket utsnitt man väljer? Cherry picking understryker behovet av att alltid söka efter *helhetsbilden* och ifrågasätta om det finns undantag eller motstridig information som har utelämnats.
- **Känsloladdat språk:** Manipulatorer använder ofta **starka ord och känsloladdade argument** för att påverka publiken på ett irrationellt plan. *Loaded language*, som det kallas på engelska, syftar till att väcka t.ex. rädsla, ilska eller entusiasm, snarare än att appellera till logik ([Loaded language - Wikipedia](#)). Genom värdeladdade termer och suggestiva formuleringar kan man få publiken att reagera utan att noga pröva fakta. Exempelvis kan en rubrik som *“Den chockerande sanningen myndigheterna inte vill att du ska veta!”* trigga misstänksamhet och klick, trots att innehållet kanske är grundlöst. Inom politiken ser vi ofta känsloladdade omdömen – en debattör kanske kallar motståndare för *“förrädare”* eller *“lögnare”* för att väcka avsky, i stället för att bemöta sakfrågan (). På sociala medier sprids virala inlägg som använder **upprörande eller rörande språk** (*“Det här hjärtskärande klippet kommer få dig att gråta”*) för att fånga uppmärksamhet (). Lärare kan här visa elever hur ordval kan manipulera: jämför neutrala beskrivningar med tendentiösa. Genom att identifiera **emotionellt laddade ord** lär sig eleverna att ta ett steg tillbaka och fråga: *“Försöker texten få mig att känna något starkt för att övertyga mig, i stället för att presentera sakliga skäl?”*.
- **Falska dikotomier:** Denna teknik – även kallad *falskt dilemma* – innebär att man **framställer en fråga som om det bara finns två alternativ** fast det i själva verket finns flera nyanser eller möjligheter () ([False dilemma - Wikipedia](#)). Det är ett logiskt felslut där komplexa frågor förenklas till ett antingen-eller. Ett klassiskt exempel är det berömda citatet *“Antingen är ni med oss, eller så är ni med terroristerna”*, som USA:s president använde efter 11 september 2001. Sådana formuleringar tvingar in mottagaren i ett hörn: *väljer du inte min sida så står du på fel sida*. I själva verket kan man förstås både kritisera terrorism **och** ifrågasätta vissa av USAs metoder – fler än två positioner existerar. I dagens debatter ser vi falska dikotomier i påståenden som *“Om du inte stöder förslaget, bryr du dig inte om X”* (där X kan vara t.ex. rättvisa, miljö, säkerhet) (). Ett annat exempel: *“Antingen litar vi blint på all vetenskap eller också är allt en lögn.”* – vilket är en falsk dikotomi då man både kan uppskatta

vetenskapens värde och samtidigt granska enskilda studier kritiskt. **Syftet** med falska dikotomier är ofta att polarisera och förenkla, så att mottagaren inte ser mellanting eller alternativa lösningar. Genom att peka ut detta knep kan elever öva sig i att hitta *tredje alternativ* och förstå att världen sällan är svartvit.

- **Andra retoriska knep:** Utöver ovanstående finns en rad andra metoder som desinformatörer använder:
 - **Ad hominem-attacker:** här går man **på personen istället för sakfrågan** – man undergräver en källas trovärdighet genom personangrepp, för att avleda från argumentens innehåll (). T.ex. “*Du kan inte lita på vad han säger om folkhälsa, han är ju inte klok!*”. Ibland kan visserligen en persons bakgrund vara relevant (t.ex. om någon har ekonomiska intressen i ämnet), men rena personangrepp utan koppling till sakfrågan är ett retoriskt vilseledande.
 - **Falsk balans:** media eller debattörer kan ge sken av att två sidor av en fråga är likvärdiga, trots att **bevisläget är ojämnt**. Ett typexempel är när 99% av experterna och all data talar för en ståndpunkt, men man ändå framställer det som “två läger” som debatterar (). “*Experter oense om jordens form – en forskare säger att den är rund, en annan menar att den är platt*” (). Genom att placera en obskyr röst mot en överväldigande vetenskaplig konsensus, luras publiken att tro att frågan är kontroversiell fast den i praktiken är avgjord. Elever kan här lära sig att väga *kvaliteten* på evidens, inte bara räkna åsikter för och emot.
 - **Konspirationsteorier:** Många desinformationskampanjer använder konspiratoriskt tänkande – de påstår att en *dold, mäktig grupp* ligger bakom händelser och att “officiell” information inte går att lita på (). Genom att hänvisa till mystiska sammansvärjningar (“*egentligen är det miljardärer som styr allt via X*”) får man publiken att tvivla på etablerade källor och i stället tro på ryktesspridarnas egen version. Konspirationsteorier uppmuntrar människor att känna sig invigda i en hemlig “sanning”, vilket kan vara psykologiskt lockande.
 - **Whataboutism och avledning:** Detta är tekniker där man svarar på kritik eller fakta med “*Men hur är det med...*” följt av ett annat ämne, för att avleda uppmärksamheten. Istället för att bemöta sakfrågan byter man fokus och försöker hitta hyckleri hos motparten eller pekar på ett annat problem.

Listan kan göras lång, men gemensamt för alla dessa knep är att de försöker **påverka informationsflödet till avsändarens fördel**, antingen genom att *vilseleda logiskt* (felslut, falska val, selektiv data) eller *påverka emotionellt* (rädsla, ilska, vi-mot-dem-retorik). Generativ AI kan potentiellt underlätta spridningen av sådan manipulerande kommunikation genom att snabbt producera mängder av vinklade texter, inlägg eller till och med AI-genererade bilder/filmer (s.k. deepfakes). Därför blir det ännu viktigare att lärare utrustar elever med ”detektorglasögon” för dessa tekniker. Genom att låta elever analysera textexempel och identifiera cherry picking, känsloladdade uttryck eller falska dikotomier i dem, tränas de i att **inte låta sig ryckas med** direkt av budskapet utan att först tänka efter: *Hur försöker texten övertyga mig? Bygger argumenten på fakta och logik, eller spelas det spratt med retoriken?*

Prebunking – förebyggande vaccin mot desinformation

Ett relativt nytt buzzword inom källkritik är **prebunking**, som kan beskrivas som *förtida debunking* – det vill säga att man **motverkar falsk information innan den får fäste**. Idén kommer från psykologisk forskning och kallas ofta för **“inokuleringsteorin”** (inoculation theory) inom misinformation. Tanken är analog med vaccinering: genom att *i förväg* exponera människor för en mild dos av vilseledande argument eller avslöja de tekniker som används för att manipulera, kan man bygga upp ett “mentalt immunförsvar” ([Why "prebunking" is the best way to fight misinformation - Big Think](#)). När personen sedan möter den verkliga desinformationen är hen redan beredd och mindre benägen att låta sig luras ([Why "prebunking" is the best way to fight misinformation - Big Think](#)).

Prebunking kan ske på olika sätt. Forskare vid Cambridge universitet och andra lärosäten har experimenterat med **små träningsmoment** – till exempel korta videoklipp eller onlinespel – som lär ut vanliga manipulationstekniker. Ett exempel är spelet *Bad News*, utvecklat i Europa, där spelaren själv får rollen som “fake news”-skapare för att lära sig vilka knep som används ([Prebunking interventions based on the psychological theory of “inoculation” can reduce susceptibility to misinformation across cultures. | HKS Misinformation Review](#)). Studier har visat att efter att ha spelat spelet blev deltagarna *mindre mottagliga* för misinformation i efterföljande tester ([Prebunking interventions based on the psychological theory of “inoculation” can reduce susceptibility to misinformation across cultures. | HKS Misinformation Review](#)). På liknande sätt har korta animerade videor tagits fram i samarbete mellan Jigsaw (en avdelning på Google) och Cambridge-forskare. Dessa 90-sekunders klipp förklarade knep som **känslomässig språkföring** och **falska dikotomier**, med konkreta exempel hämtade från populärkulturen (['Pre-bunking' Fake News on YouTube Helped Curb Misinformation ...](#)) ([Why "prebunking" is the best way to fight misinformation - Big Think](#)). När miljontals användare på YouTube fick se dessa som annonser, såg man en mätbar – om än modest – ökning (~5% högre) i deras förmåga att känna igen vilseledande retorik jämfört med kontrollgruppen ([Why "prebunking" is the best way to fight misinformation - Big Think](#)). Effekten var dessutom **bred**: den syntes hos både höger- och vänsterinriktade personer, oavsett utbildningsnivå ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Att lära ut *hur* man blir manipulerad tycks alltså ge en viss skyddande effekt över lag.

Ett uppmärksammat internationellt exempel på prebunking skedde i samband med Rysslands invasion av Ukraina 2022. Tech-bolaget Google tog initiativ till en **prebunking-kampanj i Östeuropa** för att mota bort desinformation riktad mot ukrainska flyktingar ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Korta videoklipp spreds på plattformar i Polen, Tjeckien och Slovakien med syfte att *förvarna* publiken om falska narrativ och ge dem motargument i förväg ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Tanken var att när ryska propaganda senare skulle försöka sprida anti-flykting-rykten, skulle fler redan vara vaccinerade mot retoriken och därför genomskåda den. Forskaren Beth Goldberg vid Jigsaw beskrev att strategin var *“att lära folk känna igen propaganda-manuset”*, eftersom de flesta vilseledande budskap bygger på återkommande mönster ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)) ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Genom att undervisa allmänheten om knep som t.ex. ad hominem-attacker eller syndabocksretorik (“skyll på *dem*

för allt”), hoppas man skapa **resiliens** – en motståndskraft mot att tro på och sprida falsk information i framtiden ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)).

För lärare innebär prebunking-perspektivet att källkritiksundervisningen kan läggas upp *proaktivt*. Istället för att enbart reagera på elevernas missuppfattningar i efterhand, kan man tidigt i en kurs “inokulera” eleverna genom att visa typiska exempel på manipulerade budskap. Till exempel kan man förekomma konspirationstänkande när man lär ut om källor: “Ni kommer kanske stöta på videor som påstår X – låt oss titta på varför de använder en sådan retorik, och vad som talar emot den.” Att föregripa vanliga feluppfattningar ger eleverna ett slags mentalt ramverk att känna igen strunt när de ser det. Forskning tyder på att en sådan förvarning, kombinerad med förklaring av *varför* något är vilseledande, gör eleverna mindre benägna att acceptera samma falska påstående senare ([Why "prebunking" is the best way to fight misinformation - Big Think](#)) ([Why "prebunking" is the best way to fight misinformation - Big Think](#)). Detta betyder inte att de blir helt immuna – effekten av prebunking kan avta med tiden och behöver kanske upprepas ([Why "prebunking" is the best way to fight misinformation - Big Think](#)) – men det kan vara ett effektivt komplement till vanlig faktaundervisning.

Slutligen ska det sägas: Prebunking är ingen magisk lösning på desinformation, men det adresserar problemet **i förebyggande syfte**, vilket är värdefullt i en tid då det sprids för många falska påståenden för att man ska hinna bemöta alla i efterhand. Som två ledande forskare uttrycker det: det är mer effektivt att *lära människor känna igen* tricken bakom falska budskap än att försöka *motbevisa vartenda falskt påstående* efter att det redan fått spridning ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Prebunking ger alltså källkritikundervisningen en framåtblickande dimension – att rusta eleverna **inför** framtida informationsmöten.

Debunking – att granska fakta och avfärda falsk information

Trots förebyggande insatser kommer elever (och vi alla) att stöta på felaktig information. Då krävs **debunking**, det vill säga att man *identifierar, granskar och korrigerar* falska påståenden. Debunking är den traditionella källkritikens hemmaplan: faktagranskning och mytavlivning. Flera etablerade organisationer globalt ägnar sig åt detta, såsom Snopes, FactCheck.org, PolitiFact, AFP Fact Check med flera, liksom svenska aktörer som Källkritikbyrå och Metros Viralgranskaren. Men hur gör man debunking på ett effektivt sätt? Forskning i kognitiv psykologi har gett upphov till **bästa praxis**:

- **Ge en alternativ förklaring:** När man avfärdat en felaktig uppgift skapas ofta ett tomrum – mottagaren undrar “okej, det där var falskt, men vad är då sanningen?”. Människor ogillar kognitiva luckor. Därför bör man försöka erbjuda en **rimlig alternativ förklaring** till fenomenet ([Psychological Interventions to Combat Misinformation](#)). Exempel: om man debunkar myten att vaccin orsakar en viss sjukdom, bör man samtidigt förklara vad som annars kan ligga bakom sjukdomsfallen. Annars riskerar folk att uppleva att något “hänger i luften”, vilket kan göra myten seglivad trots motbevis.
- **Förklara varför det är fel:** Att bara säga “det där stämmer inte” räcker inte långt. En effektiv debunking går in på **detalj** och presenterar bevis och resonemang som tydligt visar *varför* påståendet är felaktigt ([Psychological Interventions to Combat](#)

[Misinformation](#)). All relevant fakta ska på bordet – här gäller det att *överbevisa* misstaget. Om en konspirationsteori bygger på en förfälskad bild, visa den äkta bilden och hur man vet att den andra är manipulerad. Om ett viralt rykte påstår något tvärt emot forskningen, länka till de vetenskapliga studierna och sammanfatta resultaten. Ju mer konkreta motbevis, desto bättre.

- **Upprepa inte myten i onödan:** En knepig paradox med debunking är att **upprepad exponering** av ett påstående – även i syfte att dementera det – kan göra att det fastnar i minnet hos folk (det s.k. *illusory truth effect*). Därför rekommenderas att *inte* älta den falska utsagan mer än nödvändigt ([Psychological Interventions to Combat Misinformation](#)). Ofta behöver man nämna myten *en gång* för att klargöra vad man korrigerar, men sedan bör fokus ligga på fakta. Det kan också vara klokt att **varna läsaren/lyssnaren innan** man återger den felaktiga uppgiften, så de är mentalt beredda på att “här kommer en osanning som snart ska vederläggas” ([Psychological Interventions to Combat Misinformation](#)). På så vis minskar risken att myten förstärks i hjärnan hos mottagaren.
- **Repetera sanningen över tid:** Även om en bra debunking kan få människor att rätta sin uppfattning, visar studier att effekten kan **avta med tiden** när minnet bleknar ([Psychological Interventions to Combat Misinformation](#)). Felaktig information kan då smyga sig tillbaka i folks medvetande (särskilt om de ser myten igen någon annanstans). Därför kan det behövas att man påminner eller uppdaterar korrekt information flera gånger om ett ämne, för att vidmakthålla den rättade uppfattningen ([Psychological Interventions to Combat Misinformation](#)). I praktiken: en engångsföreläsning om att jorden är rund botar inte alla flat earth-anhängare för all framtid; fortsatt undervisning och repetition behövs.
- **Var en trovärdig korrekturläsare:** Vem som utför debunkingen spelar roll. Mottagare är mer benägna att acceptera rättelser om de **litar på källan** som levererar dem ([Psychological Interventions to Combat Misinformation](#)). Därför bör man sträva efter att etablera ett *förtroendeöverskott* – dvs. ha mer **cred** än den som spred myten. I klassrummet har läraren ofta denna auktoritet, men online kan det krävas att pålitliga tredje parter (experter, faktagranskare, erkända medier) står för rättelsen. Ibland kan det även vara effektivt att kort påpeka *varför* den ursprungliga källan hade fel, t.ex. om den hade en tydlig bias eller intressekonflikt ([Psychological Interventions to Combat Misinformation](#)). (“*Den här artikeln som påstod att klimatförändringen är en bluff visade sig vara finansierad av en oljeindustrilobby.*”) Då förstår publiken inte bara *att* påståendet var fel, utan även *hur* och *varför* felaktigheten uppstod – vilket kan öka vaksamheten framöver.

Dessa principer kan fungera som riktlinjer för undervisning i praktisk faktagranskning. En övning för elever kan vara att ta en faktisk felaktig nyhet eller myt, och låta dem konstruera en **debunking-analys** enligt modellen ovan: Vilka bevis behövs? Hur formulerar man en rubrik som inte upprepar myten men lockar till läsning? Vilken alternativ förklaring bör ges? Etc. Genom att själva prova på att “vara faktagranskare” internaliserar eleverna dessa strategier.

Det är också värt att diskutera debunkingens begränsningar. Ibland möter man personer som *inte vill* bli övertygade av fakta – starka känslor eller världsbilder kan göra att korrigeringar avfärdas som “fejkade” oavsett evidens. Rättelse kan upplevas som ett **personligt angrepp** mot någon som investerat sin identitet i en viss övertygelse, särskilt i polariserade frågor ([Social media experiment reveals potential to 'inoculate' millions of users against misinformation](#)). Då gäller det att ha tålamod, hålla en respektfull ton och kanske byta

angreppssätt (t.ex. fokusera på gemensamma värderingar först, eller låta personen själv resonera fram svagheterna i myten). Debunking är oftast *nödvändigt* för att sanningen ska segra, men det är inte alltid *tillräckligt* ensamt. Därför förespråkar experter en **kombination** av insatser för en sund informationsmiljö: *fortsatt faktagranskning* där det går, *prebunking* för att bygga motståndskraft, samt *medieutbildning* som stärker allmän källkritisk kompetens ([Psychological Interventions to Combat Misinformation](#)) ([Psychological Interventions to Combat Misinformation](#)).

Källtillit och sakkritik – att värdera källor och experter

Ett sista viktigt tema är **källtillit** och **sakkritik**, dvs. hur vi avgör *vilka* källor vi kan lita på och om en uppgiftslämnare besitter verklig sakkunskap. I en tid av informationsöverflöd – där vem som helst kan publicera vad som helst – kan elever känna att “*inget går att lita på*”. Men målet med källkritik är inte cynism eller misstro mot *all* information, utan att kunna urskilja **trovärdiga källor** från opålitliga. *Källtillit* handlar om att ha förtroende för de källor som förtjänar det, och *sakkritik* innebär att kritiskt granska **innehållet i sak** även om källan är formellt auktoritativ.

Hur bedömer vi en källas pålitlighet? Här finns etablerade checklistor och metoder (t.ex. **SIFT-modellen** eller den klassiska **CRAAP-testen**). Några grundläggande frågor elever bör lära sig ställa:

- **Vem står bakom informationen?** Identifiera källan: Är det en känd nyhetsorganisation, en myndighet, en forskare – eller är det en anonym blogg, ett konto på sociala medier utan klargjord identitet? Seriösa aktörer brukar ha kontaktuppgifter, “Om oss”-sidor, och ett rykte att upprätthålla. Oklara eller helt okända avsändare förtjänar extra skepsis ([A quick guide to spotting misinformation | UNICEF Europe and Central Asia](#)). “*Betrakta källan*” är steg ett i många guider: uppmana elever att klicka bortom själva artikeln/inlägget och undersöka webbplatsens syfte och bakgrund ([How To Spot Fake News](#)). Om en källa har en tydlig agenda – politisk, religiös, kommersiell – bör man vara uppmärksam på vinklingar. Att få sin nyhetsbild från flera olika håll är klokt; om en uppgift **bekräftas av många oberoende källor** ökar sannolikheten att den är sann ([A quick guide to spotting misinformation | UNICEF Europe and Central Asia](#)).
- **Vad är det för typ av innehåll?** Är det nyhetsrapportering, opinionsmaterial, satir, eller reklam maskerad som artikel (s.k. advertorial)? Elever bör lära sig skilja på nyhetsgenre och inte ta t.ex. en krönika eller insändare som objektiv sanning. Och är innehållet aktuellt? Äldre artiklar som delas i nya sammanhang kan vilseleda (kolla datum!) ([A quick guide to spotting misinformation | UNICEF Europe and Central Asia](#)). Ibland sprids rena skämt eller satir som om de vore äkta – en snabb koll på om källan är känd för humor eller satir kan spara mycket missförstånd ([How To Spot Fake News](#)). Frågan “*Kan det här vara avsett som ett skämt?*” är relevant inför väldigt bisarra påståenden.
- **Stämmer det överens med andra källor och expertis?** Om en häpnadsväckande nyhet bara kommer från *en* källa medan inga andra medier nämner det, bör varningsklockor ringa ([A quick guide to spotting misinformation | UNICEF Europe and Central Asia](#)). Lär elever göra en snabb sökning: “*Är andra trovärdiga aktörer*

också på detta?”. Ofta kan man på så vis avslöja bluffar – t.ex. om en studie påstås “bevisa” något anmärkningsvärt, kolla om den refereras i etablerad press eller forskningssidor. Om inte, kanske studien är dubiös eller feltolkad. **Lateral läsning** är ett användbart grepp: att parallellt söka information *om källan* och *om ämnet* i andra flikar för att sätta det i context. Elever kan träna på att kolla upp **stödbevis**: om en artikel refererar till statistik eller vittnen, finns det länkar eller källhänvisningar? Följ dem! Stödjer de verkligen det som påstås? ([How To Spot Fake News](#))

- **Vem är “experten” och är hen verkligen expert?** Ofta stöter vi på namn som presenteras med auktoritet, t.ex. “Dr. X säger att...”. Här behövs *sakkritik*: är denna person **kvalificerad inom området**? En person kan vara doktor i ett ämne men uttala sig om något helt annat där hen saknar kompetens. Ett läskigt exempel är tobaksindustrins gamla strategi med “**låtsasexperter**”: på 1980-talet hyrdes vissa forskare in – som inte forskade på rökningens effekter – för att bära *vita rockar* och inge förtroende när de sådde tvivel om rökningens faror ([How fake experts are used to mislead - Cranky Uncle](#)) ([How fake experts are used to mislead - Cranky Uncle](#)). De *såg ut* som experter men hade inte evidensen bakom sig. I modern tid ser vi liknande tendenser med t.ex. klimatförnekare eller antivaccin-profiler som åberopar någon akademisk titel men går emot den samlade expertkunskapen. **En riktig expert** kännetecknas av *djup relevant utbildning* och *erfarenhet* inom just det fältet, samt ofta *peer review*-granskade publikationer ([How fake experts are used to mislead - Cranky Uncle](#)). John Cook, forskare inom klimatinformation, föreslår en bra tumregel: en varningsflagga är om någon **åberopar sin auktoritet men samtidigt förnekar rådande forskningskonsensus** ([How fake experts are used to mislead - Cranky Uncle](#)). Då bör man definitivt “granska examen” – har personen faktiskt expertis inom området, eller är det en “*fake expert*” som försöker ge tyngd åt desinformation? Att lära eleverna kolla upp en experts bakgrund (t.ex. genom att söka på deras namn, se vad de har för publikationer eller var de arbetar) är värdefullt ([A quick guide to spotting misinformation | UNICEF Europe and Central Asia](#)). Också: har *andra* experter respekt för personen, eller anses hen kontroversiell? Samstämmighet bland oberoende experter är ofta ett gott tecken.
- **Är informationen rimlig och självkoherent?** Till sist ska elever uppmuntras att även med sunda förnuftet och kunskaper i ämnet *analysera sakinnehållet*. Ibland kan en källa vara legitim men ha fel ändå – experter kan göra misstag, nyhetsmedier kan rapportera felaktigt. Sakkritik innebär att pröva uppgifterna mot annat man vet. Ställer sig ens *logiska* eller *vetenskapliga* förståelse upp och skriker att något är fel, så är det läge att undersöka vidare. Till exempel: om en artikel påstår att en ny mirakelkur botar alla sjukdomar, går det emot allt eleverna lärt sig om biologi – alltså högst tvivelaktigt, *även om* källan försöker se seriös ut. Eller om en graf i en studie ser konstig ut (kanske skalan är manipulerad, eller data spretar): då ska man inte hänfälla åt auktoritetstro utan våga fråga, “*kan det verkligen stämma?*”. Denna djupare analys av innehållets kvalitet är essensen av sakkritik.

Sammanfattningsvis handlar källkritik i AI-eran om en balans mellan **skeptisk granskning** och **informerat förtroende**. Elever behöver lära sig att inte tro på allt de ser – men heller inte förkasta allt som “fake” i post-sanningens tecken. Istället ska de kunna bygga upp *källtillit* gentemot de informationskällor som **gång på gång visar sig pålitliga**, såsom välrenommerad vetenskapsjournalistik, uppslagsverk, forskarsamhället och andra kvalitetsgranskade forum. Genom att konsekvent använda verktyg som iakttagelser av språk, kontroll av källhänvisningar, korsjämförelser med andra källor och koll av expertis, kan eleverna utveckla ett **självsäkert kritiskt tänkande**. Målet är att de varken blir lättmanipulerade av en

välformulerad AI-text eller apatiskt misstror *all* information, utan att de navigerar informationsflödet som *aktiva detektiver*.

I en värld där generativ AI kan massproducera innehåll – sant och falskt om vartannat – är denna kompetens viktigare än någonsin. Genom att förstå fenomen som AI-hallucinationer, genomskåda retoriska manipulationer och vara rustade med strategier som prebunking, debunking och smart källvärdering, kan nästa generation internetanvändare bli vaksamma och upplysta medborgare. **Källkritik 2.0**, i AI:s tidevarv, handlar inte bara om att hitta sanningen utan också om att *förstå lögnen*. Genom att lära ut *hur* och *varför* felaktigheter uppstår – vare sig det är en dator som hallucinerar eller en propagandist som kör falska dilemman – ger vi eleverna verktygen att tänka själva och stå emot manipulation. Detta är kärnan i modern bildning och demokratisk resiliens.

Bias inom AI – varför partiska algoritmer är en utmaning för källkritik

AI används idag i allt från sociala medier till rekrytering och undervisning. Men precis som vi människor kan AI-system ha **bias** – en inbyggd partiskhet som gör att resultaten kan gynna eller missgynna vissa grupper eller perspektiv. För gymnasielärare är det viktigt att förstå hur bias i AI fungerar, både för att kunna bedöma AI-genererat material kritiskt och för att lära elever om källkritik i en digital tidsålder. I denna text går vi igenom vad AI-bias är, varför det uppstår, vilka risker det medför samt hur våra egna kognitiva bias påverkar – och påverkas av – AI. Vi ger exempel från verkligheten och hänvisar till forskning för att ge en djupare förståelse.

Vad innebär bias inom AI?

AI-bias syftar på att ett AI-system *systematiskt* favoriserar vissa utfall eller grupper på ett orättvist sätt. Det betyder att algoritmens beslut inte är neutrala. Bias kan visa sig som diskriminerande eller stereotyp beteende hos AI:t och beror ofta på att mänskliga snedvridningar "läckt in" i systemet. AI kan nämligen ärva våra fördomar – till exempel könsstereotyper eller rasistiska mönster – via de data och regler vi människor ger det ([AI Bias Examples | IBM](#)). Med andra ord producerar ett partiskt AI-system resultat som *återspeglar och förstärker* existerande mänskliga bias och ojämlikheter ([AI Bias Examples | IBM](#)).

Ett uppmärksammat exempel är när **Amazon utvecklade ett AI-baserat rekryteringssystem**. Tanken var att automatiskt ranka arbetssökande, men det visade sig att systemet diskriminerade kvinnor ([Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters](#)). Algoritmen hade tränats på tio års rekryteringsdata där majoriteten av de anställda var män, vilket ledde till att AI:t lärde sig att manliga kandidater var att föredra. Resultatet blev att CV:n som innehöll ordet "kvinnors", till exempel erfarenheter som "ordförande i **kvinnors** schackklubb", fick lägre betyg, och kandidater från kvinnosegregerade högskolor missgynnades ([Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters](#)). Här ser vi hur historisk snedfördelning i datan (fler män i tech-branschen) gjorde att AI:t blev partiskt.

Ett annat exempel är **ansiktigenkänning**. Forskning av Joy Buolamwini och Timnit Gebru vid MIT Media Lab visade att kommersiella ansiktigenkänningssystem var mycket mindre träffsäkra för svarta kvinnors ansikten än för vita mäns ([Study finds gender and skin-type bias in commercial artificial-intelligence systems | MIT News | Massachusetts Institute of Technology](#)). Detta berodde på att algoritmerna tränats på bilddatabaser med oproportionerligt många bilder av ljushyade män. Konsekvensen blev att mörkhyade kvinnor feligenkändes i upp till 35% av fallen, jämfört med nära felfri igenkänning för vita män ([Study finds gender and skin-type bias in commercial artificial-intelligence systems | MIT News | Massachusetts Institute of Technology](#)). Ett sådant bias är inte bara orättvist utan kan vara farligt – tänk om polisens ansiktigenkänning identifierar fel person på övervakningsbilder på grund av detta?

Även **generativ AI**, som chatbotar och bildgeneratorer, uppvisar bias. Språkmodeller tränas på enorma mängder text från internet, vilket gör att de kan återspegla stereotyper som finns i det materialet. Forskare vid MIT fann till exempel att en avancerad språkmodell kodade in

könsstereotyper om yrken: den ansåg att ord som "flygvärdinna", "sekreterare" och "läkarassistent" var kvinnligt kodade yrken, medan "fiskare", "advokat" och "domare" sågs som manligt kodade ([Does AI Have a Bias Problem? | NEA](#)). På liknande sätt har en studie vid Dartmouth visat att språkmodeller kan ha inbyggda fördomar som antar att vissa grupper är bättre eller sämre på vissa färdigheter baserat på stereotypa kopplingar ([Does AI Have a Bias Problem? | NEA](#)). Kort sagt kan generativ AI – från texter till bilder – skapa innehåll som *verkar* objektivt men som i själva verket lutar åt de fördomar som fanns i träningsdatan.

Varför uppstår AI-bias?

Det finns flera anledningar till att AI-system blir partiska. En huvudorsak är **träningsdatan**. AI lär sig från exempel, och om exemplen är skeva blir också AI:ts beslut skeva. Om en datamängd inte är representativ för verkligheten – säg att den främst innehåller information om en viss grupp människor och ignorerar andra grupper – kommer AI:t att prestera sämre för underrepresenterade grupper ([AI Bias Examples | IBM](#)). I fallet med ansiktsigenkänning överrepresenterades till exempel vita ansikten, vilket orsakade fel för mörkhyade personer ([AI Bias Examples | IBM](#)). Inom polisarbete har det uppmärksammats att prediktiva system ibland tränas på historisk brottsdata från områden med hög andel minoriteter, vilket riskerar att algoritmen *överbetonar* dessa områden (eftersom datan samlats in där) och därmed förstärker en skev bild av var brott begås ([AI Bias Examples | IBM](#)). Bias kan alltså "smittas" in i AI redan genom vilka data vi väljer att samla in och använda.

Även hur data **märks upp och tolkas** spelar roll. Om människorna som förbereder träningsdatan har omedvetna fördomar kan dessa leta sig in i AI:t. Inom rekrytering kan till exempel valda nyckelord eller kategoriseringar leda fel – Amazons verktyg påverkades av att ord associerade med kvinnor fått lägre värde i träningsmaterialet ([Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters](#)). På samma sätt kan AI för språkanalys tränas på texter där vissa grupper konsekvent framställs negativt. Då finns risken att modellen reproducerar dessa associationer när den genererar nytt innehåll.

En annan faktor är **algoritmens utformning och design**. Utvecklare tar en rad beslut när de skapar AI-system: vilka variabler som ska räknas in, hur modellen ska viktas, vilka mål den ska optimera mot, osv. Om dessa beslut inte är genomtänkta ur rättvisesynpunkt kan bias uppstå trots att datan i sig kanske verkar "neutral". Till exempel kan en programmerare omedvetet väga in en faktor (t.ex. postnummer, inkomst eller ordförråd) som korrelerar med kön eller etnicitet, så att algoritmens beslut missgynnar en viss grupp ([AI Bias Examples | IBM](#)). Små programmeringsval baserade på utvecklarens egna föreställningar – medvetna eller omedvetna – kan alltså få stora konsekvenser. Forskning från NIST (den amerikanska standardiseringsmyndigheten) betonar att mänskliga, institutionella och samhällsliga faktorer är *betydande* källor till AI-bias, och ofta förbises ([AI Bias Examples | IBM](#)). Med andra ord: bias handlar inte bara om tekniken i sig, utan också om de människor och strukturer som skapar tekniken.

Sammanfattningsvis uppstår AI-bias genom en kombination av skev data, designval i algoritmen och mänskliga beslut i utvecklingsprocessen. Generativa AI-modeller som ChatGPT är ett bra exempel – de är tränade på internetdata *som i sig* innehåller allt från språkliga egenheter till fördomar. Om en viss typ av åsikt eller framställning dominerar i träningsdatan, kommer modellen att reproducera det mönstret. Bias kan alltså krypa in i AI på många sätt innan systemet ens tas i bruk.

Risker och konsekvenser av AI-bias

Konsekvenserna av partiska AI-system kan bli allvarliga. För det första kan **ojämlikheter förstärkas**. När AI okritiskt återupprepar historiska mönster av exkludering riskerar det att cementera dessa i nya beslut. AI-bias tenderar att *skala upp* diskriminering – det som tidigare skedde i mindre skala kan ett automatiserat system göra i stor skala, snabbare. Ett snedvridet system kan till exempel gallra bort kvalificerade kandidater från en minoritetsgrupp vid anställning (som i Amazon-fallet) eller ge personer olika lånevillkor baserat på etnicitet eller kön, trots likvärdig ekonomi. På samhällsnivå innebär det att redan utsatta grupper kan få ännu svårare att delta på lika villkor i ekonomin och samhället ([AI Bias Examples | IBM](#)). Om bias lämnas oadresserat kan det "hårdkoda" fördomar i de beslut som påverkar människors liv, vilket urholkar tilliten till systemen. Allmänheten – särskilt marginaliserade grupper – förlorar förtroende för AI om det visar sig ge systematiskt orättvisa resultat ([AI Bias Examples | IBM](#)).

Spridning av stereotyper är en annan risk. Om en generativ AI-model konsekvent skriver om, säg, ingenjörer som män och sjuksköterskor som kvinnor, så reproducerar den gamla stereotyper och ger dem sken av att vara normala eller faktabaserade. Detta kan påverka hur unga uppfattar yrken och sig själva. Ett konkret exempel är annonsalgoritmer: en studie vid Carnegie Mellon University fann att Googles annonssystem visade höglavlönade jobbannonser betydligt oftare för män än för kvinnor ([AI Bias Examples | IBM](#)). Algoritmen förstärkte därmed en stereotyp föreställning (att toppjobben är för män) och samtidigt missgynnades kvinnor som inte fick se samma möjligheter. På liknande sätt har generativa bild-AI som *Midjourney* kritiserats för att förstärka stereotyper – i ett experiment genererades bilder på äldre yrkesutövare, och samtliga äldre personer avbildades som män, vilket implicerar att "äldre med erfarenhet" automatiskt kopplas till manligt kön ([AI Bias Examples | IBM](#)). Genom att upprepa sådana snedvridningar riskerar AI-system att normalisera fördomar och påverka våra uppfattningar utan att vi märker det.

AI-bias kan också leda till **felaktiga eller orättvisa beslut** inom kritiska områden. Inom rättsväsendet har t.ex. **prediktiv polisiära system** använts för att förutspå var brottslighet kan ske. Dessa system bygger ofta på historisk data över arresteringar. Forskning och granskningar har visat att de tenderar att skicka polisresurser till områden som historiskt varit överbevakade, ofta områden där minoriteter bor, medan brott i andra områden förbises ([AI Bias Examples | IBM](#)). Algoritmen *tror* sig vara objektiv men förstärker egentligen existerande mönster av profilering. Konsekvensen kan bli en ond cirkel: mer övervakning i redan utsatta områden leder till fler registrerade incidenter där, vilket ytterligare bekräftar algoritmens bias. Även inom rättssystemets beslutsstöd har bias uppdagats – ett känt fall är riskbedömningsverktyget COMPAS i USA som systematiskt överskattade risken för återfall i brott för svarta jämfört med vita, enligt en undersökning av ProPublica ([AI Bias Examples | IBM](#)). Det visar hur AI-beslut kan vara långt ifrån neutrala och hur viktigt det är att granska dem kritiskt.

Förutom de direkta konsekvenserna kan AI-bias *indirekt* skada samhället genom att undergräva **förtroendet för teknik och information**. Om elever upptäcker att en AI-chattbot ger partiska svar, eller om medborgare märker att endast vissa perspektiv kommer fram i flöden och sökresultat, riskerar de att bli misstroiska mot digital information generellt. Det ställer större krav på **källkritik och kritiskt tänkande**. Vi kan inte anta att "datorn har rätt" –

tvärtom måste vi vara medvetna om att även sofistikerade AI-system kan ha blinda fläckar och inbyggda värderingar.

Mänskliga bias: våra egna filter och fällor

Innan vi pekar finger åt algoritmerna är det värt att minnas att vi människor själva är fulla av **kognitiva bias**. En kognitiv bias är en systematisk tankevrupa – ett slags genväg i hjärnan som kan få oss att missa information eller bedöma verkligheten skevt. Dessa bias påverkar hur vi tolkar information, fattar beslut och även hur vi interagerar med teknik. Här är några viktiga kognitiva bias som är relevanta i mötet med AI:

- **Bekräftelsebias (confirmation bias):** Tendensen att leta efter och tro på information som *bekräftar* det vi redan tycker, och ignorera sådant som motsäger våra uppfattningar ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). Vi människor gillar när vår världsbild bekräftas – det känns bekvämt. I kontakt med AI kan detta bias göra att vi omedvetet formulerar frågor eller tolkar svar på ett sätt som stärker det vi redan tror. Till exempel kanske en elev ställer ledande frågor till en AI ("Visst är det så att...") och accepterar svaret okritiskt om det ligger i linje med elevens egen åsikt. **Bekräftelsebias** bidrar också till så kallade filterbubblor online – vi söker oss till källor som håller med oss, och algoritmerna serverar mer av samma, vilket förstärker vår uppfattning att vi har rätt hela tiden.
- **Auktoritetsbias (authority bias):** Tendensen att tillskriva *auktoriteter* (experter, ledare – eller för den delen datorer) extra stor trovärdighet, ofta utan kritisk granskning (). Vi har en benägenhet att tro på information som kommer från en källa vi uppfattar som kunnig eller officiell. När en AI presenteras som högteknologisk och objektiv kan auktoritetsbias kicka in – man litat på svaret just för att "det kom från AI:n". Elever kan till exempel anta att "om Wikipedia eller ChatGPT säger det, så stämmer det", istället för att dubbelkolla. Auktoritetsbias gör att vi ibland **övervärderar** AI:ts påståenden och glömmar att även dessa måste ifrågasättas och jämföras med andra källor.
- **Förankringsbias (anchoring bias):** Tendensen att låta den **första** informationen vi får fungera som ett ankare som påverkar vår bedömning av senare information ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). Om vi får höra en viss siffra eller uppgift först, utgår vi omedvetet från den när vi sedan tar ställning. I en AI-kontext kan detta innebära att det första svar man får från en sökmotor eller chatbot sätter agendan för vad man tror är sant. Till exempel kan den första träffen på Google, eller det första förslaget som en AI ger på en lösning, verka mest korrekt bara för att det kom först. **Förankringseffekten** kan göra oss mindre benägna att ompröva initial information – även om den skulle vara felaktig eller ofullständig.
- **Tillgänglighetsbias (availability bias):** Tendensen att överskatta värdet eller sannolikheten av något som lätt kommer till minnet, dvs. information som är *tillgänglig* i vårt medvetande ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). Vi ger alltså större vikt åt exempel vi nyligen hört om eller som var särskilt slående. I praktiken kan detta bias få oss att tro att något är vanligare eller viktigare än det är, bara för att vi råkar se det ofta. I vårt digitala informationsflöde kan tillgänglighetsbias uppstå när dramatiska nyheter eller klickvänliga historier sprids brett – de finns "överallt" i flödet och känns därför representativa, fast de kanske är extremfall. När elever googlar kan de första resultaten (eller de mest sensationella) etsa sig fast i minnet och färga deras uppfattning av ämnet. **Tillgänglighetsbias** gör att

vi riskerar att basera våra slutsatser på hur lätt något dyker upp i huvudet, snarare än på hur vanligt eller relevant det faktiskt är.

- **Halo-effekten:** Tendensen att låta en positiv egenskap hos en person eller företeelse spilla över så att vi antar att *allt* hos denna är bra ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). Om vi imponeras av en egenskap, som att någon talar väl eller att ett verktyg levererar ett bra resultat i ett fall, så antar vi lätt att de är pålitliga även i andra avseenden. I fallet med AI kan halo-effekten innebära att om en AI-tjänst (eller teknikföretaget bakom den) har ett gott rykte eller tidigare gett bra svar, så ifrågasätter vi inte dess utsagor inom andra områden. Till exempel: om en elev fått ett jättebra svar på en historiafråga från en AI, kanske hen tror blint på nästa svar AI:n ger om biologi – trots att expertisen kan variera. Halo-effekten kan alltså ge AI-system en oförtjänt "gloria" av ofelbarhet i användarens ögon.

Hur AI förstärker våra mänskliga bias

Det paradoxala med AI är att även om maskinerna i sig inte har några känslor eller ideologier, kan de *förstärka* de bias vi människor redan bär på. Dels sker detta genom själva algoritmernas urval och personalisering av information, dels genom hur vi psykologiskt reagerar på AI:s sätt att presentera information.

En tydlig mekanism är **personaliseringseffekten** på internet. Moderna AI-drivna plattformar (som sociala medier, nyhetsflöden och sökmotorer) anpassar innehållet efter våra tidigare klick och preferenser. Vi får mer av det vi redan visat intresse för. Konsekvensen blir att informationen vi möter online ofta bekräftar våra befintliga attityder, medan motstridiga perspektiv filtreras bort – upp till 20% av relevant information som inte stämmer med våra preferenser kan helt missa oss ([Extended confirmation bias: When the mind leaks into algorithms - Open for Debate - Cardiff University](#)). Detta kan skapa **filterbubblor**: vi blir inneslutna i ett eko-rum av likasinnade åsikter. AI:t förstärker här vår naturliga bekräftelsebias genom algoritmiskt urval – vi matas med innehåll som får oss att nicka instämmande, vilket ytterligare övertygar oss om att våra uppfattningar är de enda rimliga ([Extended confirmation bias: When the mind leaks into algorithms - Open for Debate - Cardiff University](#)). På så vis kan teknik som skulle bredda våra vyer ironiskt nog göra dem smalare, genom att den "förlänger" vår bekräftelsebias in i det digitala rummet.

AI-system kan också spela på **auktoritetsbias**. Många AI-tjänster presenterar sina resultat med stor självsäkerhet och precision. En chatbot kan formulera ett svar i klarspråk, utan att tveka eller redovisa osäkerhet – ibland utan att ange källor – vilket kan ge intrycket av att svaret är auktoritativt och komplett. Om vi människor redan har en tendens att lita på bestämda utsagor från auktoriteter, är risken att vi inte ifrågasätter AI:ts svar ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)). Generativa AI-modeller som skriver välformulerade och sammanhängande texter utnyttjar just detta: de *låter* som experter, och då klickar auktoritetsbias in så att lärare och elever kanske accepterar svaret utan vidare granskning ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)). Vi glömmer att modellen saknar verklig förståelse och kan ha fabricerat information (så kallade hallucinationer) – formen lurar oss att fokusera mer på *hur* svaret presenteras än *om* innehållet faktiskt stämmer.

Förankringsbias kan förstärkas av AI genom att den första informationen som en algoritmer får får oproportionerligt stor vikt. Tänk på när man söker på Google: de översta resultaten får

störst uppmärksamhet. Om de råkar vara missvisande kan de ändå prägla användarens bild av sanningen. Likaså kan den första vägvisningen från en GPS få oss att hålla fast vid en rutt, eller det första svaret från en AI-assistent få oss att utgå från en viss ståndpunkt. Studier pekar på att användare sällan ifrågasätter första förslaget en algoritm ger ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)) – det fungerar som ett ankare. I en dialog med en chatbot kan detta bli tydligt: om chatboten inleder med en viss lösning eller påstående, kan följdfrågorna vi ställer hamna inom samma tankebanor, vilket *bekräftat* det initiala svaret (även om det kanske borde ifrågasättas). Algoritmen må vara opartisk i grunden, men vårt eget sinne hakar upp sig på förstaintrycket som AI:n gav.

När det gäller **tillgänglighetsbias** bidrar AI-drivna informationsflöden till att vissa nyheter eller fakta blir väldigt framträdande (trendar, går viralt) och därmed känns mer relevanta än de kanske är. Sociala medie-algoritmer prioriterar ofta innehåll som får många reaktioner – och det är inte alltid det mest korrekta innehållet, utan det som väcker starkast känslor. Sådan content ser vi om och om igen, vilket gör det lättillgängligt i minnet. När vi sedan bedömer vad som är sant eller viktigt, finns dessa färskna intryck genast till hands och påverkar oss. Ett exempel är om ett fåtal uppseendeväckande händelser (kanske genererade eller spridda av AI) dominerar flödet; då kan elever få en skev bild av hur vanligt något är – "alla pratar om X, alltså måste X hända hela tiden". AI förstärker här vår tillgänglighetsbias genom att selektivt höja fram viss information så att den känns mer allmän än den egentligen är.

Slutligen kan **halo-effekten** göra oss mindre kritiska till AI. Om användare har positiva erfarenheter av en AI-tjänst – den kanske hjälpte dem lösa ett problem snabbt – kan de få ett generellt förtroende för tjänsten. Då finns risken att de även litar på AI:ts utsagor i frågor där AI:n egentligen inte är pålitlig. Till exempel kanske en lärare märker att en översättnings-AI fungerar utmärkt i språkundervisningen (positivt intryck), och börjar anta att samma AI ger korrekta historiska fakta – trots att faktagranskning inte är dess starka sida. Halo-effekten förstärks av AI:ns ofta *mänskliga* framtoning; vi börjar nästan betrakta AI:n som en kompetent kollega inom alla områden. Detta illustrerar hur våra mänskliga bias kan göra att vi övertolkar AI-systemens kompetens och objektivitet.

Avslutande tankar

Att förstå hur AI-bias fungerar hjälper oss lärare att guida elever i ett landskap där artificiell intelligens blir allt vanligare. Bias inom AI visar att teknik inte är ofelbar eller neutral bara för att den är avancerad – tvärtom behöver vi vara **mer** källkritiska när vi använder AI-genererad information. Samtidigt påminner de kognitiva biasen oss om att vi har våra egna filter som färgar hur vi tar till oss AI:ts output. Genom att diskutera konkreta exempel (som partiska rekryteringsalgoritmer och ansiktsgenkänning) och de psykologiska mekanismerna bakom (bekräftelse, auktoritet, förankring, tillgänglighet, halo) kan vi hjälpa elever att utveckla ett **kritiskt förhållningssätt**. Målet är inte att misstänkliggöra all AI, utan att ge eleverna verktyg att genomskåda både maskinens och människans bias. På så vis kan nästa generation navigera en värld av AI med öppna ögon – medvetna om både möjligheterna och fallgroparna – och bli kunniga digitala medborgare som tänker själva.

Källor:

1. Jeffrey Dastin. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters ([Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters](#))
2. Quartz via MIT News. "...Gender Shades project... Buolamwini and Gebru found that commercial facial recognition software was significantly less accurate for darker-skinned women than for lighter-skinned men." ([Study finds gender and skin-type bias in commercial artificial-intelligence systems | MIT News | Massachusetts Institute of Technology](#))
3. NEA (National Education Association). "Scientists from MIT found that a language model thinks that 'flight attendant,' 'secretary,' and 'physician's assistant' are feminine jobs, while 'fisherman,' 'lawyer,' and 'judge' are masculine." ([Does AI Have a Bias Problem? | NEA](#))
4. IBM Data and AI Team. "AI bias... refers to AI systems that produce biased results that reflect and perpetuate human biases within a society, including historical and current social inequality." ([AI Bias Examples | IBM](#)) ([AI Bias Examples | IBM](#))
5. IBM Data and AI Team. *Examples of bias in AI: Ansiktsgenökning som tränats mest på vita ansikten fungerar sämre för andra grupper* ([AI Bias Examples | IBM](#)); utvecklades omedvetna val (t.ex. viktning av data) kan leda till diskriminerande utfall ([AI Bias Examples | IBM](#)); prediktiv polising repeterar historiska mönster av övervakning ([AI Bias Examples | IBM](#)); Googles annonsalgoritm visade toppjobb oftare för män ([AI Bias Examples | IBM](#)); Midjourney genererade äldre experter som enbart män ([AI Bias Examples | IBM](#)).
6. Digiteket – "Bias, eller du är inte så objektiv som du tror!" (2020). Definitioner av kognitiva bias: **Bekräftelsebias:** vi tar lättare till oss info som bekräftar våra åsikter ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). **Haloeffekt:** en positiv egenskap får oss att tillskriva fler positiva egenskaper ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). **Tillgänglighetsbias:** vi ger större tyngd åt information vi minns väl (ofta ny eller uppseendeväckande) ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)). **Förankringseffekt:** den första informationen vi får utgör ett ankare för senare bedömningar ([Bias, eller du är inte så objektiv som du tror! - Digiteket](#)).
7. IMPRODOVA training material. *Exempel på fördomar.* **Auktoritetsbias:** vi tenderar att tro på och följa auktoriteters uttalanden även om de kan vara fel (). **Konfirmeringsbias:** vi söker information som stödjer våra övertygelser och ignorerar motsägande info ().
8. Uwe Peters (Cardiff University). "Extended confirmation bias: When the mind leaks into algorithms." (2022). Personalisering av innehåll online gör att vi missar upp till 20% av relevant information som inte stämmer med våra preferenser, vilket främjar politisk polarisering och "filterbubblor" ([Extended confirmation bias: When the mind leaks into algorithms - Open for Debate - Cardiff University](#)).
9. Frontier Economics. "Responsible Generative AI: Both humans and algorithms have biases." (2023). **Authority bias:** människor litar på information från auktoritativa källor, inklusive datorapplikationer, utan att ifrågasätta korrektheten – GenAI ger självsäkra svar utan källhänvisning, och användare kan då underlåta att ifrågasätta ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)). **Anchoring bias:** vi påverkas av första uppgiften vi får – användare kan fastna vid första svaret en GenAI ger, även om det är missvisande ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)). **Confirmation bias:** vi söker sådan info som bekräftar våra åsikter – GenAI som levererar svar i linje med användarens förutfattade meningar accepteras lättare, vilket

kan skapa en ond cirkel av bekräftelse ([Responsible Generative AI: Both humans and algorithms have biases | Frontier Economics](#)).

Transparens och förklarbarhet i AI – en nyckel till källkritik

I en tid där AI blir en allt större del av våra liv – från sociala medier och sökmotorer till automatiserade beslutsprocesser i arbetslivet – är det avgörande att vi förstår hur dessa system fungerar. Två centrala begrepp inom detta område är **transparens** och **förklarbarhet**. Men vad betyder de egentligen, och varför spelar de en roll för vår förmåga att vara källkritiska?

Transparens handlar om hur mycket information vi har om hur ett AI-system är byggt, vilken data det använder och hur det fattar beslut. Förklarbarhet är något mer specifikt: det handlar om att vi ska kunna förstå och förutspå AI:s beslut. Om vi inte vet hur eller varför en AI ger oss ett visst svar, hur kan vi då avgöra om det är trovärdigt?

Bristen på transparens och förklarbarhet gör oss sämre rustade att vara källkritiska. Om vi får ett nyhetsinlägg i vårt sociala medieflöde, vet vi då varför det visades just för oss? Om vi får en rekommendation om en bok eller en film, förstår vi på vilka grunder den rekommenderades? Om en AI-chattbot svarar på en fråga vi ställer, kan vi förstå varför den svarade som den gjorde?

Exempel på problem med bristande transparens och förklarbarhet

1. **AI-genererade nyheter:** Många får idag sina nyheter via AI-drivna flöden på sociala medier. Dessa flöden anpassas efter vår tidigare aktivitet, vilket kan leda till att vi bara ser vissa typer av nyheter och missar andra perspektiv. Men algoritmerna som styr detta är sällan synliga för oss.
2. **Rekommendationssystem:** Streamingplattformar som Netflix och Spotify använder AI för att rekommendera filmer och musik. Men vad bygger dessa rekommendationer på? Är det utifrån vad vi faktiskt gillar, eller för att vissa innehållsproducenter har betalat för att synas mer?
3. **AI-chattbotar:** Många använder AI-chattbotar för att få svar på frågor, men svaren kan vara påverkade av den data chattboten tränats på. Om en AI svarar felaktigt på en fråga, hur kan vi avgöra det? Hur kan vi förstå vilka källor som använts?

Att kunna undervisa om transparens och förklarbarhet innebär att hjälpa elever att utveckla ett kritiskt förhållningssätt till AI-genererad information. I nästa avsnitt tittar vi närmare på vad transparens egentligen innebär och varför det är en central aspekt av källkritik i AI-sammanhang.

Vad är transparens i AI – och varför är det viktigt?

När vi pratar om transparens i AI handlar det om hur mycket vi kan förstå och granska ett AI-system. Det innebär att kunna få insyn i hur AI:n fungerar, vilken data den har tränats på och vilka begränsningar den har.

Varför är detta viktigt? För att kunna vara källkritiska mot AI-genererad information måste vi veta var den kommer ifrån och hur den har skapats. Om en AI-chattbot ger oss ett svar, hur kan vi veta om det är baserat på trovärdiga källor? Om en AI-bildgenerator skapar en bild, hur

kan vi veta vad som har påverkat resultatet? Transparens hjälper oss att förstå vad vi kan lita på – och vad vi bör ifrågasätta.

Tänk dig två olika situationer:

1. Du frågar en lärare varför du fått ett visst betyg och läraren förklarar tydligt hur bedömningen gjorts.
2. Du får samma betyg men läraren säger bara: *"Det är ett komplext system, men du kan lita på att det är rätt."*

I det första fallet finns transparens – du får insyn i hur beslutet fattades. I det andra fallet finns ingen transparens – du förväntas lita på systemet utan att förstå det. Många AI-system fungerar just på det andra sättet: de ger svar, men vi vet inte alltid varför.

Det finns tre viktiga aspekter av transparens i AI:

1. Träningsdata och källor – var kommer informationen ifrån?

Generativa AI-modeller, som ChatGPT och bildgeneratorer, lär sig genom att analysera enorma mängder text och bilder från internet. Men vi vet ofta inte exakt vilka källor som har använts, vilket påverkar hur AI:n svarar.

Exempel:

- En AI-chattbot som tränats mest på Wikipedia kan ge andra svar än en AI som också tränats på vetenskapliga artiklar eller tidningsartiklar.
- En AI som genererar bilder kan ha tränats på vissa typer av bilder mer än andra, vilket påverkar hur den skapar nya bilder.

Varför är detta viktigt?

Om vi inte vet var AI får sin information ifrån, hur kan vi då avgöra om den är pålitlig? En AI kan låta självsäker även när den har fel – men utan transparens i källorna är det svårt att upptäcka felaktigheter.

2. Hur AI:n bearbetar frågor – hur skapas svaren?

AI fungerar inte som en människa som tänker och resonerar. I stället bygger generativa AI-modeller svar genom att förutspå vilket ord som mest sannolikt bör komma näst i en mening, baserat på all text den har tränats på.

Exempel:

- Frågar du en AI "Vem uppfann glödlampan?" kan den svara "Thomas Edison" eftersom Edison ofta nämns i samband med glödlampan. Men den kanske missar att Humphry Davy och Joseph Swan också hade viktiga roller.
- Ber du en AI skapa en bild av en "historisk kung" kan den mest sannolikt skapa en bild av en europeisk kung, eftersom det finns fler sådana bilder i dess träningsdata.

💡 **Varför är detta viktigt?**

AI hittar inte "sanningen" – den skapar sannolika svar utifrån sin träningsdata. Om vi inte förstår detta kan vi lätt tro att AI har en objektiv förståelse av världen, när den i själva verket bara gissar utifrån tidigare mönster.

3. Begränsningar och kunskapsstopp – vad kan AI inte veta?

AI har begränsningar, men de är ofta osynliga för användaren. En av de största begränsningarna är att AI-modeller har ett "kunskapsstopp", vilket innebär att de inte har information om händelser efter ett visst datum.

📌 **Exempel:**

- Frågar du en AI om "vem som vann fotbolls-VM 2026" kommer den att sakna information om detta om den bara tränats på data fram till 2023. Men den kanske ändå gissar och ger ett svar som låter trovärdigt.
- Om du frågar en AI om en aktuell politisk händelse kan den basera sitt svar på gammal information, vilket kan leda till missvisande slutsatser.

💡 **Varför är detta viktigt?**

Om vi inte känner till AI:ns begränsningar kan vi tro att den alltid har uppdaterad och korrekt information – när den i själva verket kan vara inaktuell eller felaktig.

Sammanfattning: Hur kan vi använda transparens för att undervisa i källkritik?

För att hjälpa elever att förstå och ifrågasätta AI-genererad information kan vi använda dessa tre frågor:

1. **Var kommer informationen ifrån?** (Vilka källor har AI:n tränats på?)
2. **Hur har AI:n skapat sitt svar?** (Är det en återgivning av fakta eller en sannolikhetsbaserad gissning?)
3. **Vilka begränsningar finns?** (Har AI:n uppdaterad information?)

Genom att ställa dessa frågor kan vi lära elever att tänka kritiskt kring AI:s svar och förstå att AI inte är en objektiv källa – utan en avancerad textgenerator med både styrkor och svagheter.

Vad är förklarbarhet i AI – och varför är det viktigt?

Förklarbarhet i AI handlar om att vi kan förstå **varför** en AI ger ett visst svar eller gör en viss bedömning. Det är en sak att veta att en AI har gett oss ett visst resultat – men en annan sak att kunna förstå hur den kom fram till det.

Om vi använder AI i undervisningen eller låter elever använda AI-verktyg, behöver vi kunna hjälpa dem att förstå när och hur de kan lita på AI:s svar. En AI kan låta självsäker, men om vi inte vet hur den kommit fram till ett svar, hur kan vi då avgöra om det är trovärdigt?

Låt oss ta ett exempel:

✦ **Tänk dig att en elev frågar en AI-chattbot:**

"Varför säger vissa att Napoleon var kort?"

Chattboten svarar:

"Napoleon var 1,69 m lång, vilket var en normal längd för hans tid. Men rykten om att han var kort kan ha spridits på grund av propaganda och förvirring kring franska och brittiska måttsystem."

Om AI:n har hög förklarbarhet, kan den också ge en motivering till sitt svar:

"Denna information kommer från historiska källor som beskriver Napoleons längd i relation till samtida standarder. Jag har sammanställt det baserat på vad som är mest sannolikt enligt min träningsdata."

Om AI:n har låg förklarbarhet, vet vi bara att den gett oss ett svar – men vi har ingen insyn i varför den formulerat det just så.

Olika typer av förklarbarhet i AI

Det finns två huvudtyper av förklarbarhet i AI:

1. **Post-hoc-förklaringar** – Förklaringar i efterhand
2. **Intrinsik förklarbarhet** – Modeller som är byggda för att vara förståeliga från början

1. Post-hoc-förklaringar – när AI försöker förklara i efterhand

Många AI-modeller är **svarta lådor** – vi kan se indata och utdata, men inte exakt hur AI:n kommit fram till sitt resultat. För att göra dessa system mer begripliga kan vi använda post-hoc-förklaringar, det vill säga att AI:n försöker motivera sina beslut i efterhand.

✦ **Exempel: AI som analyserar bilder**

Tänk dig att du använder en AI-bildgenerator och ber den skapa en bild av en "vetenskapsman". Bilden visar en äldre man med glasögon och labbrock. Varför blev det så?

- Om AI:n har post-hoc-förklaringar kan den berätta att de flesta vetenskapsmän i dess träningsdata såg ut så, vilket påverkat dess bildgenerering.
- Om AI:n saknar förklarbarhet, får du bara bilden utan någon insyn i varför den valde just dessa drag.

✦ **Exempel: AI som analyserar CV:n**

Tänk dig att ett rekryteringsverktyg använder AI för att granska jobbansökningar. En kandidat blir bortsorterad. Varför?

- Med **post-hoc-förklaringar** kan AI:n förklara att den valt bort kandidaten eftersom denne saknar en viss erfarenhet som de flesta tidigare anställda hade.
- Utan förklarbarhet kan rekryteraren bara lita på AI:ns beslut utan att veta varför – vilket kan leda till omedveten diskriminering eller andra problem.

💡 **Problemet med post-hoc-förklaringar:**

De kan vara hjälpsamma, men de är inte alltid helt sanningsenliga. Ibland kan en AI fabricera en förklaring som låter rimlig, men som inte exakt speglar hur beslutet faktiskt fattades.

2. Intrinsik förklarbarhet – AI som är begriplig från början

Vissa AI-modeller är konstruerade för att vara begripliga redan från början. Dessa system använder enklare beslutsmodeller som kan redovisas i tydliga steg.

✦ **Exempel: En AI som hjälper elever att rätta stavfel**

Om AI:n är byggd med intrinsik förklarbarhet kan den ge en stegvis förklaring:

- "Jag föreslår att du ändrar 'til' till 'till' eftersom ordet 'til' inte finns i svenska språket."

Här kan vi se exakt **varför** AI:n föreslår en ändring.

✦ **Exempel: En AI som rekommenderar böcker**

En AI som föreslår böcker kan vara byggd på en enkel logik:

- "Jag rekommenderar den här boken eftersom du tidigare har läst liknande böcker av samma författare."

I dessa fall är AI:s beslutsprocess lätt att förstå, vilket gör den mer förklarbar.

💡 **Problemet med intrinsik förklarbarhet:**

Dessa AI-system är ofta enklare och mindre kraftfulla än de stora generativa AI-modellerna, som t.ex. ChatGPT. Ju mer komplex en AI blir, desto svårare blir det att bygga in förklarbarhet i själva modellen.

Varför är förklarbarhet viktigt för källkritik?

Om vi inte kan förstå varför en AI ger oss ett visst svar, kan vi heller inte avgöra om vi kan lita på det. Lärare och elever behöver därför kunna ställa frågor som:

- ✓ Hur kom AI:n fram till det här svaret?
- ✓ Vilka källor har använts?
- ✓ Kan AI:n förklara sitt resonemang?

Om vi lär elever att efterfråga förklaringar och ifrågasätta AI:s svar, blir de bättre på att navigera en värld där AI genererar allt mer information.

Varför är transparens och förklarbarhet viktiga för källkritik?

I en tid där AI genererar allt från nyheter och bilder till uppsatser och översättningar är det avgörande att vi inte bara accepterar AI:s svar som sanningar. AI kan låta självsäker och auktoritativ, men utan transparens och förklarbarhet har vi ingen aning om hur den kommit fram till sina svar.

När vi undervisar elever i källkritik har vi länge lärt dem att fråga:

- 🤖 *Vem har skrivit texten?*
- 🤖 *Vilka källor har använts?*
- 🤖 *Finns det någon bakomliggande agenda?*

Men hur gör vi detta när informationen inte är skriven av en människa, utan av en AI? Det är här transparens och förklarbarhet blir viktiga verktyg.

1. Undvika "blint förtroende" för AI

AI ger svar på ett sätt som låter övertygande. Den kan formulera sig med säkerhet, även när den har fel. Om elever inte lär sig att vara kritiska kan de lätt ta AI:s svar som fakta – trots att de kan vara felaktiga eller vilseledande.

📌 Exempel: AI som expert på Shakespeare?

En elev ber en AI-chattbot om hjälp med att analysera *Hamlet* och får svaret:
"I slutet av pjäsen överlever Hamlet och blir kung av Danmark."

Problemet? Det är helt fel. Men eftersom AI:n formulerar sig med självsäkerhet kan eleven lätt tro på det. Utan transparens (vilka källor AI:n använt) och förklarbarhet (hur AI:n kommit fram till svaret) är det svårt att ifrågasätta.

💡 Lärdom:

Vi måste lära elever att alltid dubbelkolla AI:s svar och inte ta dem för givna. En AI kan inte "veta" något – den kan bara skapa sannolika svar baserade på sin träningsdata.

2. Identifiera bias och felaktigheter i AI

AI är inte neutral. Den påverkas av den data den tränats på, vilket innebär att den kan förstärka existerande fördomar och skevheter. Om vi inte förstår hur en AI fungerar kan vi omedvetet sprida dess bias vidare.

✦ Exempel: AI och bildgenerering

En bildgenerator får instruktionen: *"Skapa en bild av en läkare."*
Resultatet? Nästan uteslutande bilder av män.

Varför? För att AI:n har tränats på en stor mängd bilder där läkare oftare har framställts som män. Om vi inte har insyn i träningsdata och AI:ns begränsningar, kan vi tro att det är en objektiv sanning att läkare oftast är män.

✦ Exempel: AI och språk

En AI-chattbot får frågan: *"Beskriv en typisk ledare."*
Den svarar: *"En stark, beslutsam och strategisk man."*

Detta visar en inbyggd bias – en skevhet som speglar hur samhället ofta pratar om ledare. Om vi inte har verktyg för att förstå och ifrågasätta sådana svar, kan vi omedvetet befästa dessa föreställningar.

💡 Lärdom:

Elever behöver lära sig att AI inte är objektiv. Genom att förstå transparens och förklarbarhet kan de börja granska AI:s svar och fråga sig: *Finns det någon bias här?*

3. Lära elever att granska och ifrågasätta AI:s svar

För att AI ska bli ett användbart verktyg i skolan måste vi lära elever att ställa kritiska frågor om AI-genererad information.

✦ Övning: AI som nyhetskälla

- Be eleverna be en AI-chattbot sammanfatta en aktuell händelse.
- Jämför AI:s svar med flera olika nyhetskällor.
- Diskutera: Vilka skillnader finns? Finns det något som saknas?

✦ Övning: Utmana AI:s bias

- Be en AI generera bilder av olika yrken och analysera resultaten.
- Diskutera: Speglar bilderna verkligheten? Vilka stereotyper finns?

✦ Övning: Kan AI förklara sig?

- Be AI:n förklara hur den kom fram till ett svar.
- Diskutera: Var svaret trovärdigt? Saknades något?

Lärdom:

Genom att aktivt träna elever i att granska AI:s svar blir de bättre rustade att använda AI på ett kritiskt och medvetet sätt.

Sammanfattning

Transparens och förklarbarhet är nycklar till att förstå och ifrågasätta AI-genererad information. Om vi lär elever att:

- ✓ *Inte lita blint på AI:s svar*
- ✓ *Identifiera bias och felaktigheter*
- ✓ *Ställa kritiska frågor till AI*

... blir de bättre rustade att navigera en värld där AI har allt större inflytande över vilken information vi ser och hur vi tolkar den.

5. Exempel från undervisningen – hur kan vi konkret arbeta med detta i klassrummet?

Att undervisa om transparens och förklarbarhet i AI behöver inte vara tekniskt eller komplicerat. Genom praktiska övningar kan vi låta eleverna uppleva hur AI fungerar, ifrågasätta dess svar och förstå varför olika AI-system kan ge olika resultat.

Här är tre övningar som kan användas i undervisningen:

Övning 1: AI som domare – varför ger olika AI:er olika svar?

 **Syfte:** Lära elever att AI-modeller kan ge olika svar beroende på vilken data de tränats på och hur de är programmerade.

Instruktion:

1. Låt eleverna ställa samma fråga till två olika AI-chattbotar (t.ex. ChatGPT och Bing Chat).
2. Jämför svaren – är de identiska, eller finns det skillnader?
3. Diskutera:
 - Varför kan samma fråga ge olika svar?
 - Vilka källor kan ha påverkat AI:ns svar?
 - Hur kan vi avgöra vilket svar som är mest tillförlitligt?

Exempel:

Eleverna kan fråga två AI:er: "Vad är orsakerna till klimatförändringar?"

Om en AI ger ett detaljerat svar med vetenskapliga referenser, medan en annan ger ett kort och vagt svar, kan detta leda till en diskussion om transparens och källor.

💡 **Lärdom:** AI är inte en objektiv sanningskälla – svaren påverkas av hur de tränats.

Övning 2: Förklara en AI-process – eleverna får vara "neuroner"

🔴 **Syfte:** Hjälpa elever att förstå hur en AI-modell bearbetar information genom att låta dem själva simulera en AI-process.

Instruktion:

1. Dela in klassen i små grupper och ge varje elev en enkel regel att följa, t.ex.:
 - "Om någon säger ett ord med bokstaven A, skicka det vidare till nästa person."
 - "Om ordet har fler än fem bokstäver, stoppa det här."
2. Starta processen genom att mata in ett ord i gruppen. Eleverna följer sina regler och "bearbetar" informationen steg för steg.
3. Diskutera:
 - Hur liknar detta hur en AI bearbetar information?
 - Vad händer om en "neuron" har en dålig regel? (t.ex. alltid stoppar ord med bokstaven E)
 - Hur påverkar små förändringar i reglerna slutresultatet?

💡 **Lärdom:** AI gör inga medvetna val – den följer regler och mönster från sin träningsdata.

Diskussion: När måste vi kunna förstå AI – och när kan vi acceptera en "svart låda"?

🔴 **Syfte:** Få elever att reflektera över när transparens och förklarbarhet är avgörande – och när vi faktiskt kan acceptera att AI fungerar som en "svart låda".

Diskussionsfrågor:

- När är det viktigt att förstå exakt hur AI fungerar? (T.ex. vid medicinska diagnoser, domstolsbeslut, kreditbedömningar)
- När kan vi acceptera att AI fungerar utan att vi förstår detaljerna? (T.ex. när vi får rekommenderade filmer på Netflix)
- Vad kan vara riskerna med att AI används i beslut vi inte kan granska?

💡 **Lärdom:** Eleverna lär sig att ifrågasätta när AI-beslut måste vara transparenta och när vi kan använda AI som ett verktyg utan att förstå exakt hur det fungerar.

6. Sammanfattning och viktiga lärdomar

AI har en enorm potential, men för att använda det på ett ansvarsfullt sätt behöver vi förstå dess begränsningar.

✓ **Transparens och förklarbarhet hjälper oss att vara källkritiska**

Om vi inte vet hur en AI fungerar eller var den får sin information ifrån, blir det svårt att lita på dess svar.

✓ **Vi ska inte lita blint på AI**

AI ger svar med auktoritet, men det betyder inte att de alltid är korrekta. Elever behöver träna sig i att ifrågasätta och dubbelkolla AI:s information.

✓ **Lärare behöver inte vara tekniska experter**

Det viktigaste är inte att förstå exakt hur AI fungerar på detaljnivå, utan att kunna hjälpa elever att ställa rätt frågor och analysera AI:s svar kritiskt.

Genom att arbeta med transparens och förklarbarhet i undervisningen kan vi ge eleverna verktygen de behöver för att navigera en värld där AI spelar en allt större roll. **AI är ett kraftfullt verktyg – men bara om vi lär oss att använda det klokt!**

Domänexpertis och kritiskt tänkande – varför kunskap är avgörande för källkritik

Källkritik ses ofta som en generell färdighet, en uppsättning frågor och metoder som kan appliceras på vilken information som helst. Vem har skrivit texten? Vad har de för syfte? När publicerades den? Dessa frågor är viktiga, men de räcker inte. Källkritik kan aldrig vara helt frikopplad från kunskap om ämnet. För att förstå och värdera information krävs något mer än kritiska frågor – det krävs domänexpertis, det vill säga en djupare förståelse för ämnesområdet.

Vi kan ta ett exempel: Föreställ dig två personer som läser en artikel om klimatförändringar. Den ena har goda kunskaper i klimatvetenskap och känner till hur klimatmodeller fungerar, vilka forskare som är ledande inom fältet och hur olika faktorer påverkar temperaturförändringar. Den andra personen saknar denna bakgrund och har en mer allmän förståelse av klimatfrågan. När de läser artikeln kommer de att kunna använda källkritik på olika sätt.

Den första personen kan granska påståendena i artikeln direkt: Stämmer det som sägs om växthusgaser? Är forskningsresultaten representativa för forskningsfältet? Den andra personen, som saknar samma ämneskunskaper, kan istället behöva förlita sig på en bedömning av källans trovärdighet: Är det en erkänd forskare som uttalar sig? Är artikeln publicerad i en ansedd vetenskaplig tidskrift?

Detta illustrerar en viktig poäng: Källkritik fungerar inte likadant för alla. Utan ämneskunskap blir vi i högre grad beroende av att lita på sekundära faktorer, som vem som skrivit texten och var den är publicerad. Med ämneskunskap kan vi däremot direkt bedöma om ett påstående verkar rimligt.

Desinformation och "korn av sanning"

Ett av de största problemen med desinformation är att den sällan är helt falsk. Istället innehåller den ofta en blandning av sanning och vilseledande information. Det gör att den kan verka trovärdig vid en första anblick – särskilt för den som saknar djupare ämneskunskaper.

Ett exempel är medicinsk desinformation. Under pandemin spreds många påståenden om att vissa vitaminer eller behandlingar kunde skydda mot covid-19. Dessa påståenden byggde ofta på riktiga vetenskapliga studier – men misstolkades eller presenterades på ett sätt som gav en felaktig bild av deras faktiska betydelse. För den som inte är insatt i medicinsk forskning kan det vara svårt att avgöra vad som är relevant och vad som är missvisande.

På samma sätt används historiska fakta ofta för att sprida politisk desinformation. En konspirationsteori kan exempelvis innehålla verkliga historiska händelser men sätta in dem i ett narrativ som snedvrider deras betydelse. Personer utan historisk kontext kan ha svårt att avgöra vad som är en rimlig tolkning och vad som är en medveten manipulation.

Domänkunskap gör oss svårare att vilseleda

Studier visar att personer med djupa ämneskunskaper är bättre på att identifiera felaktigheter och att genomsåda manipulerad information. De kan snabbare se vad som saknas, vad som inte stämmer och när något verkar osannolikt.

Ett tydligt exempel på detta är vetenskapliga bedrägerier. Inom forskarsamhället upptäcks ofta förfalskade eller felaktiga studier av andra forskare inom samma fält – inte nödvändigtvis för att de har tillgång till mer information, utan för att de har en djupare förståelse för vad som är rimligt och för hur vetenskapliga processer fungerar.

På samma sätt kan en erfaren journalist snabbare se när en nyhetsartikel verkar tveksam, eftersom hen har kunskap om hur journalistik fungerar och vilka källor som brukar vara pålitliga. En historiker kan identifiera när någon försöker vinkla ett historiskt skeende på ett sätt som inte stämmer med etablerad forskning.

Detta betyder att vi, för att stärka elevernas källkritiska förmåga, inte bara kan undervisa dem i generella källkritiska metoder. Vi måste också bygga deras ämneskunskaper – ge dem en grund av fakta, förståelse och kontext så att de själva kan avgöra vad som är trovärdigt och vad som inte är det.

Kritiskt tänkande är en aktiv process

Att vara kritisk innebär inte att avfärda all information vi möter. Tvärtom handlar kritiskt tänkande om att kunna väga information, analysera den i sitt sammanhang och jämföra med andra källor. Detta är en aktiv process, där vi inte bara konsumerar information utan också ifrågasätter, undersöker och diskuterar den.

En viktig del av kritiskt tänkande är **lateral läsning** – en metod som innebär att vi snabbt lämnar en källa för att kontrollera den mot andra oberoende källor. Forskning visar att faktagranskare och journalister ofta använder denna teknik när de möter ny information: istället för att analysera en artikel i detalj går de direkt vidare och söker upp andra källor för att se om informationen bekräftas.

I undervisningen kan vi träna elever på detta genom att låta dem jämföra AI-genererade svar med olika källor, diskutera hur information kan vinklas och fundera över varför vissa typer av information sprids mer än andra.

Men för att göra detta effektivt måste vi också ge dem verktygen att förstå ämnet de granskar. Vi kan inte förvänta oss att elever ska kunna analysera en medicinsk studie om de saknar grundläggande förståelse för medicinsk forskning. Vi kan inte begära att de ska kunna identifiera historiska förvrängningar om de inte har studerat historia.

Kritiskt tänkande och domänexpertis går hand i hand. För att elever ska kunna genomsåda manipulation och desinformation behöver de både generella källkritiska verktyg och en bred faktakunskap att utgå ifrån. I en tid av AI-genererad information är detta viktigare än någonsin.

Agens i en värld där AI gör "intelligens" tillgänglig som råvara

I ett informationslandskap där generativ AI kan producera text, bilder och videor på begäran, aktualiseras frågan om vår **agens** – det vill säga vår förmåga att själva styra och förhålla oss kritiskt till information, snarare än att bara ta emot den. Agens kan beskrivas som den kognitiva, medvetna kontrollen över hur vi tänker, värderar och interagerar med information. När AI skapar innehåll som låter övertygande, och algoritmer styr vad vi ser i våra flöden, uppstår en risk: Är det tekniken som har agens över oss, eller är det vi som aktivt kan välja vad vi tar till oss och hur vi granskar det?

För att **utöva kritiskt tänkande** behöver vi just denna kognitiva agens. Den som har hög agens stannar upp och frågar sig: "Vem säger detta, och varför?" eller "Finns det andra perspektiv, och hur kan jag verifiera påståendet?" Individer utan denna reflex tenderar att förhålla sig passivt: "Jag såg det på TikTok, alltså är det sant." Kognitiv agens innebär i stället att aktivt reflektera över information, förstå att den alltid är färgad av avsändarens intentioner och kontext, och att man själv har en roll i att bedöma dess värde. Utan agens löper vi större risk att bli mottagare av ensidiga eller felaktiga budskap, i stället för att vara aktiva granskare som söker bekräftelse och alternativa källor.

Agens och källkritik hänger tätt samman. Digitala plattformar påverkar vilken information som lyfts fram, men det är individens agens som avgör hur man förhåller sig till den. Exempelvis kan en låg-agens-inställning vara: "TikTok visar mig detta, så det måste vara populärt och värt att lita på." En högre grad av agens innebär: "TikTok ger mig denna information, men av vilka skäl? Finns det ett algoritmiskt motiv att visa just detta för mig? Hur vet jag att det är sant?" Källkritik stannar därmed inte vid att granska en enskild webbplats eller en tidningsartikel – det handlar också om att förstå de bakomliggande strukturerna och affärsmodellerna som styr informationsflödet. När vi förstår hur rekommendationssystem fungerar, och varför vissa budskap sprids snabbt, blir det lättare att värdera informationen. Denna medvetenhet stärker i sin tur vår agens: vi vet hur och varför information sprids, och kan agera på det.

Samtidigt står vi inför **flera utmaningar** som kan försvaga vår agens i det digitala informationslandskapet. Ett exempel är de algoritmiska filterbubblor som uppstår när plattformar optimerar innehåll för våra förmodade intressen. Även om detta kan upplevas bekvämt – vi får snabbt upp videor eller artiklar som matchar vår smak – riskerar vi att tappa i bredd och djup. Våra val blir i praktiken styrda av algoritmer, vilket minskar vår upplevda och faktiska handlingsfrihet. Informationsöverflöd är en annan utmaning. När vi översköls av nyheter, uppdateringar och klipp från otaliga källor kan det kännas enklare att bara följa strömmen i stället för att aktivt granska och söka egna källor. Här eroderar vår agens om vi inte medvetet tränar vår förmåga att sälla, jämföra och ifrågasätta.

I dag blir detta extra påtagligt när innehållet **skapats av AI**. Om vi inte ens vet om en text eller bild är skapad av en människa eller en AI, förlorar vi en viktig bit i källkritiken: "Vem är avsändaren?" Även om AI inte nödvändigtvis har en egen "agenda", kan dess träningsdata innehålla bias eller felaktigheter, vilket då reproduceras i det genererade materialet. Utan hög agens riskerar vi att lita på ett innehåll bara för att det är välformulerat eller visuellt imponerande. Med stark agens kan vi däremot fråga oss: "Varför ser den här bilden ut som

den gör?”, ”Vilken data kan AI:n ha använt för att skapa denna text?” och ”Hur kan jag dubbelkolla dessa påståenden?”

Hur stärker vi då agens? En viktig komponent är utbildning och reflektion kring hur algoritmer och informationssystem fungerar. Det handlar inte bara om traditionell källkritik, utan även om att förstå hur reklamsamarbeten, affärsmodeller och data insamlade från vår användning driver fram det som visas i våra flöden. Dessutom är det viktigt att utveckla en balanserad skepsis, så att vi inte faller i fällan av total relativism eller konspirationstänkande, där allt blir lika misstänkt och därmed lika trovärdigt.

Här kan **skolan** spela en avgörande roll genom att gå bortom statiska checklistor som CRAAP och SIFT och i stället ge elever praktiska övningar. Att låta elever spåra källan till en viral nyhet eller själva producera fejkad innehåll med en AI ger dem erfarenhetsbaserade insikter om hur lätt information kan manipuleras. De lär sig också vikten av att kontrollera uppgifter mot flera oberoende källor (lateral läsning) och hur egna sökstrategier kan påverka resultatet. Detta utvecklar en fördjupad förståelse för informationsflödets dynamik och stärker elevernas agens – de blir inte bara passiva mottagare utan aktiva informationsjägare.

AI kan slutligen både **stärka och försvaga** denna agens. Om AI används medvetet – som ett verktyg för att snabbt hitta varierade perspektiv eller för att öva källkritik – kan det göra människor mer rustade. Det kan korta ner informationssökningen, bidra med sammanfattningar av komplexa texter och underlätta för en bredare grupp att få tillgång till information. Men om vi litar blint på AI som en objektiv expert förlorar vi vår autonoma förmåga. När AI:n blir vårt enda filter kan vår agens försvagas, eftersom vi sällan ifrågasätter svaren från ”det digitala oraklet”.

I grund och botten är **agens en fråga om makt** – makten att förstå, välja och agera i mötet med information. Även i en värld där ”intelligens” blivit alltmer tillgänglig som råvara är det mänsklig kognitiv förmåga och kritiska reflektioner som avgör vad vi gör med all denna nya kapacitet. Vi kan använda generativ AI för att berika vårt lärande och vår förståelse, men bara om vi är medvetna om våra egna beslut och håller fast vid vår rätt att välja och ifrågasätta. Så länge vi bevarar och stärker vår agens förblir tekniken ett verktyg i våra händer, snarare än att vi blir föremål för teknikens agendor.

Lektionsförslag – Källkritik i praktiken

- **Algoritmer och rekommendationssystem**
- **Informationsnätverk och hur AI påverkar sociala medier**
- **Hyperindividualisering och filterbubblor**
- **Ekokammare**
- **Hyperpersonalisering och integritet (privacy)**
- **Hyperverklighet, desinformation och deepfakes**
- **Desinformation**
- **Hallucinationer hos AI**
- **Manipulationstekniker (cherry picking, känsloladdat språk, falska dikotomier)**
- **Prebunking**
- **Debunking**
- **Källtillit och sakkritik**
- **Bias inom AI**
- **Mänskligt bias i förhållande till AI**
 - Bekräftelsebias
 - Auktoritetsbias
 - Förankringsbias
 - Tillgänglighetsbias
 - Halo-effekten
- **Transparens och förklarbarhet (XAI)**
- **Domänexpertis och kritiskt tänkande**

Lektion 1: Varför källkritik i AI-eran? (Introduktion)

Introduktion

Syftet med denna första lektion är att sätta scenen: varför källkritik och kritiskt tänkande är avgörande i en värld där generativ AI kan skapa **övertygande men potentiellt felaktig information**. Eleverna introduceras till hur dagens informationslandskap förändrats av AI. Till exempel kan en chattbot som ChatGPT leverera självsäkra svar som låter trovärdiga men ändå innehåller faktafel. Introduktionen betonar vikten av att **alltid förhålla sig källkritiskt**, oavsett om informationen kommer från en traditionell källa eller en AI. Genom att förstå riskerna – som falska nyheter, AI-hallucinationer och desinformation – motiveras eleverna att lära sig de verktyg som kommande lektioner erbjuder.

Genomförande (steg-för-steg)

1. **Inled med en diskussion:** Fråga eleverna varifrån de får sin information (nyheter, sociala medier, vänner, AI-chatbots osv). Lista dessa på tavlan.
2. **Visa exempel:** Presentera två korta texter om samma nyhet – en från en etablerad nyhetssajt och en genererad av en AI. (Du kan själv be ChatGPT skriva en nyhet om ett aktuellt ämne.) Låt eleverna läsa och reflektera över texternas stil och innehåll.
3. **Jämför källor:** Diskutera i klassen hur man kan avgöra vilken text som är mer trovärdig. Peka på skillnader som avsändare, språk och möjlighet att bekräfta uppgifter.
4. **Kort föreläsning:** Gå igenom en översikt av dagens digitala utmaningar: enorm informationsmängd, hastig spridning via sociala medier och AI-verktyg som kan generera innehåll. Förklara att **källkritik** handlar om att systematiskt granska information och dess källa, och att **källtillit** handlar om att lära sig lita på *pålitliga* källor man identifierat som trovärdiga.
5. **Elevernas förkunskaper:** Poängtera att även om ungdomar är digitalt infödda betyder det inte automatiskt att de är bra på källkritik. I själva verket visar forskning att de som tror sig vara bäst ofta presterar sämst i tester av digital källkritik. Detta kan vara ögonöppnande och understryker lektionens syfte.

Lektionsingångar & diskussion

- **Diskussionsfråga 1:** "*Hur vet du om något du läser på nätet är sant?*" – Låt eleverna först diskutera två och två, sedan i helklass.

- **Diskussionsfråga 2:** "*Kan vi lita på AI-genererade texter eller bilder?*" – Samla tankar på tavlan. Troligen får ni fram att AI kan både hjälpa och stjälpa. Lyft att AI inte har **egen förståelse eller ansvar** – det är upp till oss att tolka och värdera dess svar.
- **Lektionsingång med exempel:** Visa en viral bild (t.ex. en bild genererad av AI som många trott var äkta). Fråga eleverna om de sett bilden och vad de tror om dess äkthet. Använd detta som ingång till att prata om *varför* källkritik behövs i dag.
- **Reflektionsfråga:** "*Hur pålitliga anser ni att era vanliga informationskällor är, och varför?*" – Eleverna kan skriva korta reflektionslappar anonymt som samlas in, för att stämna av deras utgångsläge.

Didaktiska tips

- **Skapa öppenhet:** Var lyhörd för elevernas egna upplevelser av falsk information. Många har säkert stött på rykten eller virala klipp. Be dem dela exempel – det gör ämnet konkret och relevant.
- **Undvik skam:** Om någon elev spridit en felaktig nyhet tidigare, se till att ingen hängs ut. Understryk att vi alla kan bli lurade online, och att syftet är att lära sig, inte att skämmas.
- **Använd tydliga begrepp:** Definiera centrala ord tydligt (t.ex. källkritik, desinformation, generativ AI). Skriv upp dem med enkla förklaringar på blädderblock eller projektion så att ni kan återkomma till dem under kursen.
- **Interaktivitet:** Om möjligt, låt eleverna använda sina mobiler eller datorer i steg 1 för att kolla dagens nyhetsflöde eller vad som trendar på sociala medier. Detta anknyter lektionen till deras vardag direkt.

Forskning & verktyg

- Forskning visar att **överdriven självsäkerhet** i digital källkritik är farligt – de som tror sig vara bäst kan vara sämst. Detta understryker vikten av att öva källkritik praktiskt, inte bara tro att man "kan det".
<https://www.vilarare.se/nyheter/annat/sjalvsakra-unga-samre-pa-kallkritik/>
- **Åsa Wikforss** betonar behovet av att inte bara lära kritik utan även *källtillit*, dvs. hur man hittar trovärdiga källor. Skolan och hemmet spelar en roll i att barn lär sig skilja på seriösa medier och opålitliga källor.
<https://internetkunskap.se/artiklar/kallkritik/darfor-behover-du-kalltillit/>

- Digitalt verktyg: **Internetstiftelsens Internetkunskap** har enkla artiklar om vad AI är och hur det påverkar samhället. Det kan vara bra bakgrundsläsning för lärare och elever som vill förstå grunderna. **Det finns också färdiga** lektionsupplägg om AI på plattformen Digitala lektioner, vilket kan komplettera denna bok och ge interaktiva övningar.
- <https://digitalalektioner.se/amnesomrade/digital-kallkritik/>

Arbetsblad – Lektion 1: Källkritik i en värld av generativ AI

Syfte med arbetsbladet

Att eleverna ska träna på att identifiera källor, reflektera över hur man kan kontrollera trovärdighet och bekanta sig med grundläggande begrepp inom källkritik och AI.

Övning 1: Var hittade jag informationen?

I tabellen nedan fyller du i **exempel** på information du eller andra stött på nyligen (det kan vara en nyhetsartikel, en tweet, ett rykte på sociala medier etc.). För varje källa ska du besvara:

1. **Källa** (Vad heter den? Är det en person, en nyhetssajt, en AI-generator?)
2. **Var hittade jag den?** (Twitter, TikTok, nyhetssajt, skickad från en kompis osv.)
3. **Hur kan jag kolla om den är trovärdig?** (Vilka steg tar du? Vilka verktyg eller webbplatser används?)

Källa	Var hittade jag den?	Hur kan jag kolla om den är trovärdig?
Exempel: Tweet om en nyhet	Twitter	Söka efter nyheten på betrodda nyhetssajter (SVT, DN m.fl.)

Tips: Tänk på att information ofta kan vara **AI-genererad**. Hur skulle du ta reda på om tweeten, videon eller inlägget kommer från en **pålitlig källa** eller om det kan röra sig om **desinformation**?

Övning 2: Kortfall – Skärmdumpen från Instagram

Du får en **skärmdump** från en vän. Bilden påstår att en **kändis har dött**. Vad gör du?

Kryssa i **ett eller flera** alternativ nedan. Diskutera sedan i mindre grupper varför vissa alternativ är bättre än andra.

1. **A:** Delar vidare direkt i din story eller till andra vänner.
2. **B:** Söker på Google eller betrodda nyhetssajter för att se om nyheten bekräftas.
3. **C:** Besöker kändisens **officiella konto** (Instagram, Twitter) för att se om det nämns något.
4. **D:** Tänker att “kändisar fejkar sin död ibland” och avfärdar det som fejk utan att kolla.
5. **E:** Misstänker att det kan vara **fejk** och letar fler källor (t.ex. rykten, officiella uttalanden).

Diskussionsfrågor:

- Varför kan det vara riskabelt att dela vidare en påstående utan att kolla?
- Har du varit med om en liknande situation där en “chocknyhet” visade sig vara falsk?
- Hur kan AI-teknik användas för att sprida både **sanna** och **falska** rykten ännu snabbare?

Begreppslista

I rutan nedan finns några viktiga begrepp som hör till lektionen. Försök ge **egna förklaringar** utifrån genomgången i klassen och dina tidigare erfarenheter.

1. **Källkritik**
 - *Din förklaring:*
 - ...
2. **Källtillit**
 - *Din förklaring:*
 - ...
3. **Generativ AI**
 - *Din förklaring:*
 - ...
4. **Desinformation**
 - *Din förklaring:*
 - ...

Tips: När ni gått igenom begreppen, ställ gärna frågor i par eller helklass:

- Har ni sett **generativ AI** (t.ex. bilder, texter) som cirkulerar som om de vore äkta?
- Hur bygger man **källtillit** till en viss nyhetssajt eller informatör?

Avslutande reflektion

- Hur kan ni använda övningarna ovan i er vardag när ni stöter på rykten, påståenden eller spektakulära nyheter online?
- På vilket sätt kan **generativ AI** försvåra eller underlätta ert arbete med källkritik?

Kom ihåg: Huvudbudskapet i lektionen är att vi alltid ska vara vaksamma, oavsett om information kommer från AI eller en människa. Genom att **systematiskt granska källor** och aktivt söka **fler bekräftelser** höjer vi vår källtillit och minskar risken för att sprida felaktigheter.

Lektion 2: Grundläggande källkritik – vem, vad, varför?

Introduktion

Denna lektion dyker djupare i de **grundläggande källkritiska frågorna**: Vem är avsändaren? Vad är det för slags information? Varför har den skapats? och Hur når den oss? Syftet är att ge eleverna en strukturerad metod att granska information, oavsett om den kommer från en webbplats, ett socialt medieinlägg eller en AI-genererad text. I AI-eran blir dessa frågor ännu viktigare – **avsändaren kan vara en algoritm eller en anonym aktör**, och motivet bakom information kan vara att påverka snarare än att informera. Genom att öva de klassiska källkritiska kriterierna (äkthet, tid, beroende, tendens) i moderna exempel lär sig eleverna att applicera gammal beprövad metodik på nya medier.

Genomförande (steg-för-steg)

1. **Genomgång av källkritiska frågor:** Skriv upp de centrala frågorna på tavlan: *Vem? Vad? Var? När? Varför? Hur?* Gå igenom var och en med kort förklaring. Exempel: "Vem säger detta?" (avsändare), "Var publiceras det?" (kanal/medium), "Varför kan det ha publicerats?" (syfte).
2. **Gruppövning – källanalys:** Dela in klassen i smågrupper. Ge varje grupp en kort text eller inlägg att analysera (t.ex. en nyhetsartikel, en bloggtext, ett AI-genererat forum-svar). Grupperna fyller i ett **källkritik-formulär** med frågorna ovan om sin text.
3. **Diskussion av resultat:** Låt varje grupp kort redovisa vad de kom fram till. Ställ följdfrågor: *Hur pålitlig verkar källan? Vad gjorde er misstänksamma eller trygga?*
4. **Fokus på AI:** Lyft i diskussionen att om texten vore AI-genererad, *vem* är egentligen avsändaren? Kom fram till att en AI-text saknar en mänsklig avsändare med ansvar, och att det i praktiken är de som tränat/äger AI:n som står bakom – men det är ofta otydligt. Här är det viktigt att elever inser att en text utan tydlig avsändare bör behandlas med större skepsis.
5. **Reflektion individuellt:** Be eleverna skriva ned en kort plan för hur de ska kolla en *valfri* information de stött på senaste veckan (t.ex. "Jag såg på TikTok att X, jag skulle kolla det genom att..."). Detta ger dem personlig anknytning och övning.

Lektionsingångar & diskussion

- **Ingång med exempel:** Ta en falsk nyhet som faktiskt cirkulerat (gärna en humoristisk eller ofarlig för stämningens skull). Läs upp rubriken högt. Fråga: "*Vad behöver vi ta*

reda på för att veta om detta är sant?" Låt eleverna kasta ur sig förslag – sannolikt täcker de många av de grundläggande frågorna.

- **Diskussionsfråga:** "*Hur påverkar det ditt förtroende om du inte vet vem som står bakom en text?"* Koppla till AI-genererat innehåll – att transparens om avsändare saknas.
- **Case-diskussion:** Beskriv ett scenario: "Du hittar en artikel som hyllar en mirakelkur mot stress, skriven av 'Dr. Andersson'. Hur kan du ta reda på om Dr. Andersson finns och är en expert?" Här övar man *vem* och *vad*-frågorna – sök på namnet, kolla upp kvalifikationer, osv. Låt eleverna resonera.
- **Tendens/trovärdighet:** Diskutera begreppet **tendens**: "*Kan en text vara sakligt korrekt men ändå vinklad?"* (Ja, genom urval av fakta, språk osv). Elever kan ge exempel, kanske politiska texter eller reklam.

Didaktiska tips

- **Använd aktuella exempel:** Uppdatera gärna exemplen varje termin för att hålla dem relevanta. Elever engageras mer av något som nyligen hänt eller sprids.
- **Varva seriöst och lättsamt:** Källkritik kan kännas tungt; att ta någon rolig internetmyt som exempel först kan lätta upp stämningen innan man går in på allvarligare saker.
- **Stödstrukturer:** Erbjud stödmallar (t.ex. källanalys-formuläret) för elever som har svårt att komma igång. Det kan ha förfyllda exempel eller hint-frågor under varje rubrik.
- **Differentiering:** Äldre högstadieelever kan behöva mer stöttning eller kortare texter att analysera, medan gymnasieelever kan hantera mer komplexa källor. Var beredd att anpassa materialet efter gruppens nivå.
- **Samarbete:** Om skolan har ett bibliotek eller en skolbibliotekarie med specialkunskap, involvera dem. Skolbibliotekarier är ofta experter på källkritik och kan tillföra perspektiv eller resurser.

Forskning & verktyg

- Klassisk källkritik med *frågemetoden* är grunden. Enligt forskning av bl.a. **Sam Wineburg** (Stanford) är det effektivt att jämföra källor och kolla upp avsändare – något professionella faktagranskare (som journalister) gör rutinmässigt.

- **Källkritiken ska sitta i händerna:** Thomas Nygren, professor i didaktik, påpekar att elever lär sig källkritik bäst genom praktisk övning, inte enbart teori. Att göra analyser själva (som i gruppövningen ovan) är alltså nyckeln.
- **Verktyg:** Introducera **SIFT-metoden** (Stoppa, Undersök källan, Finna bättre täckning, Spåra ursprunget) som utvecklats för digital källkritik. Den är lätt att lära ut och finns översatt till svenska i olika guider online.
- Digitala verktyg:
 - **Google omvänt bildsök** – Visar var en bild förekommit tidigare; bra för att kolla om en bild är äkta eller tagen ur sitt sammanhang.
 - **Wikipedia** – Inte som källa i sig, men som verktyg för att hitta bakgrund om avsändare eller ämne (källorna i Wikipedia-artiklar kan ge ledtrådar). Påminn om att Wikipedia i sig ska hanteras källkritiskt, men ofta länkar den till primärkällor.

Arbetsblad – Lektion 2: Grundläggande källkritik

Vem, vad, var, när, varför och hur?

1. Källkritiskt analysformulär

Instruktion: Fyll i nedanstående punkter för en **given källa**. Det kan vara en artikel, ett blogginlägg, en AI-genererad text, en nyhet på sociala medier, etc. Använd gärna en riktig text eller länk som ni ska analysera i denna lektion.

1. **Vem?**
 - Vem står bakom informationen?
 - Vad vet du om avsändaren (person, myndighet, företag, anonym)?
2. **Vad?**
 - Vilken **typ** av information är det? (nyhet, debattinlägg, reklam, AI-genererad text, annat)
 - Vilka **påståenden** görs i texten?
3. **Var?**
 - På **vilken plattform** eller i vilket sammanhang publiceras det? (tidning, blogg, socialt medium)
 - Påverkar **kontexten** trovärdigheten?
4. **När?**
 - När **skapades** eller **publicerades** informationen?
 - Är den **aktuell** eller kan den vara **inaktuell**? Spelar tidpunkt roll?
5. **Varför?**
 - **Syfte:** att informera, påverka, sälja något, underhålla?
 - Finns det en **tydlig agenda** eller bakomliggande intention?
6. **Hur?**
 - Hur fick du **tag** på informationen? (Delad av en vän, algoritmflöde, egen sökning?)
 - Hur **sprids** den? Finns den på många ställen, eller bara en källa?
7. **Bedömning:**
 - Efter att ha svarat på ovan – **litar du** på informationen? Varför/varför inte?
 - Vad skulle du **göra** för att verifiera (eller förkasta) den här informationen ytterligare?

2. Övningsexempel: Fejkad nyhet

Scenario: Du har hittat en **kort text** (se nedan) som påstår att en ny mirakelkur botar stress på bara 5 minuter. Den är skriven av “Dr. Andersson” och utlovar snabba resultat. Inga källor eller referenser anges.

“Mirakelkur mot stress: Bli stressfri på 5 minuter!”

Utdrag: “Hej, jag är Dr. Andersson. Jag har upptäckt en revolutionerande metod som gör att du helt slipper stress på fem minuter. Du behöver bara ta detta kosttillskott varje morgon, och din ångest försvinner mirakulöst. Inga biverkningar har rapporterats, och 100 % av mina patienter älskar denna metod. Passa på innan vi höjer priset!”

Uppgift: Analysera den fejkade nyheten genom att fylla i formuläret ovan (Vem, Vad, Var, När, Varför, Hur + Bedömning). Gå sedan tillsammans i klassen och jämför era slutsatser.

Diskussionsfrågor:

- Varför kan denna text verka misstänksam?
- Hur kan ni kolla om “Dr. Andersson” existerar och är legitimerad?
- Vilka tecken på **tendens** (vinkling) ser ni?

Tips för genomförandet

- **Arbeta i par eller smågrupper:** Diskutera varje ruta i formuläret innan ni fyller i den.
- **Ge gärna exempel:** Om ni skriver “Avsändaren är anonym”, fundera på hur ni kan ta reda på mer om vem som ligger bakom.
- **Variera exempel:** Låt olika grupper analysera olika texter (en kan vara AI-genererad, en annan från en känd nyhetssajt, en tredje från en influencer).

Avslutande reflektion

1. **Hur ändras** din bedömning av källans trovärdighet när du vet att den kan vara AI-genererad och inte har en tydlig avsändare?
2. **Vad** i texten är mest misstänkt eller mest övertygande?
3. **Hur** kan du använda samma frågemodell (Vem, Vad, Var, När, Varför, Hur) i vardagen nästa gång du ser en “mirakelnyhet” eller ett klickbait-inlägg?

Koppla tillbaka till lektionen: Genom dessa frågor lär ni er se inte bara *vad* som sägs, utan också *vem* som säger det, *varför*, och *hur* det sprids. Detta är grunden för en modern, **kritisk mediekonsument** – särskilt viktigt i AI-eran.

Lycka till med er källkritiska analys!

Lektion 3: Verklig eller AI

(Vad är AI-genererat och vad är verkligt?)

Introduktion

Denna lektion syftar till att väcka elevernas medvetenhet om hur enkelt det är att skapa eller förvanska information med hjälp av generativ AI. Eleverna får lära sig att känna igen AI-genererat material (t.ex. bilder, ljud, musik) och upptäcka skillnaden mellan sådant som är på riktigt och sådant som är skapat av en dator. Ni kan använda en färdig **PowerPoint** (bifogas i materialet) för att leda eleverna genom lektionen, eller plocka de delar ni vill betona mest.

Syfte och mål

- **Öka medvetenhet** om att många bilder, videor och ljudklipp i dag kan vara AI-genererade.
- **Introducera begreppen** generativ AI, deepfake, och desinformation.
- **Motivera eleverna** till att vara mer källkritiska och ifrågasätta bilder och klipp de ser i sin vardag.

Förslag på lektionsgångar

1. **Kahoot om AI-bilder**
 - Använd Kahooten "[Spot the AI-generated image](#)" för att låta eleverna gissa om en bild är riktig eller AI-genererad.
 - Diskutera resultaten: Var det enkelt eller svårt att se skillnad? Hur många hade rätt/fel?
2. **Deepfake Game**
 - Spela "[Leon Furze Deepfake Game](#)" i helklass. Eleverna får se korta klipp där de ska avgöra om det är äkta eller manipulerat.
 - Reflektera kring hur trovärdiga deepfakes kan vara, och vilka risker det kan innebära (t.ex. fejkade uttalanden av kändisar eller politiker).
3. **Exempelmaterial från bifogad mapp**
 - Visa några exempel (bilder eller korta videor) ur mappen ni fått bifogad, eller använd egna klipp.
 - Låt eleverna rösta med "tumme upp" eller "tumme ner" för att markera om de tror att klippet är äkta eller fejk.

Genomförande (steg-för-steg)

1. **Introduktion (5 min)**
 - Presentera lektionens huvudfråga: "*Hur vet vi vad som är verkligt och vad som är AI-genererat?*"
 - Koppla till elevernas vardag: Har de stött på videor eller bilder online som verkat misstänkt fejk?
2. **Korta övningar eller spel (10–15 min)**
 - **Kahoot eller Leon Furze-spelet:** Låt eleverna gissa om bilder/klipp är äkta eller AI-genererade.

- Diskutera spontana reaktioner: Var det något som gjorde att de misstänkte fejk? (Konstig bakgrund, märkliga detaljer, skuggor som inte stämmer, osv.)
3. **Förklaring av generativ AI (5 min)**
- Visa och förklara i enkla ord vad generativ AI är: en algoritm som tränats på stora datamängder för att skapa nya, “unika” bilder, texter eller ljud.
 - Nämn några vanliga tjänster eller visa, om din kommun tillåter det, tjänsterna på projektorn:
 - **ideogram.ai** för att skapa AI-bilder.
 - **grok.com** för att generera bilder (mindre restriktiv, även kändisbilder).
 - **sunno.com** för att skapa musik.
 - **elevenlabs.io** för att skapa röstklipp.
4. **Diskussion i par eller smågrupper (5–10 min)**
- Låt eleverna fundera kring:
 - När skulle AI-genererat material kunna vara hjälpsamt eller roligt?
 - När kan det vara farligt eller vilseledande?
 - Be dem skriva upp **två exempel** på positiva tillämpningar (t.ex. konst, design, musik) och **två exempel** på risker (t.ex. deepfakes för att sprida rykten).
5. **Sammanfattning i helklass (5–10 min)**
- Ta upp exempel från grupperna och diskutera kring balansgången mellan nytta och skada.
 - Understryk att första steget i källkritik är att *inse att något kan vara fejk* – att våga fråga “Kan detta verkligen stämma?” innan man sprider vidare.

Diskussionsfrågor

- **Hur vanligt tror ni att AI-genererat material är i dag?**
- **Vad är det som gör AI-bilder så övertygande?** (Realistiska detaljer, fotorealistiska ansikten, etc.)
- **Kan man förbjuda AI-genererade bilder?** Hur skulle det i så fall gå till?
- **Hur påverkas vårt förtroende för online-innehåll när vi vet att sådant här är möjligt?**

Didaktiska tips

- **Utnyttja elevernas nyfikenhet:** En del elever är intresserade av nya AI-appar och kan till exempel visa varandra hur de testat att skapa musik eller bilder.
- **Ha exempel från elevernas egen värld:** Om någon i klassen redan provat att skapa en “deepfejkad” bild eller video, låt dem berätta.
- **Diskutera etik och integritet:** Passa på att fråga hur det känns att kunna återskapa någons röst eller ansikte utan deras samtycke. Var går gränsen?
- **Teknisk nivå:** Anpassa förklaringen av “hur AI fungerar” efter elevgruppens ålder och intresse. Fokus här är inte att bli AI-expert, utan att förstå möjligheterna och riskerna.

Forskning & verktyg

- **Internetstiftelsen** har översiktliga förklaringar om AI och källkritik (Digitala lektioner).
- **Skolverket** betonar att elever måste lära sig förstå och kritiskt förhålla sig till automatiskt genererat innehåll.

- **Deepfake-teknik** är sedan flera år under snabb utveckling, vilket ökar behovet av att kunna urskilja äkta från falskt.

Sammanfattning

I denna första lektion får eleverna en **startpunkt** för att förstå hur vanligt och enkelt det är att skapa AI-genererat innehåll som kan förväxlas med verkligheten. Genom Kahoot, spel och exempelmaterial övas elevernas öga för detaljer och förmåga att tvivla, ställa frågor och undersöka källan. Fokus ligger på att **väck nyfikenhet** och **motivera** dem till att vara mer källkritiska framöver.

Inför kommande lektioner: Eleverna börjar redan nu ana hur snabbt information kan manipuleras. I nästa steg kommer ni att fördjupa er i hur algoritmer styr flöden och hur man systematiskt granskar källor och innehåll.

Tips: Om tekniken tillåter, låt gärna eleverna (eller läraren i helklass) prova någon enkel AI-bildgenerator, t.ex. [ideogram.ai](https://www.ideogram.ai), för att själva se hur lätt det är att skapa “fejkmaterial”. Diskussionen blir då mer konkret och engagerande.

Lektion 3: Arbetsblad: Verklig eller AI?

Instruktion:

Du kommer att titta på olika bilder eller korta videoklipp och försöka avgöra om de är äkta eller AI-genererade. Fyll i tabellen nedan före och efter klassdiskussionen. Därefter svarar du på några reflektionsfrågor och bekantar dig med viktiga begrepp.

1. Övning: Rätt eller Fejk?

Bild/Klipp	Din bedömning (innan diskussion)	Din bedömning (efter diskussion)	Vilka tecken såg du? (t.ex. ologiska skuggor, konstig bakgrund)
Exempel 1 (länk eller fil)			
Exempel 2 (länk eller fil)			
Exempel 3 (länk eller fil)			

Gör så här:

1. Titta på bild eller klipp nr 1 och skriv i kolumnen vad du tror: äkta eller AI-genererat?
2. Diskutera sedan i par eller helklass. Efter diskussionen fyller du i om du ändrat din bedömning.
3. Lägg till korta anteckningar om vad som fick dig att misstänka att den var äkta eller fejk.

2. Reflektionsfrågor

1. **Hur skulle du förklara för en vän att bilder på nätet inte alltid är vad de verkar?**
Skriv din förklaring här (ca 2–3 meningar):
...
 2. **Vilka konsekvenser kan det få om folk inte förstår att AI kan fejka i princip vad som helst?**
Fundera på exempel inom nyheter, sociala medier, kändisar eller politik:
...
-

3. Begreppslista

Instruktion: Nedan finns viktiga ord från lektionen. Skriv korta förklaringar med egna ord. Du kan använda lektionsgenomgången, diskussioner eller söka efter en enkel förklaring på nätet.

1. **Generativ AI**
Din förklaring:
...
2. **Deepfake**
Din förklaring:
...
3. **Desinformation**
Din förklaring:
...
4. **Algoritm**
Din förklaring:
...

Tips till dig som elev:

- Var nyfiken och våga ifrågasätta sådant som ser för bra (eller för galet) ut för att vara sant.
- Diskutera gärna med en kompis: Hur kan man dubbelkolla om något är äkta?
- Spara din ifyllda begreppslista och jämför med andra lektionsanteckningar för att se hur begreppen hänger ihop.

Lycka till!

Lektion 4: Jag och min algoritm

(Vad berättar mina sociala medier om mig, och stämmer det verkligen?)

Introduktion

Allt vi gör på internet lämnar digitala spår. De sociala medieplattformar vi använder – Youtube, TikTok, Instagram eller andra – samlar information om våra beteenden för att “kurera” (anpassa) flöden som passar just oss. Denna lektion syftar till att låta eleverna undersöka sina egna sociala medieflöden och fundera över vilka intressen och vanor som algoritmerna tror att de har, jämfört med hur de själva ser på sig som personer.

Syfte och mål

- **Öka självinsikt** kring hur algoritmer baserar förslag på vårt tidigare beteende.
- **Reflektera** över skillnaden mellan verkliga intressen och vad flödet visar.
- **Stimulera** kritiskt tänkande kring hur mycket av “jag” som syns i mitt feed – och om feeden även påverkar mig att konsumera vissa typer av innehåll mer än jag tror.

Förslag på lektionsgångar

1. **Kort klassdiskussion:**
 - Fråga: “Hur många av er har sett förslag i era flöden som *egentligen inte* intresserar er?”
 - Låt eleverna dela exempel (“Jag gillade en video om bakning – nu är hela mitt flöde fullt av recept!”).
2. **Visuell illustration:**
 - Visa ett enkelt cirkeldiagram eller skärmdump av lärarens eget Youtube-flöde (om du vill). Poängtera vilken typ av videor som syns. Stämmer det med dina faktiska intressen?
 - Detta blir en konkret ingång till hur algoritmen “läser av” oss.

Genomförande (steg-för-steg)

1. **Förklaring av ‘personlig algoritm’ (5 min)**
 - Gå igenom begreppet *algoritm* i enkla ordalag: “Ett sätt för programvaran att räkna ut vad du gillar baserat på vad du klickar på, hur länge du tittar, vilka du följer etc.”
 - Förklara att detta kan vara både positivt (mer av det vi gillar) och problematiskt (vi blir ‘inlåsta’ i en viss intressesfär eller får antaganden om vilka vi är).
2. **Egen feed-analys (10–15 min)**
 - Be eleverna öppna valfritt socialt medieflöde (Youtube, TikTok, Instagram, Spotify m.fl.) på sin mobil eller dator – om möjligt.
 - **Alternativ:** Om inte alla kan/ vill öppna sina konton i klassrummet, be dem göra en översikt av “senaste 10 videorna” eller “senaste 10 kontona” de sett under gårdagen.

- Eleverna antecknar kort vilka kategorier/teman som dyker upp (t.ex. humor, gaming, smink, sport, musik, nyheter, politik osv.).
- 3. **Reflektion enskilt eller i par (5–10 min)**
 - Eleverna funderar på:
 - “Stämmer detta med hur jag ser mig själv?”
 - “Finns det något i flödet som förvånar mig?”
 - “Är det något jag inte vill ha men ändå får mycket av?”
 - De kan sedan diskutera två och två: Vilka likheter/skillnader har de upptäckt?
- 4. **Kort presentation eller skriftlig inlämning (5–10 min)**
 - **Muntlig presentation (i mindre grupper eller helklass):** Låt frivilliga berätta om vad som dominerar deras flöden och varför de tror att algoritmen visar just det.
 - **Skriftlig reflektion:** Eleverna kan skriva en halv sida om “Min feed, mina intressen och mina överraskningar”.
- 5. **Sammanfattning och diskussion (5–10 min)**
 - Gå tillbaka till frågan: “Vad säger algoritmen om oss – och stämmer det?”
 - Diskutera hur elevers självbild kan påverkas av att ständigt mötas av vissa slags videos eller inlägg.
 - Ta upp möjligheten att “styra” flöden aktivt (t.ex. att söka på nya saker, avfölja vissa kanaler, ‘gilla’ nya saker för att bredda algoritmens förslag).

Diskussionsfrågor

1. **Tror ni att algoritmen har “rätt” bild av vilka ni är?**
2. **Kan algoritmen förstärka intressen eller åsikter ni redan har? På vilket sätt?**
3. **Vilka positiva effekter kan personaliserade flöden ha för era studier eller fritidsintressen?**
4. **Finns det risk att ni missar annat viktigt eller spännande för att algoritmen fokuserar på sådant ni redan gillar?**
5. **Hur skulle ni kunna förändra ert flöde om ni vill?**

Didaktiska tips

- **Respekt för integritet:** Påminn eleverna om att de inte måste dela personliga eller känsliga detaljer om sina flöden om de inte vill. Man kan prata mer generellt (“Jag får mycket av X, men det känns inte riktigt jag”).
- **Koppla till annan undervisning:** Om ni arbetar med digital kompetens eller sociala frågor, kan ni ta upp hur algoritmer samverkar med reklam och influencers.
- **Framtidsperspektiv:** Diskutera hur algoritmer inte bara visar vad du gillar nu, utan kan forma vad du gillar i framtiden.
- **Differentiering:** För elever som vill fördjupa sig: låt dem undersöka hur man manuellt kan justera rekommendationer (t.ex. inställningar för “visa mindre av detta” på Youtube/TikTok).
- **Livslång relevans:** Påminn om att många vuxna inte heller tänker på hur algoritmer påverkar oss. Att träna kritiskt tänkande kring detta är en värdefull färdighet.

Fördjupning

- **Digital dagbok:** Eleverna kan under en vecka göra en kort “dagboksanteckning” över vilka typer av videor de får upp och hur det ändras när de aktivt söker på nya ämnen. Nästa lektion kan ni jämföra resultat i grupper.
- **Jämförelse mellan plattformar:** Eleverna kan jämföra sitt flöde på två olika appar (t.ex. Instagram vs. Youtube) för att se om algoritmerna har samma bild av dem.

Sammanfattning

I denna lektion får eleverna **undersöka och reflektera** kring hur sociala medieplattformar visar innehåll baserat på tidigare beteende. De analyserar sitt eget flöde, drar slutsatser och diskuterar eventuella skillnader mellan den bild algoritmen förmedlar och vem de själva anser sig vara. Genom att förstå den här mekanismen ökar elevernas **digitala självinsikt** och förmåga att själva påverka vilka intryck de tar del av online.

Arbetsblad – Lektion 4: Jag och min algoritm

Vad berättar mina sociala medier om mig – och stämmer det verkligen?

1. Öppna ditt flöde (eller återskapa gårdagens “Topp 10”)

Instruktion: Ta fram ditt vanliga sociala medieflöde (t.ex. YouTube, TikTok, Instagram, Spotify). Om det inte är möjligt i klassrummet, gör en **“Topp 10-lista”** över de senaste inläggen/videorna du såg igår.

1. **Notera vilka kategorier** du ser mest just nu. Exempel:
 - Humor, Gaming, Mode, Mat, Sport, Musik, Politik, Nyheter, m.m.
2. **Skriv ner minst 5–10 inlägg/klipp:**

Inlägg/Klipp	Kategori	Varför tror du att algoritmen visar detta?
1. ...	t.ex. Humor	...
2. ...	t.ex. Sport	...
3. ...	t.ex. Politik	...
4. ...	t.ex. Musik	...
5. ...		

Reflektion:

- Vilka 2–3 kategorier dominerar?
 - Varför tror du algoritmen “valde” just detta?
-

2. Vem är jag enligt algoritmen?

1. **Hur väl tycker du** att flödet stämmer överens med dina **faktiska** intressen?
 - Skriv kort: “Flödet visar mycket av....., men jag är (inte) så intresserad av det egentligen.”
 2. **Finns något i flödet** som **förvånar** dig?
 - “Hade ingen aning om att jag får så mycket av...?”
 - Varför tror du att det dykt upp?
 3. **Ser du** något i flödet som du **helst hade velat slippa**?
 - Anteckna exempel: “Jag får förslag om... men är ointresserad, ändå fortsätter algoritmen visa det.” Varför händer det?
-

3. Diskutera i par

1. **Jämför dina anteckningar** med en klasskompis.
 - Likheter: Har ni någon kategori som återkommer hos båda?
 - Skillnader: Finns någon kategori du har men inte hen?
2. **Varför kan algoritmen** ibland visa saker som inte alls stämmer med våra intressen?

- Fundera på: Reklam, slumpmässiga rekommendationer, tidigare engagemang osv.
3. **Påverkas** du av flödet mer än du tror?
- Exempel: Har du börjat gilla/skratta åt /bli orolig över saker bara för att du ser dem ofta?

4. Skriv eller berätta kort

Välj **ett** av följande alternativ:

1. **Skriftlig reflektion** (ca en halv sida):

“Min feed, mina intressen och mina överraskningar.”

- Berätta vilka kategorier som dominerar och hur du känner inför dem.
 - Reflektera över hur väl din feed stämmer överens med dig som person.
 - **Tror du** algoritmen har “rätt” bild av vem du är?
 - **Kan** algoritmen förstärka intressen eller åsikter du redan har? Hur?
 - **Missar du** kanske annat som kunde vara spännande (ny musik, andra hobbies) för att din feed fokuserar på gamla vanor?
 - **Hur skulle du** kunna bredda ditt flöde? (t.ex. söka på nya ämnen, avfölja vissa konton, aktivt gilla nya saker)
 - Reflektera över vad du tycker är bra eller dåligt med algoritmen.
2. **Muntlig minipresentation** (i mindre grupp):
- Visa vilka kategorier som syns mest i ditt flöde och **varför** du tror de dyker upp.
 - Berätta om något **förvånande** eller **störande** du inte förstår varför du får se.

5. Sammanfattande frågor

1. **Tror du** algoritmen har “rätt” bild av vem du är?
2. **Kan** algoritmen förstärka intressen eller åsikter du redan har? Hur?
3. **Missar du** kanske annat som kunde vara spännande (ny musik, andra hobbies) för att din feed fokuserar på gamla vanor?
4. **Hur skulle du** kunna bredda ditt flöde? (t.ex. söka på nya ämnen, avfölja vissa konton, aktivt gilla nya saker)

6. Tips för att balansera ditt flöde

- **Sök medvetet** på nya ämnen du är nyfiken på, även om de inte är vanliga i din feed.
- **Rensa** bort ämnen du inte gillar genom att trycka “Jag är inte intresserad” (finns i många appar).
- **Prata med andra**: Se vad dina vänner har i sina flöden för att få nya idéer.

- **Var nyfiken:** Algoritmen belönar det du klickar på – våga klicka på något oväntat ibland.

Lycka till med att utforska din algoritm – och kom ihåg att du kan påverka vad som visas för dig online!

Lektion 5: Kurera algoritmen

(Hur snabbt anpassar sig algoritmer, och hur kan vi styra dem?)

Introduktion

Efter att ha funderat kring vad algoritmer “tror” om oss i förra lektionen, ska vi nu se **hur kraftfulla** dessa algoritmer är och **hur vi aktivt kan styra** våra flöden. Algoritmer är inte alltid “onda” eller “goda” – de är verktyg som visar innehåll baserat på våra beteenden. Denna lektion visar hur snabbt ett flöde kan fyllas av bara en typ av innehåll (t.ex. smink, fotboll, konspirationsteorier, politiska budskap) om vi konsekvent söker och interagerar med just det.

Syfte och mål

- **Förstå algoritmers drivkrafter:** Varför föreslår en plattform nyheter, videor eller produkter?
- **Upptäcka hur enkelt** det är att “mata” algoritmen med en viss typ av innehåll för att snabbt få ett homogent flöde.
- **Diskutera positiva och negativa effekter** av att algoritmen anpassar sig så snabbt.

Lektionsingång

1. Exempel från SVT Dokument inifrån, Priset vi betalar

- Berätta kort om hur SVT testade TikToks algoritm för att få endast sminktips i flödet, genom att medvetet klicka på sminkvideor och scrolla förbi annat. Diskutera hur fort flödet ändrades.

Avsnitt 3 cirka 30 minuter in

<https://www.svtplay.se/video/e4oaVbz/dokument-inifran-priset-vi-betalar/3-att-vara-en-vara?video=visa&position=1785>

2. Klassrumsdemonstration (frivillig)

- Om läraren vill: skapa ett nytt konto (eller radera all historik/”fortsätt titta” på ett redan befintligt konto) och **visa i realtid** på projektorn hur man “kurerar” algoritmen genom att bara söka på ett specifikt ämne, gilla liknande videor etc. Notera hur fort förslagen förändras.

Genomförande (steg-för-steg)

1. Introduktion till algoritmers logik (5 min)

- Påminn om att algoritmer är utformade för att hålla oss kvar så länge som möjligt på plattformen, genom att visa mer av det vi verkar gilla.
- Betona att **algoritmen inte är dålig i sig**; i bästa fall kan den hjälpa oss att hitta nya, meningsfulla läresurser.

2. Jämföra plattformar (10–15 min)

- Diskutera i mindre grupper hur olika plattformar (TikTok, Youtube, Instagram, Twitter m.fl.) anpassar flöden.
- Eleverna kan dela erfarenheter (eller observera lärarens exempel) om hur snabbt man kan få ett “filterbubbla-liknande” flöde.
- Frågeställningar:

- “Vad skiljer algoritmen på TikTok från den på Youtube?”
 - “Vilken plattform verkar snabbast anpassa sig?”
3. **Diskussion om aktiv kuration (5–10 min)**
 - Hur kan eleverna själva **aktivt styra** sina flöden? (t.ex. söka på nya ämnen medvetet, avfölja vissa konton, klicka på “jag är inte intresserad av detta innehåll” etc.)
 - Poängtera att det inte är fel att gilla en viss typ av innehåll, men att man kan missa andra intressanta saker om ens flöde blir för ensidigt.
 4. **Gruppövning eller helklass (10–15 min)**
 - **Scenario:** “Vad händer om vi bestämmer oss för att göra en ‘fotbollsklass’? Vi väljer att på projektorn bara söka på fotbollsklipp och gilla fotbollsrelaterade inlägg. Hur lång tid tar det innan flödet blir helt fotbollsdominerat?”
 - **Reflektera:** Vad hade hänt om man i stället matat algoritmen med politik, mordhistorier eller Andrew Tate-videor? Skulle det gå snabbare eller långsammare? Varför?
 5. **Sammanfattning (5 min)**
 - Påminn om att vi **inte** uppmanar elever att skapa nya konton privat, men att en klassrumsdemonstration kan belysa hur algoritmen fungerar.
 - Diskutera: “Hur kan denna kunskap hjälpa oss att få ett mer balanserat, lärorikt flöde i stället för att passivt matas med vad algoritmen ‘tror’ vi vill se?”

Diskussionsfrågor

1. **Varför kan algoritmer vara bra för lärande?** Kan ni ge exempel på hur ni hittat bra studieresurser tack vare en plattform?
2. **Finns det risker med att algoritmerna blir ‘för bra’ på att förutsäga vad vi gillar?**
3. **Är vissa plattformar “farligare” när det gäller att skapa ensidiga flöden? Varför?**
4. **Hur skulle du råda en vän som märker att deras flöde blivit väldigt ensidigt?**

Didaktiska tips

- **Var tydlig med syftet** när ni gör en klassrumsdemonstration av ny/”ren” algoritm: målet är att se hur systemet snabbt anpassar sig, inte att uppmana eleverna att göra konton eller experiment med egen data.
- **Knyt an till digital källkritik:** Understryk att när en algoritm bara visar en slags innehåll, kan det påverka hur vi uppfattar världen. Detta är en del av filterbubblor och ekokammare.
- **Använd “behind the scenes”-exempel:** Vissa plattformar berättar (i sin policy/FAQ) om hur deras algoritm fungerar. Visa korta citat för eleverna, så de förstår att detta inte är magi, utan programmerade regler.
- **Olika grad av restriktivitet:** Diskutera hur olika appar har olika “rekommendationsmotorer” – TikToks är ökänd för sin snabbhet, medan andra appar kanske kräver flera sökningar/gillningar innan den “läser in” en profil.

Fördjupning

- **Långa experiment över tid:** Om tid finns, kan man under en vecka (på projektorn vid lektionens start) konsekvent “mata” algoritmen med ett visst ämne. Följ sedan upp förändringarna i flödet dag för dag.

- **Koppla till nya ämnen:** Testa att “kurera” algoritmen mot ett intressant skolrelaterat ämne (ex. historieklipp, språklektioner) och se om det gör lärandet roligare. Diskutera därefter om elever upplever att de lärt sig något nytt.
-

Sammanfattning

I denna lektion lär sig eleverna hur **otroligt snabbt** algoritmer kan formas för att visa en viss typ av innehåll – både på gott och ont. Genom konkreta exempel eller experiment i klassrummet förstår de att vi kan **kurera** (styra) våra flöden genom våra beteenden online. Lektionen betonar att algoritmer i sig inte är dåliga, men att vår medvetenhet är avgörande för att undvika ensidiga flöden och filterbubblor.

Arbetsblad – Lektion 5: Kurera algoritmen

Hur snabbt anpassar sig algoritmer, och hur kan vi styra dem?

1. Inledning

Den här lektionen handlar om **hur algoritmer på sociala medier och videoplattformar fungerar**. De är inte goda eller onda i sig, utan *verktyg* som visar mer av det vi klickar på och interagerar med. Syftet är att visa **hur snabbt** flödet kan “låsa in” sig på ett enda ämne (t.ex. fotboll, smink, konspirationer) – och att vi själva kan **påverka** vad vi ser.

2. Övning: Hur “kurerar” man algoritmen?

Här är några **steg** du kan fylla i individuellt eller diskutera i grupp. Tanken är att förstå **hur** ett flöde kan bli ensidigt och **hur** du kan styra det.

1. Mina favoritplattformar:

- Vilka appar/platser (TikTok, Youtube, Instagram, Twitter osv.) använder du mest?
- Ser du att de föreslår liknande typer av innehåll hela tiden?

2. Scenario:

- “Du vill att flödet på en ny plattform snabbt ska handla om **bara** en sak – t.ex. fotboll eller smink.”
- Hur **gör** du? Sätt kryss i rutorna som beskriver hur du skulle “mata” algoritmen:

Åtgärd	Beskrivning	Kryssa i (✓/✗)
Sök bara på samma sak (t.ex. fotboll)	Jag skriver in “fotboll” i sökrutan och letar upp klipp om fotboll.	
Gilla/like samma kategori av inlägg	Jag gillar, kommenterar och delar fotbollsinlägg.	
Scrollar förbi annat	Jag ignorerar aktivt allt som inte är fotboll.	
Följer konton inom samma ämne	Jag följer flera fotbollskonton eller fotbolls-YouTubers.	

3. Reflektion:

- Hur lång tid **tror** du att det skulle ta innan ditt flöde blir *nästan* bara fotboll? (Några dagar, timmar, en vecka?)
- Varför **tror** du det går så fort?

3. Diskussion: För- och nackdelar

Fyll i tabellen nedan. Diskutera hur ett snabbt “filtrerat” flöde kan vara både positivt och negativt.

Fördelar (positivt)	Nackdelar (negativt)
Exempel: Man hittar mer innehåll inom sitt intresse snabbt.	Exempel: Man kan fastna i en “filterbubbla” och missa annan info.

4. Tips för aktiv “kurering”

1. **Hur kan du styra algoritmen till *mer varierat innehåll*?**
 - *Exempel:* Medvetet söka på nya ämnen, gilla olika saker, klicka “jag är inte intresserad” på vissa klipp.
2. **Vad händer** om du inte gör något alls – bara svajpar vidare?
 - Skriv en mening om hur algoritmen troligtvis **själv** formar ditt flöde.
3. **Har du någonsin försökt** påverka din feed på detta sätt? Hur gick det?

5. Fallstudie: “Fotbollsklass”

Scenario: Er lärare demonstrerar på projektorn att ni skapar ett nytt konto på en plattform. Ni letar bara upp fotbollsklipp och gillar dem en hel lektion. Diskutera hur snabbt ni tror klippen blir ensidiga.

1. **Hypotes:** Hur många videor/inlägg krävs innan flödet blir “monotont”? ...
2. **Alternativ:** Om ni **bara** matat algoritmen med *mordhistorier* eller *konspirationsteorier*, hade det gått snabbare/långsammare? Varför?

6. Sammanfattande frågor

1. **Varför kan algoritmer vara bra för lärande?**
...
2. **Finns det risker** med att algoritmerna blir för snabba på att förutsäga vad du gillar?
Hur kan det påverka dig?
...
3. **Vilken plattform tycker du är snabbast** på att “låsa in” sig på ett ämne? Motivera!
...
4. **Hur skulle du hjälpa en vän** vars flöde blivit väldigt ensidigt – t.ex. bara fotbollsrelaterat?
...

7. Avslutande reflektion

- **Vad tar du med dig** från den här lektionen om hur du kan styra (eller “kurera”) ditt eget flöde?

- **Hur** påverkar detta din källkritik? Kan det bli svårare eller lättare att hitta bred info om du låter algoritmen styra för mycket?

Kom ihåg: Algoritmerna är *programmerade* för att erbjuda mer av det du interagerar med. De i sig är varken goda eller onda, men du kan hamna i en filterbubbla utan att märka det. Genom att vara **medveten** kan du istället ta makten över ditt flöde – och se till att du upptäcker nya, intressanta saker!

Lycka till med att 'kurera' ditt flöde och hålla ögonen öppna för nya perspektiv!

Lektion 6: Räkna AI-genererade klipp

(Medvetenhet om ‘sanningsvakuum’ och hur våra sinnen kan luras)

Introduktion

Hur påverkas vår uppfattning av vad som är sant när digital teknik kan skapa hyperrealistiska bilder, videor och ljud? I denna lektion tittar vi närmare på fenomenet “sanningsvakuum”: vissa tror inte på något (allt är “fejk”), medan andra tror på allt (de ifrågasätter aldrig). Vi diskuterar hur lätt vi kan luras även när vi “vet” att något är falskt. Muller-Lyer-illusionen blir en symbol för hur våra sinnen fortsätter att se fel, trots att vi teoretiskt sett förstår sanningen.

Syfte och mål

- **Öka medvetenhet** om att vår hjärna kan bli lurad – även om vi “vet” bättre.
- **Utforska fenomenet** att vissa kan bli hyperkritiska (tror ingenting) medan andra är extremt godtrogna (tror allt).
- **Diskutera AI-genererat innehåll** och hur vi kan upptäcka det, trots vår benägenhet att tro på sådant som bekräftar vår världsbild.

Lektionsingång

1. **Visa en kort AI-genererad film eller klipp** som är tydligt överdrivet (men ändå trovärdigt). Fråga klassen:
 - “Hade ni kunnat gissa att detta är AI-genererat om ni inte fått veta det i förväg?”
2. **Introducera Muller-Lyer-illusionen** (de två parallella linjerna med pilar åt olika håll). Även när vi vet att de är lika långa, *ser* vi dem som olika långa. Diskutera parallellerna till hur AI-skapat material kan lura oss känslomässigt.

Genomförande (steg-för-steg)

1. **Förklaring av ‘sanningsvakuum’ (5 min)**
 - Beskriv hur en del människor blivit så skeptiska till allt på nätet att de inte litar på någon källa, medan andra är så godtrogna att de tror på allt.
 - Koppla detta till hur AI-genererat innehåll kan både öka misstron (“Inget är äkta”) och vilseleda de godtrogna (“Allt låter trovärdigt!”).
2. **Eleverna räknar AI-genererade klipp i sina egna flöden (10–15 min)**
 - **Instruktion:** Be eleverna öppna (valfritt) Instagram, TikTok eller en liknande app. Under t.ex. **5 minuters scrollande** ska de försöka identifiera om något de ser skulle kunna vara AI-genererat (bilder, korta videoklipp, filter, djupfejkat ljud m.m.).
 - **Notera antal fynd:** Eleverna antecknar hur många klipp eller bilder de uppfattar som eventuellt AI-genererade.
 - **Räkna procentsats:** När 5 minuter gått, räkna hur många klipp eleverna hann se totalt, och hur många de tror var AI-genererade. Varje grupp eller individ tar fram en grov **procentsats** (t.ex. “5 av 20 klipp = 25 %”).
 - **Reflektion:** Fråga om det var förvånande att siffran blev så låg/hög. Hur säkra är de på sina bedömningar?

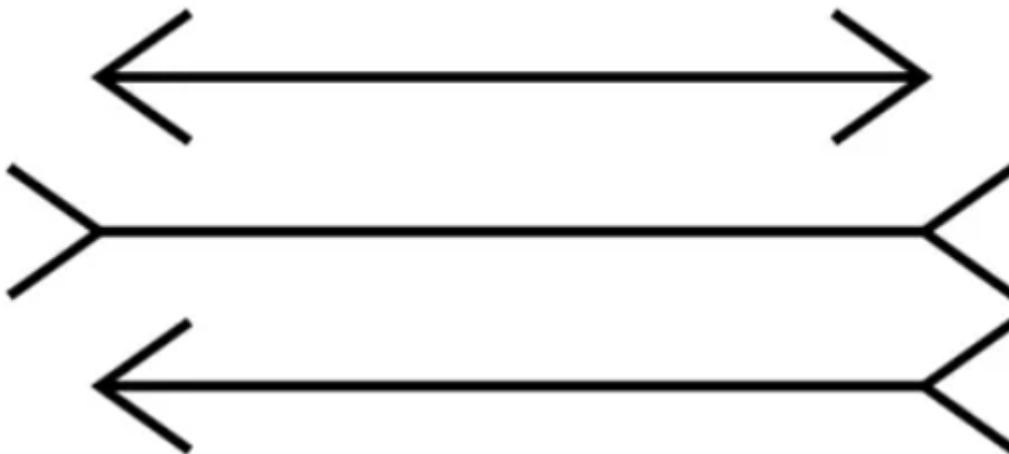
3. Reflektion: Varför blir vi lurade? (5–10 min)

- Diskutera i par eller smågrupper:
 - Vilka “tecken” letade ni efter för att avgöra om något var AI-genererat?
 - Var det svårare eller lättare än ni trodde att hitta AI-innehåll i ert flöde?
 - Hur mycket är bedömning och magkänsla, och hur mycket är *faktisk* kunskap?

4. Muller-Lyer-illusionen (5 min)

- Visa illusionen (projicera eller rita på tavlan).

<https://www.verywellmind.com/how-the-muller-lyer-illusion-works-4111110>



- Påminn om att trots att vi *vet* att linjerna är lika långa uppfattar vi dem ofta som olika långa.
- Dra parallellen till AI: Vår hjärna kan “se” något som äkta för att vi *vill* att det ska vara äkta (eller tvärtom), trots att det kanske är fejk.

5. Diskussion om bekräftelsebias (5–10 min)

- Förklara kort att vi tenderar tro på sådant som bekräftar vår redan existerande världsbild.
- Eleverna ger exempel på när de (eller andra) glatt delat en story/video som passade deras åsikt men sedan visade sig vara fejk.
- Poängtera sambandet med algoritmer: Om vi ofta klickar på innehåll som matchar våra åsikter, visar flödet mer av liknande innehåll – vilket kan förstärka vår tro på fejknyheter.

6. Sammanfattning (5 min)

- Betona vikten av att hitta en medelväg: *varken* automatiskt misstro allt *eller* svälja allt.
- Påminn eleverna om att även experter kan bli lurade av AI-förfalskningar. Källkritik är en process, inte något vi “alltid” klarar perfekt.

Diskussionsfrågor

1. Varför tror ni en del människor säger: “Jag tror ingenting på nätet”, medan andra verkar tro på allt?

2. Hur hjälpte 5 minuters scrollande er att inse hur ofta (eller sällan) AI-innehåll dyker upp i ert feed?
3. Hur spelar bekräftelsebias in när vi avgör om ett klipp är äkta eller fejk?
4. Vad kan hända om vi regelbundet sprider bilder eller videor som är AI-genererade men bekräftar vår åsikt?

Didaktiska tips

- **Var hänsynsfull mot integritet:** Ingen elev ska behöva visa sitt personliga flöde om de inte vill. Denna övning kan också göras med en **frivillig** elevs flöde visat på skärm, eller lärarens eget för demonstration.
- **Antal klipp:** Justera tiden och antalet klipp beroende på hur snabbt elever scrollar. 5 minuter eller 20 klipp brukar räcka för en översikt.
- **Anpassa efter ålder:** Yngre elever kan ha enklare att bara "räkna" misstänkt AI-innehåll, medan äldre elever kan fördjupa sig mer i diskussionen om bekräftelsebias och illusionerpsykologi.
- **Fortsättningen:** Koppla gärna denna lektion till fler exempel på optiska eller kognitiva illusioner för att bredda förståelsen av hur lätt våra sinnen påverkas.

Fördjupning

- **Statistik i helklass:** Sammanställ hur många procent AI-innehåll klasserna hittade och se om det skiljer sig mellan olika sociala medier. Diskutera varför TikTok vs. Instagram vs. Youtube kanske visar olika mycket AI-innehåll.
- **Knyt an till forskning:** Om tid finns, visa korta artiklar eller videor om hur vår hjärna bearbetar falsk information. Diskutera "kognitiva bias" och hur de påverkar vår kritiska förmåga.

Sammanfattning

Under denna lektion får eleverna konkret uppleva hur ofta (eller hur sällan) AI-genererat innehåll dyker upp i deras egna flöden, och hur svårt det kan vara att säkert identifiera vad som är äkta eller fejk. Genom Muller-Lyer-illusionen förstår de att hjärnan kan luras även när man *vet* bättre, och genom diskussion om bekräftelsebias inser de att vi ibland *vill* tro på visst innehåll. Syftet är att balansera mellan total misstro och naiv godtrogenhet, för att utveckla en mer **nyanserad och medveten** källkritisk hållning.

Arbetsblad – Lektion 6: Räkna AI-genererade klipp

Hur lätt är det att bli lurad och vad kan vi göra för att balansera mellan total misstro och blind tillit?

1. Uppvärmning: “Sanningsvakuum”

Instruktion: Läs kortfattat om begreppet “sanningsvakuum.” Fundera över varför vissa människor inte litar på något (allt är “fejk”) medan andra tror på allt (inget ifrågasätts).

Vad är ett “sanningsvakuum”?

I en värld där AI kan skapa övertygande bilder, videor och texter blir det allt svårare att avgöra vad som är sant. Detta har lett till ett fenomen kallat **sanningsvakuum**, där människor hamnar i två ytterligheter:

1) Total misstro – Vissa börjar ifrågasätta **allt** och tror att ingenting på nätet är sant. De avfärdar även trovärdig information, vilket kan leda till att de inte litar på nyhetskällor, forskare eller experter.

2) Blind godtrogenhet – Andra tror på **allt** de ser och delar vidare information utan att kontrollera källan. De luras lätt av AI-genererat innehåll, deepfakes eller manipulerade nyheter.

Båda ytterligheterna är problematiska – att misstro allt kan leda till cynism och konspirationstänk, medan att tro på allt kan göra att vi sprider felaktig information. Det viktiga är att utveckla en **balanserad och nyanserad källkritisk hållning** där vi ifrågasätter *lagom* och lär oss att känna igen både pålitliga och vilseledande källor.

Frågor:

1. Har du själv märkt av att vissa i din omgivning antingen är **extremt** kritiska mot allt eller **aldrig** ifrågasätter något online?
2. Hur kan AI-genererat material bidra till att öka både tvivel och godtrogenhet?

Skriv dina tankar här:

- ...
- ...

2. Praktisk Övning: Räkna AI-genererade klipp

Instruktion: Använd en app du gillar (TikTok, Instagram, YouTube, etc.). Under **5 minuter** ska du scrolla och försöka identifiera om något du ser *kan* vara AI-genererat (bild, video, ljud, eller filter). Ingen behöver visa sitt eget flöde öppet – du kan göra detta i tysthet.

1. **Anteckna antal klipp** du ser totalt (T):
 - T = ...
2. **Anteckna hur många** du misstänker är AI-genererade (A):
 - A = ...
3. **Räkna procentsats**

$$\text{Procent AI} = \frac{A}{T} \times 100\% \quad \text{\texttt{\text{Procent AI}} = \frac{\{A\}}{\{T\}} \times 100 \%}$$

- Procent AI \approx \texttt{\text{Procent AI}} \approx ... %

Reflektion:

- Var du **förvånad** över siffran? Var den högre eller lägre än du trodde?
- Hur **säker** är du på att klippen var AI-genererade? Var det tydligt eller mest magkänsla?

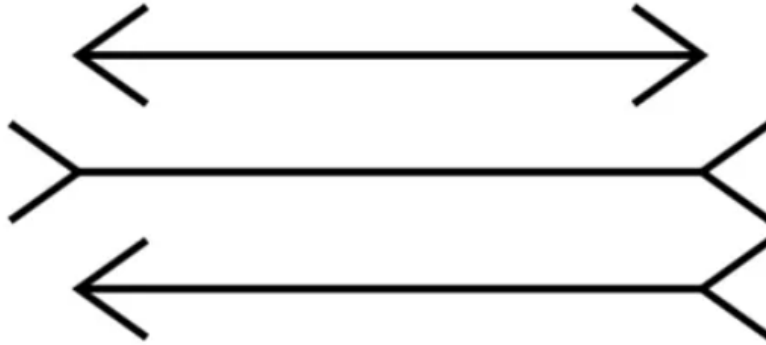
3. Varför blir vi lurade?

Instruktion: Diskutera i par eller grupp: Vilka “tecken” letade ni efter för att avgöra om något kunde vara AI-genererat? Hur mycket handlade om ren **intuition** och hur mycket om **konkreta ledtrådar** (t.ex. bildfel, onaturligt ljud, märkliga detaljer)?

1. **Tecken på AI:**
 - ...
 - ...
2. **Hur säker** var du på dina bedömningar?
 - () Mycket säker
 - () Ganska säker
 - () Osäker
3. **Var det svårare** eller **lättare** än du trodde att hitta misstänkt AI-innehåll? Varför?

4. Muller-Lyer-illusionen

Titta på **Muller-Lyer-illusionen** (två linjer med pilar åt olika håll). Även om vi *vet* att linjerna är lika långa, *ser* de olika långa ut.



Koppling till AI:

- Vår **hjärna** kan luras. Även om vi vet att något är falskt, så spelar det kanske inte någon roll om det bekräftar vår världsbild, eller våra åsikter. Därför är det inte säkert att det spelar någon roll om vi kan identifiera en bild eller en video som AI-genererad eller inte.

5. Bekräftelsebias – Vi tror gärna på det som passar oss

Instruktion: Fundera på när du eller någon du känner har delat en bild/video online för att den “stämde med” er åsikt, men sedan visade sig vara felaktig.

1. **Exempel:**
 - ...
2. **Hur** kopplar detta till “filterbubblor” och att algoritmer ofta visar oss mer av det vi redan gillar?
3. **Varför** kan AI göra att våra bubblor blir ännu starkare?

6. Avslutande frågor

1. **Varför säger vissa:** “Jag litar inte på något som läggs ut” medan andra “tror på allt”?
2. **Efter 5 minuters scrollande**, hur ofta verkar AI-innehåll faktiskt dyka upp i ditt flöde?
3. **Har** Muller-Lyer-illusionen hjälpt dig att förstå att vi kan “se fel” även när vi vet bättre? På vilket sätt?
4. **Vilka konsekvenser** ser du om vi rutinmässigt sprider AI-genererat fejk som stödjer vår världsbild?

7. Sammanfattning

Genom denna lektion har du:

- Upptäckt **hur ofta** (eller sällan) AI kan dyka upp i ditt flöde.
- Insett att våra hjärnor kan luras av illusioner och **bekräftelsebias**.

- Förstått att **varken extrem misstro eller blind tillit** är en bra strategi.

Reflektion: Hur tänker du i fortsättningen när du ser ett klipp som känns *för bra* eller *för konstigt* för att vara sant?

- ...

Tips:

- Var alltid **källkritisk** men behåll **nyfikenheten**. Prata med andra om dina fynd!
- Om du **misstänker** AI, jämför med andra källor eller fråga någon du litar på.

Lycka till med att navigera mellan total misstro och överdriven godtrogenhet – målet är en nyanserad och medveten hållning!

Lektion 7: AI och informationsspridning – hastighet, räckvidd och risker

Introduktion

Generativ AI har gjort det möjligt att producera enorma mängder innehåll på nolltid. Denna lektion syftar till att belysa **AI:s påverkan på informationsspridning** i samhället. Introduktionen kan inledas med att beskriva hur nyheter och rykten förr spreds långsammare via traditionella medier med granskning, medan idag kan en **AI-genererad text eller bild spridas viralt** inom timmar utan mänsklig faktagranskning. Poängtera att AI kan förstärka både det positiva (snabb spridning av viktig information) och det negativa (blixtsnabb spridning av falsarier). Exempel: Under kriser (som pandemin) spreds både livsviktig information och falska rykten mycket snabbt på nätet, delvis drivet av botar och automatiserade konton. Eleven bör förstå att hastigheten och volymen av information idag kräver att vi är extra vaksamma och kritiskt tänkande.

Genomförande (steg-för-steg)

1. **Start med konkret exempel:** Välj en nyhetshändelse (t.ex. en naturkatastrof eller stor nyhet) och berätta hur den spreds på olika plattformar. Nämn gärna om AI var inblandad (t.ex. deepfake-bilder som cirkulerade, eller automatiska konton).
2. **Visa statistik/bild:** Om möjligt, visa en enkel grafik över informationsspridning – t.ex. en kurva som jämför hur fort en nyhet nådde 50% av befolkningen år 1990 vs 2023. (Kan beskrivas muntligt om ej bild finns). Detta för att eleverna visuellt ska förstå skillnaden i hastighet.
3. **Diskutera drivkrafter:** Fråga klassen varför vissa inlägg sprids viralt. Kom in på algoritmer (mer om det i nästa lektion, men nämn att plattformar som YouTube/Facebook prioriterar visst innehåll) och även mänskliga psykologiska drivkrafter (vi delar sånt som väcker starka känslor). En UVA-studie visade t.ex. att Facebooks algoritm belönar känsloladdat, “inflammatoriskt” innehåll, vilket kan driva upp extrema inlägg news.virginia.edu
4. **Gruppövning – rykten i ringen:** Gör en analogiövning: viskleken. Eleverna står i cirkel och läraren viskar ett meddelande till en elev som skickar det runt. När det nått sista personen säger denne högt vad hen hörde – sannolikt har budskapet förändrats. Använd detta som analogi: på internet kan information förvrängas eller bli fel när den sprids snabbt och av många, **speciellt** om AI-robotar eller okontrollerade delningar är inblandade.
5. **Fallstudie med AI-botar:** Presentera ett konkret fall (t.ex. Twitter och COVID-19 misinformation). Berätta att forskare fann att en stor andel konton som spred ett visst

rykte var botar. Fråga eleverna: *"Varför skulle någon programmera botar att sprida falska rykten?"* (Möjliga svar: påverka opinion, ekonomisk vinning via klick, skapa kaos etc.)

6. **Sammanfattning med råd:** Summera att AI kan vara en **megafon** – det kan sprida saker jättefort, så vi måste vara snabba med källkritik också. Ge rådet att vid stora nyheter, sök information hos betrodda källor (myndigheter, stora nyhetsmedier) snarare än att lita på första bästa inlägg.

Lektionsingångar & diskussion

- **Ingångsvideo:** Om tids finns, visa en kort nyhetsinslag eller YouTube-klipp (2-3 min) som sammanfattar hur snabbt en viss desinformation spreds via sociala medier. Alternativt en animerad video om "the speed of fake news". Detta fångar intresset visuellt.
- **Diskussionsfråga:** *"Vilka fördelar ser ni med att information sprids blixtnsnabbt idag? Vilka nackdelar?"* – Låt båda sidor komma fram. Fördelar: tidig varning vid fara, fler röster kan höras. Nackdelar: svårt att hinna kontrollera, falskt kan spridas lika fort som sant.
- **Elevfråga:** *"Har ni själva vidarebefordrat/delat något online som ni senare insåg inte stämde?"* – Den som vill kan beskriva (frivilligt). Hur kändes det när man insåg det? Denna reflektion gör att de personligen inser värdet av att bromsa innan man delar.
- **Debatt i klassen:** Dela klassen i två och kör en mini-debatt: ena halvan argumenterar för att "AI och internet gör informationsspridning bättre", andra för att "AI och internet gör informationsspridning sämre". Efter några minuter får de byta sida för att se båda perspektiven. Detta tränar kritiskt tänkande – att se för- och nackdelar.

Didaktiska tips

- **Var försiktig med skrämself:** Det är lätt att denna lektion bara betonar faror. Försök ge en balanserad bild – teknik ger möjligheter också. Annars kan elever antingen bli rädda för allt online eller avfärda lektionen som domedagsprat.
- **Lokala exempel:** Om något rykte/flöde hänt i er stad eller skola (t.ex. ett falskt bombhot på sociala medier eller liknande), och det är lämpligt att diskutera, kan det göra lektionen väldigt verklig. Se bara till att hantera eventuella oroskänslor.
- **Samarbeta med samhällskunskapen:** Den här lektionen tangerar ämnen som yttrandefrihet, demokrati och mediepåverkan. Kanske samhällsläraren kan följa upp med en lektion om hur opinioner formas, eller historiska exempel på propaganda vs dagens situation.

- **Visualisera:** Använd tavlan eller digitala verktyg för att skissa informationsflöden (t.ex. rita upp en karta över världen och pilar hur en nyhet sprids via retweets). Många elever tar till sig bättre med visuellt stöd.

Forskning & verktyg

- **Falsa nyheter sprids snabbare än sanna:** En känd studie i Science (MIT-forskare, 2018) visade att falska nyheter på Twitter sprids märkbart snabbare och bredare än sanna nyheter, ofta på grund av att de var mer sensationella. Detta var dock drivet främst av människor som delade, inte enbart botar. Att förstå denna "dragningskraft" hos det falska är viktigt för elevernas medvetenhet.
- **Sociala botar och infodemi:** Forskning under COVID-19-pandemin fann att ~29% av kontona i vissa disinformationskampanjer var automatiserade botar, vilka genom retweets kunde få falska budskap att se populära ut. Detta kallas ibland en "infodemi", när en epidemi av felaktig information sprids parallellt med en sjukdom.
- **Algoritmers roll (intro):** Redan här kan nämnas (mer i nästa lektion) att algoritmer på sociala plattformar väljer ut vad du ser baserat på engagemang, vilket kan förstärka extremt innehåll news.virginia.edu

Exempel: Facebooks algoritm (före 2016) belönade inlägg med mycket reaktioner, vilket gynnade polariserande material.

- **Digitala verktyg:**
 - **Botometer (Indiana University)** – ett online verktyg (på engelska) där man kan ange ett Twitterkonto och få en bedömning om det verkar drivas av en bot. Elever kan testa med fejkade konton (kanske skapa ett eget experimentkonto).
 - **CrowdTangle (Meta)** – ett verktyg (för utbildare/journalister) som visar hur innehåll sprids på Facebook/Instagram. Om lärare har tillgång, kan man visa realtids-data av virala poster.
 - **Nyhetsvärderaren** (nämnd tidigare) – vissa uppgifter där handlar just om att se hur en nyhet spridits och var den dök upp först.

Arbetsblad

Arbetsbladet för Lektion 4 innehåller:

- **Flödesschema:** En övning där eleverna ska fylla i ett flödesschema över hur information sprids. De får en startpunkt ("En bloggare postar en AI-genererad bild med falsk bildtext") och en slutpunkt ("Många i klassen har sett bilden och tror på budskapet"). Mellan finns tomma rutor där eleverna i smågrupper fyller i möjliga steg

(t.ex. "någon delar på Twitter", "en YouTuber inkluderar bilden i en video", "journalister plockar upp det", etc). Detta tränar deras insikt i hur något kan "gå runt" på nätet.

- **Ordförståelse:** Kort definitionsruta där elever matchar begrepp med rätt förklaring: *viralt, bot, algoritm, infodemi, exponering*. T.ex. "*Bot*" – "*Ett automatiserat program som kan agera som en användare på internet*". Bra för att säkerställa att de hänger med i terminologin.
- **Korta frågor:** Ett par flervalsfrågor att svara enskilt på som repetition:
 1. "Vilket påstående stämmer bäst? A) AI gör att information alltid blir mer tillförlitlig eftersom den är datorgenererad. B) AI kan förstärka spridningen av både korrekt och felaktig information, så man måste vara källkritisk. C) AI-teknik används nästan aldrig i nyhetsspridning än." (*Rätt svar B*).
 2. "Vad menas med att en nyhet blir 'viralt spridd'? A) Att den sprids lika fort som ett datavirus infekterar datorer. B) Att många människor delar den vidare på kort tid. C) Att nyheten handlar om ett virus." (*Rätt svar B*).

Efter att arbetsbladet fyllts i kan ni gå igenom svaren tillsammans och diskutera där det blev olika svar, för att befästa kunskaperna.

Arbetsblad – Lektion 7: AI och informations spridning

1. Inledning – Snabbare än någonsin

AI har förändrat hur information sprids i världen. Tidigare gick nyheter genom journalister och redaktioner, medan de idag kan spridas direkt via sociala medier – ibland utan någon faktakontroll.

✅ **Fördelar:** Vi kan snabbt få viktig information vid kriser. ✅ **Nackdelar:** Falska rykten kan spridas lika snabbt som sanningar.

2. Flödesschema: Hur sprids information?

Instruktion: Nedan ser du en startpunkt och en slutpunkt. Fyll i mellanrummen med möjliga steg som kan ske när en AI-genererad bild eller text börjar spridas online.

Startpunkt: En bloggare postar en AI-genererad bild med falsk bildtext.

↓ Steg 1: _____
↓ Steg 2: _____
↓ Steg 3: _____
↓ Steg 4: _____

Slutpunkt: Många i klassen har sett bilden och tror på budskapet.

✦ Frågor att diskutera:

- Vilka kanaler (Facebook, TikTok, nyhetssajter, influencers) påverkar spridningen?
 - Hur kan AI bidra till att snabba upp spridningen?
 - Hur kan en enskild person bromsa eller förstärka spridningen?
-

3. Ordförståelse – Matcha begrepp med rätt förklaring

Dra streck mellan begreppet och dess förklaring.

1. **Viralt**
2. **Bot**
3. **Algoritm**
4. **Infodemi**
5. **Exponering**

- A) Ett automatiserat program som kan agera som en mänsklig användare online.
- B) En spridning av felaktig eller överdriven information som påverkar samhället negativt.
- C) När ett inlägg eller en nyhet snabbt får stor spridning på internet.
- D) Hur ofta en person ser en viss typ av innehåll online.
- E) Ett datorprogram som avgör vilka inlägg du ser i ditt flöde på sociala medier.

4. Gruppövning – Viskleken 2.0

Instruktion:

1. Sitt i en ring. En elev får ett meddelande viskat till sig av läraren.
2. Meddelandet viskas vidare från person till person.
3. Den sista eleven säger meddelandet högt.
4. Diskutera: Hur mycket förändrades budskapet? Vad påminner detta om i verkliga sociala medier?

Frågor att svara på skriftligt:

- Hur kan en nyhet förändras när den sprids av många människor?
- Vad kan AI och botar göra för att förstärka eller förvränga information?

5. Fallstudie – AI-botar och falsk information

Läs följande exempel:

Under COVID-19-pandemin spreds en konspirationsteori om att 5G-master spred viruset. En analys visade att en stor andel av inläggen kom från bot-konton, programmerade för att sprida rykten.

Diskutera:

- Varför skulle någon vilja programmera AI-botar att sprida falska rykten?
- Hur kan vi som individer motverka detta?

6. Korta frågor – Testa dina kunskaper

1. **Vilket påstående stämmer bäst?**
 - A) AI gör att information alltid blir mer tillförlitlig eftersom den är datorgenererad.
 - B) AI kan förstärka spridningen av både korrekt och felaktig information, så man måste vara källkritisk.
 - C) AI-teknik används nästan aldrig i nyhetsspridning.

(Rätt svar: B)
2. **Vad menas med att en nyhet blir 'viralt spridd'?**
 - A) Att den sprids lika fort som ett datavirus infekterar datorer.
 - B) Att många människor delar den vidare på kort tid.

C) Att nyheten handlar om ett virus.
(Rätt svar: B)

7. Avslutande reflektion

✎ Skriv ett kort svar på följande:

- Har du själv någon gång delat något på nätet som du senare insåg var felaktigt? Hur kändes det?
 - Hur kan du använda det du lärt dig idag för att bli mer medveten om informations spridning online?
-

📌 **Slutpoäng:** AI kan vara en megafon – det kan sprida saker jättefort, men vi måste vara lika snabba med vår källkritik! ✅

Lektion 8: Algoritmernas roll – från sökresultat till flöden

Introduktion

I den här lektionen zoomar vi in på **algoritmer** – de osynliga motorerna som avgör vilken information som dyker upp i våra sökresultat och sociala medieflöden. Syftet är att eleverna ska förstå att det inte är en neutral spegel av verkligheten de ser online, utan ett *kuraterat innehåll* format av algoritmer med olika syften (t.ex. maximera användartid, klick eller anpassa efter intresse). Introduktionen kan förklara vad en algoritm är i enkla ord: en serie instruktioner eller regler som datorer följer för att lösa en uppgift. Ge exemplet att Google använder algoritmer för att ranka träffar när vi söker, och Instagram/Facebook använder algoritmer för att välja inlägg i våra flöden. Emphasize att algoritmer kan skapa **filterbubblor**, där man mest ser sånt som bekräftar ens intressen/åsikter. Men forskning nyanserar också bilden: vi skapar delvis våra egna bubblor genom vad vi klickar på och följer

play.gu.se

. Denna lektion ger insikt i hur eleverna kan *bryta* ut ur bubbla och söka mer aktivt.

Genomförande (steg-för-steg)

1. **Fråga eleverna:** "Har du någon gång undrat varför just de här klippen dyker upp på din TikTok/Instagram/Snap Discover?" Låt dem kort diskutera med bänkgrannen. Introducera sedan idén att det är en algoritm som lär sig vad du gillat innan.
2. **Mini-föreläsning med exempel:** Förklara olika typer av algoritmer:
 - Sökalgoritmer (t.ex. Google PageRank + hundratals faktorer som påverkar vilka sajter som kommer först).
 - Sociala medie-algoritmer (t.ex. YouTubes rekommendationer eller TikToks "For You"-flöde baserat på tidigare beteende).
 - Flödesalgoritmer vs kronologiska flöden (Twitter var tidigare kronologiskt, gav större blandning; algoritmiska flöden visar det "mest relevanta" men kan leda till en snävare diet).
Nämn UVA-studien som fann att olika plattformars algoritmer hade olika effekter: Facebooks nyhetsflöde tenderade att göra att användare såg mer ensidigt (mer polariserat) innehåll, medan Reddit paradoxalt nog exponerade användare för mer varierade nyheter.
3. **Demo-sökning:** Kör en snabb live-demonstration: be två elever söka på Google efter samma nyckelord (t.ex. "klimathotet" eller "vaccin") på sina mobiler – om de är

inloggade på Google med olika sökhistorik, se om resultaten skiljer sig i toppträffar. Detta visar personalisering i praktiken. (Alternativt, visa skillnad mellan att söka i inkognito-fönster vs inloggat.)

4. **Algoritmexperiment i klassen:** Dela upp eleverna i grupper och ge varje grupp i uppgift att under veckan *manipulera* en algoritm lite: t.ex. en grupp skapar ett nytt YouTube-konto och bara tittar på vissa typer av videos (t.ex. bara nyheter vs bara gaming) och noterar hur rekommendationerna förändras; en annan grupp gör liknande på TikTok. Redan nästa dag kan vissa resultat ses. (Om en vecka är för långt, kan man diskutera hypotetiskt).
5. **Diskussion & strategier:** Samtala om hur man kan motverka att fastna i filterbubbla. Tips kan vara: följa medvetet olika typer av konton, använda olika nyhetskällor, söka aktivt efter alternativa perspektiv, rensa cache/inloggning för neutrala sökningar, och vara medveten om att "populärt på ditt flöde" inte är samma som objektivt populärt.
6. **Avslut:** Sammanfatta att algoritmer inte är "onda", men de har mål som inte alltid stämmer med sanningssökande. Elevens kritiska roll är att vara medveten och aktiv i sitt informationssökande, inte passiv konsument.

Lektionsingångar & diskussion

- **Ingång med quiz:** Gör ett snabbt handuppräkningsquiz: "Hur många här använder Google som sökmotor? Hur många har någon gång gått till sida 2 i sökresultaten?" (Sannolikt få). Diskutera att de flesta klickar på toppresultaten – vilket gör att de som hamnar där får all uppmärksamhet, tack vare Googles algoritm.
- **Diskussionsfråga:** "Tror ni två personer som har helt olika intressen får olika nyhetsflöden på t.ex. Facebook?" De flesta inser nog att ja, så är det. Följd: "Vad kan det få för konsekvenser för hur de ser på världen?" – Led in på att de kan få helt olika bild av verkligheten (t.ex. politiskt).
- **Mytsamtal:** "Är filterbubblor en myt?" – Nämn att vissa forskare säger att effekten överdrivits biblioteksbladet.se, för vi exponeras fortfarande för olika saker via vänner och nätverk. Elever kan få tycka till – har de själva sett motstridiga inlägg? (Troligen har de, men kanske man lättare ignorerar det som inte passar ens intresse). Bra att nyansera att algoritmer påverkar, men mänskligt val spelar roll också.
- **Case:** Beskriv en händelse: "Två kompisar, Ali och Bea, söker efter 'klimatförändringar'. Ali som brukar klicka på vetenskapsnyheter får upp forskning och FN-rapporter. Bea som mest klickar på sensationsnyheter får upp en YouTube-video med 'klimatskeptiker avslöjar bluffen'. Hur kan det bli så?" Låt eleverna analysera och koppla till algoritmer och tidigare beteende.

- **Framtidsfråga:** "Tycker ni att sociala medier borde visa allt i kronologisk ordning istället för med algoritmer? Varför/varför inte?" – Elevernas svar kan leda till diskussion om för- och nackdelar med personalisering vs kronologi.

Didaktiska tips

- **Förenkla tekniken:** Algoritmer kan låta högteknologiskt – använd analogier. T.ex. jämför med en bibliotekarie (algoritm) som lär känna dig och varje gång du kommer föreslår böcker liknande de du gillat innan. Elever förstår då hur rekommendationssystem fungerar.
- **Dataskydd:** Var diskret om du gör elevexperiment (punkt 4). Elever bör inte avslöja privata saker; syftet är att observera algoritmen, inte elevernas personliga beteende. Respektera integritet – kanske kör experimenten på gemensamma klasskonton istället för personliga.
- **Koppla till matte/programmering:** Om skolan erbjuder programmering kan du samarbeta – kanske låta elever som kan lite programmering demonstrera en enkel "algoritm" de kodat, eller diskutera begreppet träningsdata. Det kan bredda förståelsen.
- **Inkludera alla:** En del elever är mindre aktiva på sociala medier/sökning. Kanske finns någon som inte har TikTok – se till att de också känner sig inkluderade. Fokusera då mer på t.ex. YouTube eller Google som nästan alla använder.
- **Utmaning:** För de snabba eleverna, ge i uppgift att hitta inställningar på sina appar för att resätta eller ändra algoritmens rekommendationer (många appar har "intresseinställningar" eller möjligheten att klicka "visa mindre sånt här"). De kan sedan lära resten av klassen tipsen.

Forskning & verktyg

- **Algoritmpåverkan på åsikter:** En rapport i MIS Quarterly (Kitchens et al.) fann som sagt att plattformars design påverkar diversiteten i nyheter. Facebooks fokus på engagemang ledde till mer ensidigt nyhetsintag, Reddit gav mer balanserat. Det visar att alla algoritmer inte är lika – viktigt att diskutera att *hur* en algoritm är byggd spelar roll.
- **Filterbubblor – blandad evidens:** Peter Dahlgren (doktorand) menar att "filterbubbla" som term är förenklad; begreppet i forskning är hellre *selektiv exponering* play.gu.se Människor tenderar att söka sig till information som bekräftar deras åsikter, vilket algoritmer kan förstärka. Men många exponeras ändå för motstridiga nyheter via sociala nätverk. Summan är att vissa hamnar i bubblor, andra inte – det beror på både teknik och användarbeteende.

- **Sökkritik:** Skolverkets modul *Sökkritik och algoritmers betydelse* betonar att undervisning bör täcka hur sökresultat rankas och att elever bör öva på avancerade sökverktyg (som citationstecken, site:-sökning etc) för att inte bara nöja sig med toppresultaten.
- **Digitala verktyg:**
 - **DuckDuckGo** – en alternativ sökmotor som inte personaliserar lika mycket. Elever kan prova samma sökterm i Google och DuckDuckGo och se om resultaten skiljer sig.
 - **Feed synchronizing add-ons** – det finns webbläsartillägg som gör om t.ex. Twitter/X till kronologisk eller filtrerar bort vissa rekommendationer. Inte nödvändigt i undervisningen, men bra att känna till för intresserade elever.
 - **Algoritm-visualiseringar** – vissa sajter erbjuder interaktiva demo av hur en nyhet sprids i nätverk (t.ex. Bodynet projekt). Om möjligt kan man visa en sådan för att visuellt se "kluster" av likasinnade noder (som motsvarar en bubbla).

Arbetsblad

Arbetsbladet till Lektion 5:

- **Frågesport "Hur funkar algoritmen?"** – En rad påståenden där eleverna kryssar för Sant/Falskt:
 1. "En algoritm är samma sak som en AI." (Falskt – AI kan använda algoritmer men algoritm är generellt instruktioner för datorer)
 2. "Din egen aktivitet online påverkar vad algoritmer visar dig senare." (Sant)
 3. "Alla ser samma innehåll på en helt opersonlig plattform." (Falskt – de flesta plattformar personaliserar något, utom kanske Wikipedia)
 4. "Man kan själv påverka algoritmen genom vad man klickar eller aktivt väljer att följa." (Sant)
- **Mini-analys:** Två skärmdumpar (beskrivna i text om bild ej används) av YouTubes startsida för två hypotetiska användare – en som gillar sport, en som gillar politik. Eleverna får identifiera skillnader och skriva varför de tror skillnaderna uppstår.
- **Reflektionsruta:** "Vad tar du med dig?" – Eleverna skriver 2-3 meningar om något nytt de lärt sig om algoritmer och hur de ska tänka framöver när de scrollar sitt flöde.

Arbetsbladet syftar till att repetera begrepp och få eleverna att själva formulera sina insikter om algoritmernas roll.

Arbetsblad: Lektion 8 – Algoritmernas roll

Instruktioner Läs igenom arbetsbladet och besvara frågorna. Diskutera med din grupp där det är relevant. Syftet är att förstå hur algoritmer påverkar informationen vi ser online och hur vi kan bli mer medvetna om vårt informationsflöde.

Del 1: Vad är en algoritm?

1. Hur skulle du förklara begreppet "algoritm" för någon som aldrig hört ordet förut?
2. Ge ett exempel på en algoritm du använder i din vardag (det kan vara på en sökmotor, sociala medier eller en streamingtjänst).
3. Vilka faktorer tror du påverkar vad som visas i ditt flöde på TikTok, Instagram eller YouTube? Lista minst tre.

Del 2: Sant eller falskt?

Markera om följande påståenden är **sanna** eller **falska** genom att sätta ett kryss (✓).

Påstående	Sant (✓)	Falskt (✓)
En algoritm är samma sak som en AI.		
Algoritmer påverkar vad du ser på sociala medier.		
Alla sökresultat på Google är samma för alla användare.		
Filterbubblor kan göra att vi bara ser innehåll som bekräftar våra egna åsikter.		
Man kan själv påverka algoritmer genom vad man klickar på.		

Del 3: Analys av algoritmer

1. **Jämför sökresultat**
 - a) Gör en Google-sökning på "klimatförändringar". b) Jämför dina topp 5 sökresultat med en klasskamrats. c) Finns det skillnader? Vad kan det bero på? Diskutera.
2. **Filterbubblor**
 - a) Vad är en filterbubbla?
 - b) Hur tror du att filterbubblor kan påverka samhället? Diskutera i grupp.
3. **Sociala medieflöden**
 - a) Hur skulle ditt TikTok- eller Instagram-flöde se ut om du bara tittade på en viss typ av videor (t.ex. sport, politik, mode) i en vecka?
 - b) Vad kan du göra för att få ett mer varierat flöde?

Del 4: Reflektion

1. Vad var det mest överraskande du lärde dig i dagens lektion?
2. Hur kan du använda denna kunskap för att förbättra ditt sätt att söka information?

3. Tycker du att algoritmer borde vara mer genomskinliga så att vi förstår hur de fungerar? Varför/varför inte?

Extra utmaning

Vill du testa hur algoritmer fungerar i praktiken?

- Skapa ett nytt YouTube-konto och titta endast på en viss typ av videor i en vecka. Hur påverkas dina rekommendationer?
- Prova att söka på samma sak i Google och DuckDuckGo. Ser du skillnader?

Lämningsinstruktion:

- lämna in arbetsbladet till din lärare efter lektionen.
- Diskutera dina svar i grupp innan du skriver ner din slutliga reflektion.

Sammanfattning:

Algoritmer styr vad vi ser online – så vår digitala värld är inte alltid en spegel av verkligheten, utan en anpassad version av den. Genom att förstå detta kan vi bättre hantera information och bryta ut ur filterbubblor!

Lektion 9: Källtillit – att hitta och lita på trovärdiga källor

Introduktion

Källkritik handlar inte bara om att tvivla – utan också om att lära sig **vad man kan lita på**. Denna lektion fokuserar på **källtillit**, alltså konsten att identifiera trovärdiga källor och bygga upp en sund skepsis utan att bli cynisk. Syftet är att eleverna ska förstå att om man misstror *allt* riskerar man att missa sann information, men om man är för godtrogen kan man luras av falsk information. Balansen är viktig. I introduktionen kan du lyfta exempel på etablerade källor (t.ex. public service, vetenskapliga tidskrifter) och kontrastera mot obskyra bloggar eller okända avsändare. Poängen är att eleverna ska utveckla en känsla för **vad som kännetecknar en pålitlig källa**, särskilt i en tid där manipulerade medier och AI-fabricerat innehåll förekommer.

Genomförande (steg-för-steg)

1. **Brainstorming " trovärdig källa"**: Be klassen spontanropa eller lista på papper vad de anser är exempel på trovärdiga informationskällor (t.ex. SVT Nyheter, NE.se, Skolverket). Samtidigt, lista exempel på mindre trovärdiga källor (t.ex. anonyma forum, reklam-mail). Diskutera varför de placeras i olika kategorier.
2. **Kriterier för trovärdighet**: Gå igenom faktorer som ger källtillit: transparens (att det framgår vem som ligger bakom), expertis (att avsändaren har relevant kunskap), granskning (t.ex. nyheter med redaktionell process), oberoende (ej uppenbar bias/tendens). Du kan skriva upp dessa kriterier och förklara med exempel. Åsa Wikforss har t.ex. påpekat att reda ut att medier med ansvarig utgivare och faktagranskning generellt är mer tillförlitliga.
3. **Övning "Hitta källan"**: Dela ut ett påstående eller nyhet till varje grupp (eller låt alla jobba med samma). Exempel: "Forskare har bevisat att ungdomars IQ sjunker av sociala medier". Låt eleverna leta reda på *ursprungskällan* till påståendet. Kanske hittar de en kvällstidningsartikel som hänvisar till en studie. Be dem bedöma studiens trovärdighet utifrån kriterierna (är det en vetenskaplig publikation? var?). Detta tränar förmågan att leta bakåt till källan och värdera den – centralt för källtillit.
4. **Diskutera källhierarki**: Förklara begreppen primärkälla, sekundärkälla och tertiärkälla. Ge exempel (primär: forskningsrapport; sekundär: en tidningsartikel om rapporten; tertiär: en blogg som citerar tidningen). Fråga: "*Vilken är bäst att lita på, och varför?*"

5. **Reflektion:** Avsluta med att eleverna enskilt skriver ner *tre kännetecken* de själva kommer att titta efter för att avgöra om en källa är pålitlig. Detta blir deras personliga checklista framöver.

Lektionsingångar & diskussion

- **Ingång med fabel:** Berätta en kort analogi: "*Föreställ er att ni är i en främmande stad och ska äta lunch. Skulle ni lita mest på en rekommendation från er resehandbok, från en okänd person på gatan, eller från en vän som bor i staden?*" Låt eleverna svara och koppla det till informationskällor – vi litar olika mycket på källor baserat på erfarenhet och relation.
- **Diskussion:** "*Kan man lita på Wikipedia?*" – En klassiker som öppnar för resonemang. Några kanske säger ja med förbehåll ("kolla historiken, referenserna"), andra nej. Lyft fram att Wikipedia har många ögon på sig men också varierande kvalitet; och att den kan vara en bra utgångspunkt om man verifierar via referenserna.
- **Diskussionsfråga:** "*Vad gör att du litar på en nyhetssida?*" – Är det designen, att du hört namnet förut, att den länkas från någon du litar på? Detta får eleverna att medvetandegöra sina egna kriterier (eller brist på kriterier).
- **Aktuellt exempel:** Om det finns en pågående debatt om "fake news" eller misstro mot media i samhället, koppla in det. Exempel: diskussionen om "filterbubblor" och minskad tillit till media – hälften av svenskar oroar sig för filterbubblor enligt Internetstiftelsen svt.se

Fråga eleverna om de känner oro för att *de* kanske fastnar i ensidiga flöden och hur det påverkar deras tillit.

Didaktiska tips

- **Elevmedverkan:** Låt eleverna själva komma med exempel på källor de litar på och inte litar på. Detta gör diskussionen mer engagerande än att läraren levererar alla exempel.
- **Konkreta checklistor:** Dela ut eller visa en enkel checklista för att bedöma källor (det finns färdiga modeller från t.ex. Statens Medieråd eller brittiska CRAAP-testet – Currency, Relevance, Authority, Accuracy, Purpose – översatt till svenska i många skolmaterial).
- **Rollspel:** Som alternativ övning kan man göra rollspel där några elever spelar "källa" (ex. en journalist, en bloggare, en AI-bot) och klassen får ställa frågor för att avgöra trovärdigheten. Det blir ett annorlunda sätt att levandegöra källtillit.

- **Koppla till studieteknik:** Påpeka att källtillit också är viktigt i skolarbeten – t.ex. när de skriver uppsatser behöver de välja bra källor. Så detta är inte bara för nyheter, utan även för studier.
- **Lugn och realism:** Försök motverka att eleverna blir alltför cyniska eller paranoida. Poängtera att de flesta stora etablerade nyhetskällor *försöker* vara korrekta och är oftast pålitliga – men att misstag och bias kan ske. Balans är nyckeln.

Forskning & verktyg

- **Källtillitens betydelse:** Forskning av Åsa Wikforss m.fl. visar att ungdomar behöver lära sig *hitta* pålitliga källor, inte bara ifrågasätta allt. Utan den förmågan riskerar man att bli cynisk eller falla för konspirationsteorier för att man saknar orientering i vad som är trovärdigt.
- **Public service och medievana:** Studier indikerar att unga från hem med mycket nyhets- och mediekonsumtion har bättre både källkritik och källtillit. De har alltså genom övning lärt sig vilka källor som oftast är tillförlitliga. Det är något skolan kan hjälpa de andra eleverna med – att introducera dem till kvalitetssajter och källor.
- **Dunning-Kruger-effekten:** Nära kopplat till källtillit – de minst kunniga överskattar ofta sin förmåga. Att känna till den biasen kan få elever att förstå vikten av ödmjukhet: man ska vara nyfiken och öppen för att man kan ha fel och att det finns bättre källor.
- **Digitala verktyg:**
 - **Nyhetsvärderaren (online test)** – ett verktyg utvecklat i forskning (bl.a. av Nygren) där elever kan testa sin källkritiska förmåga genom att bedöma autentiska nyheter och sedan få facit.
 - **Källkritikbyrån** – en svensk resurs av journalister som granskar virala påståenden. Deras artiklar visar konkret hur experter tänker när de faktakollar.
 - **Media Bias Chart** (finns online) – om ni vill diskutera politisk vinkling kan en mediekarta över höger/vänster och faktaopålitlighet illustrera var olika källor befinner sig. Var försiktig att inte presentera det som absolut sanning, men som ett diskussionsunderlag.

Arbetsblad – Lektion 9: Källtillit

(Hur hittar vi trovärdiga källor, och varför är det viktigt?)

1. Övning: Källranking

Nedan följer en lista på **10 källor**. Din uppgift är att **rangordna** dem från 1 till 10, där **1** är den källa du **litar mest på** och **10** är den du **litar minst på**. Tänk på **hur** du bedömer dem: Har de redaktörer? Ansvarig utgivare? Är de kända för kvalitet eller tendens?

1. En **större dagstidning** (t.ex. DN/SvD)
2. **Wikipedia**
3. **Forskning.se** (svensk vetenskaplig nyhetsportal)
4. Ett **populärt Instagram-konto** med nyheter
5. **Skolverket.se**
6. En **AI-chatbot** (ex. ChatGPT)
7. **Bloggen "HealthyLife"** (okänd skribent)
8. **SVT Nyheter** (public service)
9. **Källkritikbyrån** (en svensk faktagranskningstjänst)
10. En **anonym insändare** på ett forum

Instruktion:

1. Rangordna källorna i den ordning du litar mest → minst på.
2. Välj dina **topp 3** (1:a, 2:a, 3:e) och förklara **varför** du placerar dem i toppen.
3. Välj dina **botten 3** (sist, näst sist, tredje sist) och motivera **varför** du har minst förtroende för dem.

Utrymme för dina anteckningar:

- **Rangordning:**
 1. ...
 2. ...
 3. ...
 4. ...
 5. ...
 6. ...
 7. ...
 8. ...
 9. ...
 10. ...
- **Mina topp 3 – Motivering:**
 1. ...
 2. ...
 3. ...
- **Mina botten 3 – Motivering:**
 1. ...
 2. ...

3. ...

2. Checklista för Trovärdighet

Nedan följer fyra **kriterier** som ofta används för att bedöma en källas pålitlighet: **transparens**, **expertis**, **granskning** och **oberoende**. I rutan under varje kriterium kan du skriva exempel på källor som **uppfyller** kriteriet.

1. **Transparens**

- Förklaring: Det är tydligt **vem** som ligger bakom källan, och **hur** de arbetar. En tydlig “Om oss”-sida, ansvarig utgivare, kontaktuppgifter etc.
- **Exempel på källa** som är transparent:

...

2. **Expertis**

- Förklaring: Källan har **faktiskt kunnande** inom området, t.ex. forskargrupp eller proffs inom ämnet.
- **Exempel på källa** som visar expertis:

...

3. **Granskning**

- Förklaring: Källan låter sig **faktagranskas**, har ett redaktionellt arbete eller peer review (vetenskaplig tidskrift).
- **Exempel på källa** med tydlig granskning:

...

4. **Oberoende**

- Förklaring: Källan är **icke-partisk**, har ingen uppenbar ekonomisk eller politisk agenda som påverkar innehållet.
- **Exempel på källa** som verkar oberoende:

...

Diskussionsfrågor:

- Vilket av dessa kriterier är **viktigast** för dig när du söker information?
- Kan en källa vara “bra” på ett område men “dålig” på ett annat? Hur märker du det?

3. Fallstudie: “SmartStudentBlog.com”

Scenario: En klasskompis delar en artikel i klassens chatgrupp om en ny *mirakelmetod* för att plugga dubbelt så effektivt. Artikeln påstår att “du bara behöver följa denna 3-stegsmetod i 2 veckor, så höjs dina betyg automatiskt.” Källan anges som **SmartStudentBlog.com**.

Svara på följande frågor med **stöd av checklistan** ovan (transparens, expertis, granskning, oberoende) och din egen källranking.

1. **Verkar källan ha expertis i ämnet (studieteknik)?**

...

2. **Kan du hitta om någon oberoende part** (t.ex. en universitetslärare eller faktagranskare) nämnt eller granskat metoden?
...
3. **Skulle du lita på denna information?** Varför eller varför inte?
...
4. **Var kan du hitta mer pålitlig information om studieteknik?**
...

Reflektion: Vilka **varningsklockor** ringer för en blogg som “lovar” fantastiska resultat utan att hänvisa till någon forskning? Skulle du behöva **mer bevis** innan du tror på den?

Användning av arbetsbladet

1. **Individ eller grupp:** Övningarna kan göras på egen hand eller i smågrupper.
2. **Diskussion:** Diskutera svaren i helklass efteråt. Var ni oense kring vissa källors pålitlighet?
3. **Koppla till verkligheten:** Uppmana eleverna att kopiera in **riktiga** exempel nästa gång de stöter på påståenden eller mirakelmetoder online.
4. **Ständig uppdatering:** Kom ihåg att nya källor dyker upp hela tiden. Denna **källranking** och **checklista** är ett levande dokument som ni kan återkomma till och ändra efterhand.

Lycka till med att utveckla er källtillit – och kom ihåg att kritiskt tänkande handlar om både skepsis och förtroende för rätt slags källor!

Lektion 10: Lär känna verktygen

(Praktisk inblick i hur enkelt det är att skapa fejkade bilder och videor)

Introduktion

I denna lektion får ni översikt över några **kostnadsfria** och **inloggningsfria** AI-verktyg som gör det möjligt att snabbt och enkelt skapa fejkade bilder eller videor. Syftet är att visa **hur lätt** det kan vara att skapa (eller manipulera) material som sedan kan spridas vidare som “nyhet” eller “bevis”. Därmed blir det tydligt varför vi behöver god källkritik och medvetenhet.

Obs: Dessa verktyg **rekommenderas inte** för eleverna att använda på egen hand, särskilt inte om det krävs att de skapar konton eller lämnar personliga uppgifter. Syftet är snarare att läraren ska kunna **demonstrera** verktygen i helklass, så att alla förstår hur smidigt det går att fabricera övertygande fejk.

Syfte och mål

- **Visa hur enkelt** det är att fejka bilder och videor utan avancerade kunskaper.
- **Diskutera** etiska och moraliska aspekter kring användningen av sådana verktyg.
- **Stärka** elevernas förståelse för varför källkritik är avgörande i en digital värld.

Lektionsingång

1. **Kort diskussion:**
 - Fråga eleverna: “Har ni sett exempel på fejkade bilder eller videor på nätet?” och “Trodde ni att man behöver proffsig utrustning eller dyr mjukvara för att göra det?”
 - Poängtera att nu ska ni **tillsammans** se hur snabbt och enkelt det faktiskt kan vara.
2. **Koppla till föregående lektion (prebunking):**
 - Påminn om hur man kan använda fejknyheter i propaganda. Dessa verktyg gör det ännu tydligare hur lätt manipulation kan ske.

Genomförande (steg-för-steg)

1. **Demonstration av Deepfake.civai.org (5–10 min)**
 - Gå till [Deepfake.civai.org](https://deepfake.civai.org) och visa hur man utan inlogg laddar upp eller tar en bild av sig själv och automatiskt genererar en “fejkad version” (t.ex. som fånge i ett fängelse, som militär, som uteliggare).
 - Diskutera: “Hur övertygande ser detta ut?” “Skulle ni kunna luras om detta dök upp i ett nyhetsflöde?”
2. **Live Portrait (5–10 min)**
 - Besök [Liveportrait.org](https://liveportrait.org) och ladda upp en stillbild och en kort video. Visa hur programmet “animerar” bilden till att göra gester eller ansiktsuttryck från videon.
 - Reflektera: “Vad skulle hända om vi använde en kändisbild?” “Hur kan detta utnyttjas för falska videoklipp?”

3. **Pika.art (5–10 min)**
 - Titta på [Pika.art](#) och demonstrera hur snabbt man kan göra korta videoklipp eller “inpainta” (lägga till nya objekt, t.ex. en alien) i en befintlig video.
 - Diskutera hur detta kan användas för skämt, men också för desinformation.
4. **Praktisk översikt (valfritt i helklass eller liten grupp)**
 - Låt eleverna (frivilligt) ställa upp med en bild för att testa hur verktyget förändrar deras utseende eller omgivning.
 - Var noga med **integritet**: ingen tvingas dela bild på sig själv. Det kan räcka med en kändisbild eller fria stockbilder från nätet.
5. **Etisk diskussion (5–10 min)**
 - Ställ frågor:
 - “Varför kan det vara problematiskt att kunna ‘klä ut’ någon till fånge eller militär så lätt?”
 - “Vad händer om man använder detta för elaka skämt (mobbing) eller politisk manipulation?”
 - “Hur kan vi se till att elever (eller vuxna) lär sig hantera dessa möjligheter på ett ansvarsfullt sätt?”

Diskussionsfrågor

1. **Vilka möjligheter** ser ni med dessa verktyg (t.ex. kreativt filmskapande, konstprojekt)?
2. **Vilka risker** är uppenbara? Kan ni komma på fler negativa användningsområden än bara fejknyheter?
3. **Borde det finnas lagar** eller regler som begränsar användningen av deepfake-verktyg?
4. **Hur kan vi** (individer, skolor, samhället) öka medvetenheten om hur enkelt det är att skapa fejk och därmed stärka källkritik?

Didaktiska tips

- **Använd projektor**: Låt läraren själv genomföra demon, för att undvika situationer där elever laddar upp bilder de inte bör dela.
- **Fokus på lärande, inte på bus**: Förtydliga att ni visar tekniken för att öka digital medvetenhet. Varna för olaglig eller kränkande användning.
- **Knyt an till andra ämnen**: I bild/estetiska ämnen kan man se fördelarna (kreativitet), i samhällskunskap betonar man riskerna (propaganda, integritetskränkningar), i teknik kan man diskutera AI-modeller i stort.
- **Omtanke om integritet**: Om någon elev vill testa med en egen bild, säkerställ att de gör det frivilligt och förstår konsekvenserna. Diskutera även rätten till sina egna bilder.

Fördjupning

- **Gå tillbaka till er fejknyhet** från föregående lektion (prebunking). Förbättra den nu med AI-genererade videoklipp eller animerade porträtt – se hur “ trovärdig ” ni kan göra en totalt påhittad story.
-

Arbetsblad – Lektion 10: Lär känna verktygen

Hur enkelt är det att skapa fejkade bilder och videor?

1. Syfte

I denna lektion ska ni lära er **hur enkelt** det kan vara att manipulera eller fejka bilder och videoklipp med hjälp av AI. Ni kommer att se exempel på kostnadsfria och inloggningsfria verktyg. Förståelsen för hur **lätt** detta är att göra ökar behovet av **källkritik** och etiskt ansvar.

2. Översikt av AI-verktyg

Nedan följer **tre** exempel på AI-sajter som kan användas för att snabbt generera eller manipulera bilder och videor. Er lärare kan demonstrera i helklass, eller ni kan observera en redan inspelad demo.

1. **Deepfake.civai.org**
 - Ladda upp en bild eller ta ett foto och gör om dig själv (t.ex. fånge, soldat).
 2. **Live Portrait** (liveportrait.org)
 - “Animerar” en bild utifrån en kort video (gester, ansiktsuttryck).
 3. **Pika.art**
 - Skapa korta videoklipp, lägg till objekt i befintliga filmer (“inpainta”).
-

3. Demonstration (Lärrarledd)

1. **Titta på** hur läraren (eller en frivillig) laddar upp en **ofarlig** bild i Deepfake.civai.org.
 2. **Se resultatet:** Hur trovärdig är den fejkade varianten?
 3. **Diskussion:** Skulle du ha kunnat genomskåda att bilden var manipulerad?
-

4. Etik och integritet

Instruktion: Svara kort på frågorna nedan (enskilt eller i par).

1. **Vad tycker du** är problematiskt med att kunna “klä ut” någon i en fejkad bild eller video?
 - ...
 2. **Hur kan detta utnyttjas** för elakt beteende (mobbing) eller politisk manipulation?
 - ...
 3. **Skulle du vara bekväm** med att någon annan la upp en manipulerad bild av dig utan din tillåtelse? Varför/varför inte?
 - ...
-

5. Diskussion: Möjligheter och risker

Instruktion: Notera **minst två** möjligheter och **minst två** risker med dessa AI-verktyg.

Möjligheter (positiva användningsområden)

- ...
- ...

Risker (negativa användningsområden)

- ...
- ...

Frågor att tänka på:

- Bör det finnas lagar som begränsar hur man får använda dessa tekniker?
- Vem bär ansvaret: den som **skapar** fejk eller den som **sprider** den?

6. Koppling till källkritik

1. **Kan** du se hur detta anknyter till tidigare lektioner om fejknyheter och prebunking?
 - ...
2. **Hur** kan förmågan att skapa deepfakes på några minuter förändra vårt förtroende för foto/video som bevis?
 - ...

7. Fördjupning (valfritt)

1. **Fortsätt med er fejknyhet** (från lektion om prebunking):
 - Förbättra den med AI-skapade bilder eller animerade porträtt.
 - Diskutera hur mycket mer trovärdig – eller skrämmande – den blir.
2. **Samhällsperspektiv:** Hur bör myndigheter, skolor och medier informera om dessa verktyg?
 - ...

8. Sammanfattande reflektion

Instruktion: Avsluta med att svara på dessa frågor:

1. **Vad tar jag med mig** från denna lektion?
 - ...
2. **Hur** kan jag använda kunskapen om hur enkelt det är att skapa fejkade klipp när jag nästa gång ser en “chockerande” bild eller video?
 - ...
3. **Vilka steg** kommer jag ta för att dubbelkolla äktheten av bildmaterial i framtiden?
 - ...

Kom ihåg: Syftet med att visa dessa AI-verktyg är att **öka medvetenheten** om hur lätt det är att manipulera innehåll. Med den vetskapen blir du en **vaksam** och **ansvarstagande** digital medborgare!

Lektion 11: Omvänd bildsökning

(Hur och varför vi kontrollerar en bilds ursprung för bättre källtillit)

Introduktion

I en värld där bilder sprids blixtnabbt i sociala medier är det lätt att bli lurad av ett manipulerat foto eller ett helt påhittat sammanhang. Denna lektion fokuserar på **omvänd bildsökning** – en metod där man söker på själva bilden för att ta reda på var den kommer ifrån, om den använts i andra sammanhang eller om den är modifierad.

Syftet är att visa eleverna hur de kan **högerklicka på en bild** och göra en “visuell sökning” (t.ex. Google Lens) för att sedan **diskutera källtillit**: Varför är vissa källor mer pålitliga än andra? Hur vet vi att en nyhetsplattform (eller person som postar bilden) faktiskt berättar hela sanningen?

Syfte och mål

- **Lära eleverna** att använda omvänd bildsökning för att snabbt hitta bildens ursprung eller annan användning online.
- **Skapa medvetenhet** kring hur bilder kan tas ur sitt sammanhang eller manipuleras.
- **Diskutera källtillit**: Varför man ofta kan lita på vissa källor mer än andra, och vilka faktorer som spelar in (t.ex. transparens, bakgrund, track record).

Lektionsingång

1. Exempel på en bild

- Välj gärna en bild som påstås visa något sensationellt. Exempelvis ett manipulerat foto av en känd politiker (t.ex. “Trump for Swifties” – Donald Trump med en Taylor Swift-tröja), en känd artist i en konstig situation, eller en historisk bild som getts en ny falsk kontext.
- Fråga: “Ser detta äkta ut? Hur kan vi ta reda på om bilden är riktig eller fejk?”

2. Kort diskussion

- “Hur ofta ser ni bilder i era sociala flöden som verkar lite misstänksamma? Har ni någon gång ifrågasatt en bild och försökt hitta källan?”

Genomförande (steg-för-steg)

1. Visa hur man gör en omvänd bildsökning (5–10 min)

- Öppna Google i en webbläsare, klicka på “Bilder”. I vissa webbläsare kan man högerklicka direkt på en bild och välja “Sök efter bilden på webben” eller “Sök med Google Lens”.
- Alternativt: gå till images.google.com, klicka på kameran/ladda upp bild och välj den misstänkta bilden.
- Förklara att Google (eller andra sökmotorer) då visar var bilden (eller liknande bilder) finns publicerade.

2. Eleverna testar själva (10–15 min)

- Låt eleverna i par eller smågrupper välja en bild (antingen en de själva hittar online eller en bild du delar i digital form).
- De gör en omvänd bildsökning och undersöker:
 1. **Var** dyker bilden upp?
 2. **Hur** beskrivs bilden på olika webbplatser?
 3. **Finns** det olika versioner av bilden (redigerad vs. original)?

3. Diskussion om källtillit (5–10 min)

- När eleverna hittat var bilden först publicerades, eller hur den spridits, diskutera:
 - Är **ursprungskällan** en etablerad nyhetsorganisation, en blogg, ett forum eller en privatperson?
 - Vilka **intressen** kan avsändaren ha (politisk agenda, kommersiella motiv, skämt)?
 - Hur **bedömer** vi om en källa är mer tillförlitlig än en annan (t.ex. god historik, faktagranskning, tydliga kontaktuppgifter)?

4. Exempel på kontroversiella bilder (valfritt)

- Visa gärna ytterligare exempel, t.ex. en bild på en kändis “i fel sammanhang” eller en gammal bild som nu påstås vara något annat.
- Eleverna får träna på att snabbt göra omvänd bildsökning och jämföra olika källor.

5. Sammanfattning (5 min)

- Betona värdet i att snabbt kunna **kolla bildens äkthet**.
- Koppla till tidigare lektioner om AI-generering: om man **vet** att även bilder kan vara AI-skapade, blir det ännu viktigare att kunna kontrollera ursprunget.

Diskussionsfrågor

1. **Varför litar vi ofta mer på vissa källor än andra?** Vad kan det bero på (historik, professionalism, faktakoll, ryktbarhet)?
2. **Hur pålitlig är en omvänd bildsökning?** Kan den också lura oss om bilden inte finns indexerad någonstans?
3. **Har ni något personligt exempel** där ni sett en bild på nätet som visade sig vara ur sitt sammanhang eller fejk?
4. **Vilka konsekvenser** kan det få om en fejkad bild sprids och alla tror den är äkta?

Didaktiska tips

- **Teknisk vägledning:** Hjälp eleverna som är ovana att högerklicka och söka med Google Lens eller liknande. På mobiler kan man behöva hålla inne fingret på bilden för att få upp alternativet.
- **Integritet och hänsyn:** Påminn eleverna att vara försiktiga med personliga bilder. Vi övar i klassen på allmänna bilder för att lära oss metoden.
- **Koppla till mediekunskap:** Ni kan kombinera denna övning med diskussion om hur journalister arbetar för att verifiera bilders ursprung.
- **Anpassa exempel:** Välj bilder som är relevanta för elevernas intressen eller som nyligen varit virala i medier, för att öka engagemanget.

Fördjupning

- **Källkritiska kriterier:** Låt eleverna analysera omvänt bildsökta resultat med klassiska frågor: Vem? Vad? Varför? När? Hur?
- **Praktisk uppgift:** Ge eleverna en påstådd "nyhetsbild" och be dem med omvänd bildsökning kartlägga var den kommer ifrån, vilka sajter den spridits på, och om originalversionen är samma eller skiljer sig.

Sammanfattning

Genom att **öva på omvänd bildsökning** lär sig eleverna ett av de mest grundläggande verktygen för att motverka bildbaserad desinformation. Lektionen lyfter också **källtillit** – hur man bedömer olika källors trovärdighet och varför det är så viktigt att **inte** ta bilder för givna, särskilt i en tid när bilder enkelt kan redigeras eller AI-genereras. Målet är att eleverna ska känna sig mer rustade att ifrågasätta och granska visuella påståenden på nätet.

Arbetsblad – Lektion 11: Omvänd bildsökning

Hur och varför vi kontrollerar en bilds ursprung för bättre källtillit

1. Syfte

I den här övningen lär du dig att använda **omvänd bildsökning** för att undersöka var en bild kommer ifrån och om den använts i olika sammanhang. Det hjälper oss att:

- ✓ Avslöja om en bild är tagen ur sitt sammanhang.
- ✓ Se om en bild har blivit redigerad eller manipulerad.
- ✓ Förstå varför vissa källor är mer trovärdiga än andra.

2. Steg 1 – Testa omvänd bildsökning

Så här gör du på dator:

1. **Högerklicka** på en bild på en webbsida.
2. Välj "**Sök bild med Google**" eller "**Sök med Google Lens**".
3. Google visar var samma eller liknande bilder finns på nätet.

Så här gör du på mobil:

1. Håll inne fingret på bilden.
2. Välj "**Sök på Google efter denna bild**" eller "**Sök med Lens**".
3. Granska resultaten.

✦ **Alternativ metod:** Gå till images.google.com, klicka på kameran  och ladda upp en bild!

3. Steg 2 – Analysera en bild


Instruktion: I grupp eller par, välj en bild som verkar misstänkt. Det kan vara en bild du hittar online eller en som läraren delar ut.

Fyll i tabellen nedan efter din sökning:

Fråga	Svar
Vad föreställer bilden?	...
Var hittade du den först?	...
Dök bilden upp på andra webbplatser?	...
Har bilden använts i olika sammanhang?	...
Är bilden redigerad eller manipulerad? Hur vet du det?	...
Vilken källa verkar vara ursprunget till bilden?	...

Hur trovärdig verkar denna källa?	...
-----------------------------------	-----

4. Steg 3 – Diskutera källtillit


 **Diskutera i grupp** och besvara följande frågor:

1. **Hittade ni flera versioner av bilden? Vad var skillnaden?**
 - ...
2. **Hur kan vi avgöra om en källa är mer pålitlig än en annan?**
 - ...
3. **Har ni någon gång sett en bild på sociala medier och senare insett att den var tagen ur sitt sammanhang? Beskriv kort.**
 - ...
4. **Vilka konsekvenser kan det få om en manipulerad bild sprids och folk tror att den är äkta?**
 - ...

5. Steg 4 – Reflektion & lärdomar

 **Svara individuellt:**

1. **Hur sanningsenlig är en bild egentligen?**
 - ...
2. **Kommer du att använda omvänd bildsökning i framtiden? När skulle det vara extra viktigt?**
 - ...
3. **Hur kan du hjälpa andra att bli mer källkritiska till bilder de ser online?**
 - ...

 **Kom ihåg!** Bara för att en bild ser äkta ut betyder det inte att den är det. Använd omvänd bildsökning för att kontrollera fakta innan du delar något vidare!

Lektion 12: InVid WeVerify

(Ett allt-i-ett-verktyg för källgranskning av videor och bilder)

Introduktion

I denna lektion får eleverna bekanta sig med **InVid WeVerify** – ett digitalt verktyg som hjälper användaren att granska videor och bilder genom bland annat omvänd bildsökning, metadata-analys och fragmentering av videoklipp. Det är skapat för journalister och faktagranskare, men även elever kan ha stor nytta av det för **källkritik**, särskilt när de stöter på videomaterial vars äkthet de vill kontrollera.

Syfte och mål

- **Lära sig använda** InVid WeVerify för att utföra omvänd bildsökning, dela upp videor i bildrutor, och analysera innehåll mer ingående.
- **Förstå** hur verktyget kan användas för retorisk analys och bedömning av tillförlitlighet.
- **Tillämpa** källkritiska färdigheter på olika typer av nyhetsinlägg – gärna från både “höger” och “vänster” medier, för en mer varierad bild.

Lektionsingång

1. Diskussion:

- Fråga eleverna: “Hur många av er har sett en kort videosekvens på nätet och undrat om den verkligen är sann eller tagen ur sitt sammanhang?”
- Förklara att InVid WeVerify kan användas för att *klippa upp* en video i bildrutor, hitta ursprungliga källor och göra en **mer noggrann** analys.

2. Kort presentation av verktyget:

- Visa hemsidan (sök på “InVid WeVerify plugin” eller besök <https://www.invid-project.eu>) och förklara att det finns som webbläsartillägg (Chrome/Firefox) eller via en webbsida.

Genomförande (steg-för-steg)

1. Introduktion till InVid WeVerify-funktionerna (5–10 min)

- **Video fragmentation:** Man kan ladda upp eller länka en video och få den uppdelad i bildrutor för att sedan göra omvänd bildsökning på enskilda rutor.

- **Keyframe extraction:** Verktöget försöker hitta “nyckelbilder” i en video som bäst representerar innehållet.
- **Omvänd bildsökning:** Direkt i verktöget kan man göra sökningar på Google, Yandex, Bing m.fl. för att hitta var en bild använts tidigare.
- **Metadata-analys:** Vissa filers metadata kan ge ledtrådar om när och var bilden togs.

2. Eleverna hittar inlägg från olika nyhetsmedier (10–15 min)

- Dela in klassen i grupper (2–3 elever). Varje grupp väljer **ett inlägg** från en “vänsterorienterad” källa och **ett** från en “högerorienterad” källa (eller bara två olika tidningar/källor med olika profil).
- Tips: Låt dem leta upp en kort video eller bild som känns intressant att granska. (T.ex. en viral klippsnutt, en politiskt laddad bild, eller en sensationalistisk rubrik.)

3. Analys med InVid WeVerify (10–15 min)

- **Steg 1:** Eleverna använder **omvänd bildsökning** på stillbilder ur klippet för att se om den förekommer i andra sammanhang.
- **Steg 2:** Kolla **metadata** (om möjligt) och se om det finns några avvikelser (t.ex. felaktiga datum).
- **Steg 3 (valfritt):** Om man hittar ursprungswebbplatsen eller källan, gör en **retorisk analys**:
 - Vilka ord eller uttryck används för att förstärka ett budskap?
 - Verkar avsändaren ha en tydlig agenda?
 - Är videon klippt så att något väsentligt kan ha utelämnats?

4. Gruppdiskussion och sammanställning (5–10 min)

- Varje grupp sammanfattar vad de hittade:
 - Kunde de se om bilden/videon var äkta eller manipulerad?
 - Hur skiljde sig “vänsterkällans” respektive “högerkällans” rapportering?
 - Fanns det retoriska knep eller partiska vinklingar?

5. Gemensam reflektion (5 min)

- Diskutera hur ett verktyg som InVid WeVerify kan höja vår källkritiska nivå när det gäller videoinnehåll, som ofta är svårare att analysera än text.
- Påminn om att även de mest avancerade verktyg har begränsningar – men att detta ändå är ett viktigt stöd för faktagranskare.

Diskussionsfrågor

1. **Hur skiljer sig** en retorisk analys av ett videoinslag från en analys av en text?
2. **Har ni märkt** att vissa klipp eller bilder används av både “höger” och “vänster” medier men med olika vinkling? Hur framkom det i er undersökning?
3. **Varför är det viktigt** att göra både teknisk analys (metadata, bildsök) och innehållsanalys (retorik, agenda)?
4. **Tror ni att** ett verktyg som InVid WeVerify kommer bli vanligare även för “vanliga” internetanvändare, inte bara journalister? Varför eller varför inte?

Didaktiska tips

- **Förbered källor** i förväg: Om eleverna snabbt behöver hitta “höger” vs. “vänster” artiklar, ge exempel på några nyhetssajter eller inlägg.
- **Tid och tillgänglighet:** Se till att InVid WeVerify är installerat på skolans datorer eller visa hur man använder en webbaserad version om ni inte kan installera tillägg.
- **Variera innehållet:** Om ni vill, låt några grupper titta på sportklipp, andra på politiska klipp, andra på kändisskvaller, för att visa verktygets bredd.
- **Stöd och struktur:** Vissa elever kan behöva ett arbetsblad med stegvis instruktion (t.ex. “1) Video fragmentation, 2) Keyframe extraction, 3) Reverse image search, ...”) för att hålla reda på processen.

Fördjupning

- **Följ länkarna vidare:** Be grupperna följa de sökträffar de hittar och se om man kan spåra klippet/bilden till en ursprungskälla. Diskutera vilken roll aggregatorer (sajter som samlar nyheter) spelar i att sprida material utan full kontext.
- **Jämför** InVid WeVerify med andra källkritiska verktyg (t.ex. Google Fact Check Tools). Vilka fördelar/nackdelar finns?

- **Lägg till** en emotionell/retorisk komponent: Låt eleverna räkna hur många “laddade ord” eller “värdeomdömen” som förekommer i en artikel eller i texten som följer en video. Koppla till begreppet “tendens”.
-

Sammanfattning

I denna lektion utforskar eleverna **InVid WeVerify** och lär sig hur det kan användas för att **nysta upp** varifrån en bild eller video kommer, samt om/var den förekommit tidigare och i vilket sammanhang. Genom att låta dem analysera material från olika politiska eller ideologiska vinklar, får eleverna både **teknisk** och **innehållsmässig** källkritik i praktiken. Syftet är att de ska inse att verktyget kan hjälpa dem att ställa bättre frågor och fatta mer **informerade** bedömningar av innehåll online.

Lektion 13: Källkritiksassistenten

(Använd AI för att analysera retorik och upptäcka manipulationstekniker)

Introduktion

Denna lektion visar hur man kan använda en AI-baserad chattbot (t.ex. ChatGPT eller liknande) som en **“källkritiksassistent”** för att analysera och synliggöra retoriska grepp och manipulationstekniker i texter eller inlägg. Eleverna får pröva att kopiera in ett inlägg – t.ex. från en känd politiker eller influencer – i en chattbot och ställa frågor som “Vilka manipulationstekniker och retoriska grepp används här?” Lektionen kopplar ihop **människans** källkritiska förmåga med **AI-verktygets** textanalys och ska leda till en mer nyanserad bild av hur vissa budskap kan påverka oss.

Syfte och mål

- **Visa hur AI** kan användas som ett stöd i retorisk analys och källkritik.
- **Uppmärksamma** olika manipulationstekniker: t.ex. personangrepp, guilt by association, överdrift, känslobaserad retorik m.m.
- **Utmana eleverna** att jämföra sina egna iakttagelser med AI:ns, för att bredda sin förståelse av textens budskap och syfte.

Lektionsingång

1. Diskussion:

- Ställ frågan: “Har ni sett inlägg på nätet där det känns som att avsändaren ‘vet exakt vad man ska säga’ för att övertyga?”
- Förklara att AI kan användas både för att **skapa** övertygande texter men också för att **analysera** dem.

2. Exempel på inlägg:

- Välj ett tydligt exempel, t.ex. en tweet eller Facebookpost från en känd person (Donald Trump, Elon Musk eller någon annan kontroversiell figur). Visa det för klassen (text eller skärmdump).

Genomförande (steg-för-steg)

1. Förbered en prompt till chattbotten (5 min)

- Exempel på prompt: *“Analysera följande text. Identifiera och beskriv vilka manipulationstekniker och retoriska grepp som används, och förklara hur de kan påverka läsaren.”*

2. Testa med ett exempel i helklass (5–10 min)

- Kopiera in texten från det valda inlägget i chattbotten tillsammans med prompten.
- Läs upp eller visa AI:ns svar: Vilka tekniker pekar AI:n ut? Är ni överens i klassen?

3. Eleverna väljer egna flöden eller inlägg (10–15 min)

- Dela in eleverna i smågrupper. Varje grupp väljer ett **aktuellt inlägg** från sin egen flödeskoll (t.ex. Twitter, Instagram, YouTube-kommentarer) eller från en nyhetssida.
- De kopierar texten (eller en sammanfattning om texten är för lång) och använder en AI-chattbot för att be om en retorisk analys.
- **OBS:** Varna för att inte klistra in **personlig** eller **känslig** info. Fokus är på offentliga inlägg.

4. Jämför AI:ns analys med gruppens egna iakttagelser (5–10 min)

- Diskutera: “Vilka metoder såg AI:n som vi inte tänkte på?” “Håller vi med AI:ns slutsats?”
- Be eleverna skriva ner **minst 2 saker** de tycker AI:n analyserade bra, och **1 sak** de ifrågasätter.

5. Sammanfattning och reflektion (5 min)

- Fördelar: AI kan snabbt hitta ordval, tonfall och mönster i text som vi kanske missar.
- Nackdelar: AI kan ha fel eller missa kontext. Elevernas **egna** förkunskaper och sunt förnuft är fortfarande viktiga.
- Understryk vikten av att AI är ett **verktyg**, inte en absolut auktoritet.

Diskussionsfrågor

1. **Varför kan AI vara bra** för att upptäcka retoriska grepp? Vad saknar AI som en människa kanske ser?

2. **Kan AI chattbotar själva** använda manipulativa tekniker när de svarar på frågor?
Hur vet vi?
3. **Hur påverkas vår tillit** till en text om vi ser att den använder starka retoriska knep, personangrepp eller känslomässig övertalning?
4. **Finns det risk** att vi litar för mycket på AI-analysen och glömmer bort vår egen kritiska läsning?

Didaktiska tips

- **Konkreta retoriska grepp:** Förklara eller påminn eleverna om vanliga termer (t.ex. ad hominem, rädslargument, halmgubbe, polarisering, etc.) så att de förstår AI:s terminologi.
- **Exemplifiera gärna** med både “högersidans” och “vänstersidans” argumentation, eller med inlägg från olika branscher (politik, marknadsföring, influencers).
- **Begränsa textlängden:** För att AI snabbt ska kunna ge ett överskådligt svar, kan ni välja kortare inlägg eller utdrag.
- **Tekniskt försprång:** Om ni har tid, låt eleverna pröva olika promptar (korta vs. mer detaljerade instruktioner) och se hur AI:s analys förändras.

Fördjupning

- **Kvalitetsbedömning av AI-svar:** Låt eleverna gemensamt diskutera “Hur korrekta tycker vi AI:ns analyser är?” och “Hur kan vi dubbelkolla med fakta eller andra källor?”
- **Skapa en checklista:** Be eleverna (som en gruppuppgift) att lista 5–10 punkter att alltid kolla när man ser AI:s retoriska analys (t.ex. “är texten medvetet vinklad?”, “kan AI ha glömt kontext?”, “finns satir?”). **Ett förslag på checklista finns nedan**
- **Historiska tal eller texter:** Vill ni gå djupare, låt eleverna köra historiska politiska tal genom AI för att se vilka retoriska knep som varit vanliga genom tiderna.

Sammanfattning

Genom att be en AI-chattbot om retorisk analys eller identifiering av manipulationstekniker i ett inlägg, får eleverna **stöd** i att uppmärksamma språkliga och psykologiska knep som kan påverka deras åsikter. Samtidigt tränar de sin **egna** källkritiska blick genom att jämföra AI:ns analys med sin egen uppfattning. Lektionen belyser både **fördelar** (snabb och systematisk

analys) och **begränsningar** (ingen mänsklig kontextkänsla) med att använda AI som “källkritiksassistent.”

Checklista: Analysera AI:s retoriska analys

När AI har gjort en retorisk analys av en text, gå igenom dessa punkter för att säkerställa att bedömningen är trovärdig och nyanserad.

✓ 1. Är texten medvetet vinklad?

- Verkar AI peka ut vissa åsikter som “goda” eller “dåliga”?
- Saknas en neutral bedömning av innehållet?

✓ 2. Kan AI ha glömt kontext?

- Analyserar AI bara texten som den är, utan att förstå bakgrunden?
- Finns det en händelse, debatt eller tidigare uttalanden som AI inte nämner?

✓ 3. Finns satir eller ironi?

- Är texten en skämtpost eller ett satiriskt inlägg?
- Har AI förstått detta, eller behandlar den det som seriös fakta?

✓ 4. Påverkar AI:s svar ditt förtroende för texten?

- Skulle du ha tolkat texten på samma sätt utan AI:s analys?
- Ger AI en rättvis bedömning, eller blir du mer misstänksam bara för att den säger att texten använder “manipulativa tekniker”?

✓ 5. Har AI hittat manipulationstekniker, men missat innehållets sanning?

- Även om en text använder känslomässiga ordval eller retoriska knep, kan den fortfarande vara sakligt korrekt.
- Fokuserar AI mer på hur något sägs än på om det som sägs faktiskt är sant?

✓ 6. Vilka retoriska grepp nämns?

- Använder AI tydliga begrepp som “överdrift”, “rädslometod” eller “guilt by association”?
- Finns det exempel i texten som stödjer detta?

✓ **7. Finns det några partiska uttryck i AI:s analys?**

- Använder AI värdeladdade ord i sin egen analys?
- Beskriver den vissa argument som “farliga”, “vilsledande” eller “extrema” utan att förklara varför?

✓ **8. Skulle en människa analysera texten annorlunda?**

- Jämför AI:s analys med din egen och klasskamraternas.
- Är det något AI har missat eller överdrivit?

✓ **9. Finns det motstridiga tolkningar?**

- Om du skulle ge texten till två olika personer, skulle de tolka den på samma sätt?
- Påverkas AI:s analys av vilken chattbot du använder?

✓ **10. Behöver du faktakolla något själv?**

- AI analyserar språk, men kan den verifiera fakta?
- Om AI nämner en specifik händelse eller person, dubbelkolla via pålitliga källor.

Instruktion för eleverna:

- Använd checklistan **efter att ni fått AI:s analys** av ett inlägg.
- Svara på frågorna i par eller smågrupper.
- Diskutera vad ni håller med om och vad ni tycker AI missat.
- Reflektera: **Skulle du ha tolkat texten annorlunda om AI inte fanns?**

Denna checklista hjälper eleverna att vara **källkritiska även mot AI**, och att förstå att en AI-analys **inte alltid är en slutgiltig sanning** utan ett verktyg att använda tillsammans med egen bedömning.

Arbetsblad – Lektion 9: Källkritiksassistenten

Hur kan AI hjälpa oss att analysera retorik och upptäcka manipulationstekniker?

1. Syfte

I denna lektion ska du och dina klasskamrater **använda en AI-chattbot** (ex. ChatGPT) för att avslöja **manipulationsknep och retoriska grepp** i olika texter eller inlägg på nätet. Målet är att visa hur AI kan vara en **“källkritiksassistent”**, men också att AI har sina brister och saknar mänsklig kontextkänsla.

2. Förberedelser

1. Välj en text eller inlägg

- Det kan vara ett **kort** inlägg från en känd politiker, influencer eller en offentlig debatt.
- Undvik att klistra in **privat** eller **känslig** information. Vi fokuserar på **offentliga** poster.

2. Formulera en prompt till AI:n

- Exempel:

“Analysera följande text. Beskriv vilka manipulativa och retoriska tekniker som används, och förklara hur de kan påverka läsaren.”

3. Kopiera texten in i AI-chattboten, invänta svaret.

3. Steg-för-steg-analys

Instruktion: Följ stegen nedan, antingen individuellt eller i smågrupper.

Steg 1: AI:s analys

1. **Klistra in** din valda text i chattboten + prompten.
2. **Läs AI:ns svar** noggrant:
 - Vilka retoriska grepp pekar AI:n ut?
 - Tycker du AI:n förklarar dem tydligt?

Steg 2: Din egen bedömning

1. **Diskutera i grupp:** “Håller vi med AI:ns slutsatser?”
2. **Notera** minst två punkter AI tog upp som ni inte hade tänkt på själva.
3. **Fundera** på om AI verkar ha missat något tydligt.

Steg 3: Jämför

1. **Gör en enkel lista** över knep AI säger finns i texten vs. knep ni själva observerat.
2. Vilka “+” (fördelar) ser ni med AI:s metod? Vilka “-” (nackdelar) ser ni?

4. Diskussionsfrågor

1. **Varför** kan AI vara bra på att upptäcka retoriska mönster?
2. **Vad** saknar AI som en människa kanske ser? (t.ex. kontext, tonfall, “mellan raderna”-budskap)
3. **Kan AI själv** vara manipulerande i sina svar? Hur vet vi det?
4. **Vilken risk** finns det med att lita för mycket på AI-analysen?

5. Checklista: Analysera AI:s retoriska analys

Instruktion: Efter att du fått AI:s bedömning, gå igenom checklistan nedan för att avgöra hur trovärdig eller fullständig du tycker analysen är.

Punkt	Dina noteringar
1. Är texten medvetet vinklad?	...
2. Kan AI ha glömt kontext?	...
3. Finns satir eller ironi?	...
4. Påverkar AI:s svar ditt förtroende för texten?	...
5. Har AI hittat retoriska knep, men missat innehållets saklighet?
6. Vilka retoriska grepp nämns?	...
7. Finns det partiska uttryck i AI:s analys?	...
8. Skulle en människa analysera texten annorlunda?	...
9. Finns det motstridiga tolkningar?	...
10. Behöver du faktakolla något själv?	...

Tips: Om du finner brister i AI:s analys, pröva att **omformulera** prompten och se om AI ger andra svar.

6. Reflektion

Instruktion: Fundera enskilt eller samtala i grupp.

1. **Hur förändras** din syn på en text när AI påpekar manipulationsknep som du själv missade?
2. **Vilken vikt** lägger du vid AI:s analys jämfört med din egen?
3. **När** är det särskilt viktigt att undersöka en texts retorik (t.ex. politiska inlägg, reklam, etc.)?

7. Sammanfattning

- AI kan vara ett **stöd** för att avslöja retoriska knep och manipulation, men **du** som människa måste tänka kritiskt och bedöma sammanhang och fakta.
 - Retoriska knep är inte alltid fel – ibland kan de förstärka ett ärligt budskap. Men när de används för att vilseleda eller förvränga verkligheten, blir de **manipulativa**.
 - **Balansen:** Lita varken blint på AI eller en textavsändare, men **kombinera** AI:s metodiska genomgång med din egen förmåga att tolka, värdera och faktagranska.
-

Lycka till med att utforska hur AI kan bredda ditt källkritiska tänkande, utan att ersätta det!

Lektion 14: Lateral läsning

(Hur man kontrollerar trovärdighet genom att läsa “utåt” istället för “inåt”)

Introduktion

Lateral läsning innebär att man, i stället för att bara analysera en källa “vertikalt” (läsa den från början till slut och försöka bedöma sanningshalten utifrån dess innehåll), även “läser utåt” genom att aktivt söka information från andra källor parallellt. Metoden är vanlig bland professionella faktagranskare och journalister: man jämför vad flera oberoende källor säger om samma ämne för att bedöma om något verkar stämma.

I denna lektion förklarar vi hur **AI-verktyg** som **Perplexity.ai**, **Grok Deep Search** eller **ChatGPT** (med möjlighet att söka på nätet) kan vara ett stöd i lateral läsning. Samtidigt är det viktigt att betona att AI-modeller ibland gör fel (s.k. “hallucinationer”) och att eleverna alltid behöver vara källkritiskt vaksamma.

Syfte och mål

- **Förstå vad lateral läsning är** och varför det skiljer sig från traditionell, “vertikal” granskning av en text.
- **Lära sig hur AI-verktyg** kan användas för att snabbt hitta kompletterande eller motsägande uppgifter om ett påstående.
- **Utveckla en rutin** för att alltid dubbelkolla viktiga fakta eller påståenden i flera källor, även med AI:s hjälp.

Lektionsingång

1. Diskussion:

- Fråga: “Hur brukar ni kolla om något är sant när ni ser det online? Blir det bara ‘kolla källan’ eller söker ni runt på andra sidor?”
- Presentera idén med “lateral läsning”: i stället för att förlita sig på textens egen framställning, letar man samtidigt (lateral = åt sidan) efter bekräftelse eller motsägelser i **andra** källor.

Genomförande (steg-för-steg)

1. Förklaring av lateral läsning (5–10 min)

- Visa ett kort exempel: ett inlägg om en “ny mirakelkur” eller en “sensationell nyhet”.
- Istället för att bara läsa inlägget noggrant, öppnar ni flera **flikar** i webbläsaren och söker på ämnet på andra sajter.
- Poängtera att professionella faktagranskare sällan “djupdyker” först i källan, utan startar med att se vad andra källor säger om saken.

2. Demonstrera AI-verktyg (10–15 min)

- **Perplexity.ai**: Visa hur ni kan mata in ett påstående och få en sammanställning från olika webbplatser. Diskutera: “Får vi relevanta länkar? Hur tydliga är källhänvisningarna?”

- **Grok Deep Search:** Om tillgängligt, visa hur plattformen kan gräva fram olika perspektiv på samma ämne.
 - **ChatGPT** (med webbläsar-plugin eller integrerad sök): Förklara att en viss version (t.ex. ChatGPT med Bing-integration eller ChatGPT “utan inlogg” om den har webbsök) kan sammanfatta olika källor – men att AI ibland kan hitta på referenser.
3. **Elevövning: Lateral läsning med AI (15–20 min)**
1. Dela in eleverna i grupper. Varje grupp får välja **ett påstående** eller **en nyhet** som de vill granska (gärna något halvkontroversiellt eller intressant).
 2. Grupperna använder **minst två** av följande metoder för att göra en lateral läsning:
 - **Söka manuellt** i Google/Bing efter andra källor.
 - **Fråga Perplexity.ai** eller liknande verktyg om att sammanfatta olika källor kring ämnet.
 - **Fråga ChatGPT** och se vad den säger och vilka länkar den ger.
3. Var uppmärksam på: Vilka **källor** nämner AI? Är de trovärdiga? Finns det direkta länkar eller “hallucinerar” AI och hittar på?
4. **Reflektion i grupp (5–10 min)**
- Eleverna sammanställer **kort** hur många källor de hittade, om källorna var överens eller motstridiga, och hur väl AI verkade ge korrekta fakta.
 - Diskutera: “Var det tydligt när AI hade fel, eller var det svårt att avgöra?”
5. **Sammanfattning i helklass (5 min)**
- Understryk att AI kan vara ett **snabbt verktyg** för att få fram en översikt av olika perspektiv (dvs. “lateral läsning i expressfart”), men att AI:s svar alltid måste kollas.
 - Påminn om att det är en **vana**: innan man tror på en påstående, är det bra att “läsa utåt” och se vad andra säger.

Förslag på lektionsupplägg/övningar

- **Mini-debunk:** Låt eleverna välja en viral nyhet/meme och “lateral-läsa” den med AI-hjälp för att se om det finns stöd i flera källor.
- **Tidsbegränsad sökning:** Ge grupperna 5 minuter att snabbt hitta “3 källor som bekräftar eller motsäger påståendet”. Diskutera resultaten.
- **AI vs. Människa:** En grupp använder AI-verktyg, en annan söker manuellt. Jämför vem som hittar mest relevant info på kort tid. Diskutera för- och nackdelar.

Diskussionsfrågor

1. **Varför är lateral läsning** mer effektiv än att bara lita på textens egna källhänvisningar?
2. **Vad gör ni** om AI ger er länkar eller referenser som inte verkar existera? Hur dubbelkollar ni?
3. **Har ni upplevt** att AI blandar ihop fakta i sina svar? Varför kan det hända?
4. **Hur kan** lateral läsning bli en “vana” i ert dagliga surfande? Vilka situationer i livet kan detta vara extra användbart?

Didaktiska tips

- **Börja enkelt:** Visa gärna ett mycket tydligt falskt påstående (“Kändis X är från planeten Mars!”) och göra lateral läsning som en lekfull demonstration.
- **Betona källkritik:** Visa hur man kan bedöma källorna AI rekommenderar (t.ex. Wikipedia vs. en liten blogg).
- **Var uppmärksam på AI:s brister:** Påminn eleverna att AI kan vara överdrivet självsäkert eller hitta på källor (fejkreferenser).
- **Koppla till tidigare lektioner:** Om ni använt omvänd bildsökning eller retorisk analys tidigare, poängtera att lateral läsning ofta kombineras med flera källkritiska metoder.

Fördjupning

- **Bygg en “källbank”:** Eleverna kan skapa en lista med trovärdiga sajter (faktagranskare, seriösa nyhetsmedier, fackbloggar) som de snabbt kan kolla när AI ger dem referenser.
- **Jämför AI-verktyg:** Låt en grupp använda Perplexity.ai, en annan ChatGPT med webbsök, en tredje “Grok Deep Search”. Jämför hur olika svaren blir på samma fråga.
- **Prova utan internet:** Diskutera hur lateral läsning gick till förr (journalister som ringde upp experter, besökte bibliotek). Ger AI & nätet oss helt nya möjligheter eller nya risker?

Sammanfattning

Lateral läsning är en **metod** för att dubbelkolla påståenden genom att titta “åt sidan” – jämföra vad andra källor säger om samma sak. AI-verktyg som Perplexity.ai, Grok Deep Search eller ChatGPT med webbsök kan **snabba upp** denna process och ge länkar till fler källor på kort tid. Samtidigt måste eleverna **fortsätta vara källkritiska**, då AI ibland felciterar eller hallucinerar information. Målet är att eleverna ska träna upp **en vana**: se AI som ett **stöd** i att läsa “lateralt” och inte som en absolut sanning.

Arbetsblad – Lektion 10: Lateral läsning

Hur kontrollerar vi trovärdighet genom att “läsa utåt” och inte bara lita på en enda källa?

1. Vad är Lateral läsning?

Instruktion: Läs följande stycke och fundera på varför “lateral läsning” kan ge en mer pålitlig bild av vad som faktiskt är sant.

Lateral läsning innebär att du parallellt letar upp **andra källor** om samma ämne för att se om de bekräftar eller motsäger det du läser. I stället för att bara lita på en texts *egen* framställning eller källhänvisningar (vertikal läsning), öppnar du nya flikar och jämför vad **oberoende** eller **andra** källor säger om samma sak.

Fråga:

Hur brukar du kolla om något är sant när du ser en post eller nyhetsartikel online? Skriv **minst två** sätt som du gör i dag, och **minst ett** sätt du kan förbättra (t.ex. “jag läser även andra sidor, eller jag kollar faktagranskningssajter”).

- Mitt sätt idag:
 1. ...
 2. ...
- Hur jag kan förbättra:
 - ...

2. Använda AI som hjälp i lateral läsning

Instruktion: Läraren visar hur man kan använda ett AI-verktyg (t.ex. Perplexity.ai eller ChatGPT med webbsök) för att snabbt hitta **fler källor**. Ni ska själva pröva detta i smågrupper.

1. **Exempel:**
 - Om ni har ett påstående: “Forskare har bevisat att [X] botar huvudvärk på 2 minuter.”
 - Fråga AI: “Sammanfatta vad flera **olika källor** säger om detta påstående.”
2. **Var vaksam:**
 - Ger AI “spök-länkar” som inte finns?
 - Är källorna trovärdiga eller är det bara bloggar och anonyma forum?

3. Praktisk övning: Lateral läsning i grupp

Instruktion: Arbeta 2–3 per grupp. Välj **ett påstående eller nyhet** ni är nyfikna på (det kan vara lite kontroversiellt eller på gränsen).

1. Manuell sökning:

- Sök i Google/Bing. Hitta minst 2–3 källor.
- Är dessa källor eniga eller motsägande?

2. AI-sökning:

- Använd t.ex. Perplexity.ai eller ChatGPT (med webbsök).
- Be om en sammanfattning av olika källors påståenden i ämnet.
- Kontrollera de källor AI nämner – existerar de? Är de pålitliga?

3. Jämför & notera:

- Hur många **unika** källor hittade ni totalt?
- Vilka **skillnader** ser ni mellan att göra manuell sökning vs. AI?

4. Sammanställning

Skapa en **kort lista** över de källor ni fann. Skriv ett “+” om ni bedömer den som trovärdig, och “?” om ni är osäkra.

Källa	Trovärdig?	Motivering
Exempel: “healthnews4u.com”	?	Okänd blogg, ingen “om oss”-sida, ej expertverifierad
Exempel: “Läkemedelsverket.se”	+	Officiell myndighet, tydlig info
...		

5. Reflektion i grupp

1. **Vad lärde ni er** om påståendet ni undersökte? Var det sant, överdrivet eller falskt?
2. **Hur pålitliga** var AI:s förslag? Blev ni lurade av “hallucinerade” länkar?
3. **Varför** kan lateral läsning bli en “vana” i er vardag? Nämn **ett** tillfälle utanför skolan då detta är extra viktigt.

6. Diskussionsfrågor

1. **Varför** är det bättre att läsa “utåt” (flera källor) än att bara tro på textens egna källhänvisningar?
2. **Vad gör du** om AI hittar på referenser som inte finns?
3. **Har AI ibland blandat ihop fakta** i sina svar – varför händer det?
4. **Hur** kan du bli snabbare på att göra lateral läsning när du ser en konstig nyhet i flödet?

7. Sammanfattning: Dina “lateral-läsningssvanor”

Instruktion: Skriv en **kort plan** (2–3 meningar) för hur du i fortsättningen tänker tillämpa lateral läsning när du stöter på en nyhet eller ett påstående online.

- **Min plan:**
 - ...

Tips: Om en rubrik är chockerande eller osannolik, öppna **flera flikar**, sök manuellt och kolla vad AI eller andra källor säger. Kom ihåg att AI bara är ett **hjälpmedel** – du måste göra den slutgiltiga bedömningen!

Bra jobbat! Genom att läsa **“utåt”** och jämföra källor, undviker du att förlita dig på en enda text som kan vara vinklad eller felaktig.

Nu är du ett steg närmare att bli en fullfjädrad “faktagranskare” i din vardag.

Lektion 15: Kritiskt granska innehåll med AI som verktyg

Syfte

I denna lektion ska eleverna utforska hur generativ AI kan användas som ett *stöd* för kritisk granskning av texter och information. Syftet är att eleverna lär sig använda AI för att analysera och granska innehåll ur flera perspektiv – utan att för den skull tumma på sitt eget kritiska tänkande. De ska erfara hur AI kan hjälpa till att sammanfatta komplex information, hitta argument och upptäcka eventuella svagheter i ett material, samtidigt som de blir medvetna om AI-verktygens begränsningar (t.ex. risk för fel eller vinklingar). Kort sagt: AI blir här en *kritisk läskompanjon* snarare än en facitmaskin.

Lektionsingång

- **Presentera en utmanande text:** Starta lektionen med att ge eleverna ett stycke ur en komplex text – kanske en debattartikel, nyhetsanalys eller vetenskaplig populärartikel – som innehåller en tydlig tes eller argumentation. Be en frivillig elev läsa upp ett stycke högt. Fråga sedan klassen: *“Vad var huvudpoängen här, och hur välstödd verkar den?”*. Det är ofta svårt att snabbt sätta fingret på det, vilket öppnar för att introducera AI som ett verktyg: *“Låt oss se hur en AI kan hjälpa oss att nysta i detta.”*
- **Förväntningar vs verklighet:** Fråga eleverna hur de *tror* att en AI skulle kunna hjälpa till att granska texten. Samla förslag på tavlan (t.ex. “kolla fakta”, “hitta huvudidéer”, “ge motargument”). Anteckna även eventuella farhågor de nämner (“kanske hittar på saker?”). Denna brainstorming ger en bild av elevernas förhandsuppfattningar och gör dem medvetna om både potential och risker med AI i källkritik.

Instruktioner (steg-för-steg)

1. **Introducera AI som analysverktyg:** Förklara att eleverna nu ska arbeta med en AI-tjänst (t.ex. ChatGPT) för att kritiskt granska en text. Poängtera att AI här inte är *svaret* utan ett *verktyg* – ungefär som en stavningskontroll eller miniräknare fast för textanalys.
2. **Individuell läsning:** Dela ut en kopia av den utvalda texten (t.ex. debattartikeln) till varje elev eller låt dem öppna den digitalt. Ge dem några minuter att läsa igenom texten på egen hand och stryka under vad de uppfattar som huvudtesen samt eventuella argument eller källor som används. Detta för att de ska bilda sig en egen uppfattning först.

3. **Använd AI för sammanfattning:** Be eleverna mata in texten eller relevanta utdrag i AI-verktyget och ge en prompt i stil med: *“Summarize the main argument of this text and list any supporting points.”* (Det går bra att göra detta på engelska även om texten är på svenska – alternativt använder ni en svenskspråkig AI). Eleverna noterar AI:ns sammanfattning.
4. **Jämför med egen uppfattning:** Låt eleverna jämföra AI:ns sammanfattning med sin egen förståelse. Stämmer AI:ns identifierade huvudargument med det de själva ansåg? Tog AI:n upp något de missade, eller utelämnade AI:n något viktigt? Denna jämförelse belyser hur AI kan fånga (eller missa) nyanser.
5. **AI för kritiska frågor:** Nästa steg – be eleverna använda AI för att granska textens tillförlitlighet. De kan ställa frågor som: *“Vilka antaganden gör författaren i texten?”*, *“Vilka motargument skulle man kunna ge?”* eller *“Finns det någon bias i resonemanget?”*. Alternativt instruera AI:n: *“Analysera följande text kritiskt och påpeka eventuella svagheter eller fel.”* Eleverna antecknar AI:ns svar.
6. **Faktakoll (ev. extension):** Om texten innehåller specifika påståenden eller statistik, demonstrera hur man **inte** kan lita blint på AI för faktakoll. Be t.ex. AI att verifiera en specifik siffra eller uppgift. Diskutera risken för hallucinationer – AI kan hitta på källor eller svar om den inte har direkt fakta. Detta moment kan kort visas av läraren om tiden är knapp, men det är nyttigt för eleverna att se begränsningen.
7. **Gruppdiskussion – AI:s hjälp och brister:** Dela in klassen i mindre grupper om 3–4. Ge dem 5–10 minuter att diskutera utifrån frågor: *“Hur hjälpte AI er att förstå eller granska texten?”*, *“Vilka insikter fick ni som ni inte hade förut?”* samt *“I vilka avseenden litar ni respektive litar ni inte på AI:ns analys?”*. Be dem komma fram till **minst ett sätt** som AI förbättrade deras analys och **minst ett problem** de måste vara vaksamma på.
8. **Återsamling och reflektion:** Låt grupperna dela sina svar. För varje positiv aspekt som nämns (t.ex. *“AI:n hjälpte oss se dolda antaganden”*), fråga klassen om de håller med och om någon fick annat resultat. För varje risk som nämns (t.ex. *“AI:n höll med om allt i texten”* eller *“AI:n kanske missade sarkasmen i texten”*), diskutera hur man kan komma runt det – t.ex. genom att ställa bättre frågor eller dubbelkolla med andra källor.
9. **Sammanfatta lärdomar:** Avslutningsvis sammanfatta: AI kan vara en användbar assistent för att snabbt dissekera och ifrågasätta innehåll, men eleven **måste** vara den som tar det slutgiltiga avgörandet om trovärdighet. Lyft att forskning faktiskt visar att om man använder generativ AI okritiskt kan det *minska* den kritiska förmågan vilarare.se

– man blir bekväm och frågar inte lika mycket själv. Men om man istället använder AI *medvetet och ställer kritiska motfrågor*, kan det bli ett verktyg som *höjer* den kritiska analysen.

Didaktiska tips

- **Välj lämplig text:** Anpassa svårighetsgraden på texten efter elevernas nivå. För gymnasiet kan det vara en debattartikel i en större dagstidning; för högstadiets äldre elever kanske en lättare insändare eller nyhetsartikel. Det viktiga är att texten har *något* att analysera – t.ex. ett argument, ett kontroversiellt påstående eller tydliga fakta – så att AI:ns insatser blir meningsfulla att utvärdera.
- **Gruppstorlek och stöttning:** Om eleverna är ovana vid den här typen av övning, kan pararbete vara bättre än individuellt initialt – två elever kan hjälpas åt vid datorn. Gå runt och stötta, särskilt i momentet där de ska formulera frågor till AI:n. Visa exempel på hur man kan skriva en prompt för att få ut en kritisk analys om de kör fast.
- **Teknikförberedelser:** Säkerställ att alla har tillgång till ett AI-verktyg. Om inte alla elever har egna enheter kan övningen göras i helklass med läraren som operator vid projektorn, men det är mer engagerande om de får prova själva. Alternativt använd skolans datorsal eller låt dem köra i omgångar.
- **Betona källkritisk attityd:** Påpeka under arbetets gång att AI:n inte är en *kunskapskälla* i sig utan en *bearbetare* av information. Allt AI föreslår måste kontrolleras mot originaltexten eller andra källor. Detta är ett gyllene tillfälle att påminna om källkritikens grundregler: vem är avsändare (här: en AI som saknar egen förståelse), vilka källor stöder påståenden etc. En bibliotekarie eller skolbibliotekarie kan med fördel involveras här för att koppla till informationssökning.
- **Hantering av AI-fel:** Om (eller snarare *när*) AI:n ger ett tveksamt eller felaktigt svar under övningen, se det som ett pedagogiskt guldläge. Lyft direkt: *“Här ser vi att AI:n kanske fabulerade lite – hur kan vi upptäcka det och undvika att gå i den fällan?”*. Eleverna lär sig då att ha ett sunt ifrågasättande även gentemot AI the74million.org

Diskussionsfrågor

- *“Skulle ni lita på en analys som en AI gjort av en text utan att ni själva läst texten? Varför eller varför inte?”*
- *“På vilket sätt gjorde AI:n er mer eller mindre källkritiska när ni granskade artikeln?”*
- *“Tror ni att AI kan upptäcka alla problem i en text? Finns det saker en människa ser som inte AI kan fånga upp?”*

- “Hur kändes det att läsa texten först själv och sedan med AI:ns hjälp? Påverkade AI ert sätt att tänka på innehållet?”
- “Hur skulle ni kunna använda en sådan här metod i framtiden, t.ex. när ni skriver egna uppsatser eller projekt?” (T.ex. som en självkoll: låta AI opponera på ens eget utkast.)

Forskning och arbetsblad

- **Forskning:** Under lektionen kan du nämna att forskning från Microsoft Research Cambridge (2025) indikerar att elever *tror* att de blir mer källkritiska av att använda AI, men att deras faktiska kritiska förmåga riskerar att minska om de förlitar sig blint på AI vilarare.se

. Detta understryker varför just denna typ av övning – där man aktivt jämför och ifrågasätter AI:ns output – är så viktig. Andra experter poängterar att kritiskt tänkande blir ännu mer nödvändigt i AI-eran för att använda tekniken på rätt sätt avidopenaccess.org. Med rätt pedagogik kan AI användas för att *skärpa* snarare än *slöa* elevernas analysförmåga.

- **Arbetsblad:** Förbered gärna ett arbetsblad eller en mall som stöd under övningen. Det kan bestå av sektioner som: **1) Egna anteckningar om textens huvudpoäng, 2) AI:ns sammanfattning (utrymme att skriva av eller klistra in), 3) AI:ns påpekade styrkor/svagheter i texten, 4) Mina reflektioner – håller jag med AI:ns analys? samt 5) Föreslagna motargument eller följdfrågor.** En sådan struktur hjälper eleverna att systematiskt jämföra och reflektera över skillnaden mellan egen analys och AI-analys. Det blir också ett bra underlag för redovisning och efterdiskussion.

Arbetsblad – Lektion 15: Kritiskt granska innehåll med AI som verktyg

Introduktion

I denna övning ska du använda en AI-tjänst (t.ex. ChatGPT) som ett stöd för att analysera och kritiskt granska en text. Tanken är inte att AI ska ge dig alla svar, utan att fungera som en "kritisk läskompanjon" som hjälper dig att se texter ur nya perspektiv. Genom att jämföra din egen analys med AI:ns output kommer du att upptäcka både styrkor och begränsningar med AI – och samtidigt utveckla ditt eget kritiska tänkande.

Instruktioner

- Din lärare ger dig en text att arbeta med (t.ex. en debattartikel eller nyhetsanalys).
- Följ stegen nedan och anteckna dina svar i varje avsnitt.
- Arbeta individuellt med läsningen och AI-analysen, men diskutera sedan i grupp.
- Tänk på: AI kan hjälpa dig att sammanfatta, hitta argument eller svagheter, men det är du som avgör vad som är trovärdigt.

Steg 1: Egen läsning och anteckningar

Läs texten noggrant och stryk under vad du uppfattar som **huvudtesen** (textens viktigaste påstående) samt eventuella **argument eller källor** som stödjer den. Anteckna sedan:

Egen analys	Anteckningar
Huvudtesen	
Argument eller källor	
Frågor eller svagheter du själv ser	

Steg 2: AI-sammanfattning

Mata in texten (eller ett utdrag) i AI-verktyget. Använd denna prompt:

- "Sammanfatta textens huvudsakliga argument och lista eventuella stödjande punkter." (Om du använder en engelskspråkig AI, skriv: "Summarize the main argument of this text and list any supporting points.")
Anteckna AI:ns svar nedan.

AI-sammanfattning Anteckningar

Huvudargument

Stödjande punkter

Steg 3: Jämför med din egen uppfattning

Jämför AI:ns sammanfattning med det du själv skrev i Steg 1. Besvara frågorna:

Jämförelse	Anteckningar
------------	--------------

Stämmer AI:ns huvudargument med ditt?	
Tog AI:n upp något du missade?	
Utelämnade AI:n något viktigt?	
Vad tycker du om AI:ns sammanfattning?	

Steg 4: Kritisk granskning med AI

Be AI:n att analysera textens tillförlitlighet. Använd en eller flera av dessa prompts:

- "Vilka antaganden gör författaren i texten?"
- "Vilka motargument skulle man kunna ge?"
- "Finns det någon bias i resonemanget?"
- "Analysera texten kritiskt och påpeka eventuella svagheter eller fel."
Anteckna AI:ns svar nedan.

Kritisk granskning AI:ns svar

Antaganden

Motargument

Bias eller svagheter

Övriga observationer

Steg 5: Reflektion

Tänk på hur AI:n påverkade din analys och besvara följande frågor:

1. **Hur hjälpte AI dig att förstå eller granska texten?**
[Skriv ditt svar här]
2. **Vilka insikter fick du som du inte hade förut?**
[Skriv ditt svar här]
3. **I vilka avseenden litar du på AI:ns analys? Varför?**
[Skriv ditt svar här]
4. **I vilka avseenden litar du inte på AI:ns analys? Varför?**
[Skriv ditt svar här]

Steg 6: Gruppdiskussion

Diskutera i en grupp om 3–4 personer i 5–10 minuter. Besvara:

- **Ett sätt som AI förbättrade er analys:**
[Skriv här]
- **Ett problem eller en risk med AI som ni måste vara vaksamma på:**
[Skriv här]

Steg 7: Sammanfattning och lärdomar

Efter gruppdiskussionen och klassens återsamling, skriv ner en viktig lärdom du tar med dig från övningen.

Min viktigaste lärdom Anteckningar

Avslutning

Bra jobbat! Genom att arbeta med detta arbetsblad har du lärt dig att använda AI som ett verktyg för att analysera och granska texter – men också att det är ditt eget kritiska tänkande som avgör vad som håller måttet. AI kan vara en hjälpsam assistent, men du är den som bestämmer i slutändan. Spara detta arbetsblad som en påminnelse för framtida analyser!

[Valfritt: Extra tankar eller frågor]

[Skriv eventuella funderingar här]

Not till läraren:

- **Textval:** Välj en text med en tydlig tes och argumentation, t.ex. en debattartikel från en dagstidning. Anpassa svårighetsgraden efter elevernas nivå.
- **Teknik:** Se till att eleverna har tillgång till en AI-tjänst (t.ex. ChatGPT). Om det inte finns tillräckligt med enheter, kan du köra AI:n via projektorn i helklass.
- **Stöd:** Gå runt under övningen och hjälp eleverna att formulera prompts till AI:n om de kör fast. Ge gärna exempel som "Analysera texten kritiskt" för att visa hur de kan få ut mer av verktyget.
- **Diskussion:** Avsluta med en kort klassdiskussion där grupperna delar sina insikter. Lyft både positiva exempel (t.ex. "AI hittade ett dolt antagande") och risker (t.ex. "AI missade tonen i texten") för att förstärka budskapet om vaksamhet.

Lektion 16: Sakkritik och faktafokus – att granska påståenden

Introduktion

Här skiftar vi fokus till **sakkritik**, d.v.s. att granska själva innehållet i ett påstående: *Är det sant eller falskt?* Sakkritik handlar om faktagranskning – att kontrollera om uppgifterna stämmer, om evidens finns, om resonemanget håller. I en tid av generativ AI och "fake news" är sakkritik en central färdighet. Lektionen introducerar också konceptet **prebunking** – att i förväg rusta sig mot vanliga falska narrativ genom att känna till dem och hur de luras. (Prebunking kan ses som en vaccinering mot misinformation misinforeview.hks.harvard.edu.)

I introduktionen kan du ge ett exempel: T.ex. "5G sprider coronaviruset" var ett konspirationsteori-påstående som cirkulerade. En sakkritisk blick skulle fråga: Finns det någon vetenskaplig grund? (Nej). Vem säger detta? (Ofta okända källor). Var är bevisen? Genom att öva sådana frågor lär sig eleverna nysta upp falskari. Prebunking-delen kan inledas med att nämna att vi kommer lära oss om vanliga knep som används för att vilseleda, så att man känner igen dem direkt nästa gång.

Genomförande (steg-för-steg)

- Påståendets anatomi:** Börja med att skriva ett kontroversiellt eller misstänkt påstående på tavlan, t.ex. "*Vaccin orsakar autism enligt ny studie*". Gå igenom hur man skulle **sakkritiskt** granska det:
 - Kolla om det verkligen finns en vetenskaplig studie (sök på nyckelord + "study debunked").
 - Se vad etablerade hälsoorganisationer (Folkhälsomyndigheten, WHO) säger.
 - Titta på själva påståendets logik och språk (överdrifter? konspirationstön?). Detta steg visar konkret hur man angriper ett misstänkt påstående.
- Faktakolla tillsammans:** Dela ut två korta artiklar eller inlägg med olika påståenden (gärna en sann, en falsk men övertygande). T.ex. en riktig nyhet om klimatförändringar vs ett blogginlägg som förnekar dem med felaktiga "fakta". Jobba gemensamt med första: visa hur man kan markera nyckeluppgifter och söka upp dem. Kanske kopiera en mening och googla, eller använda en faktakoll-sajt. Elever får se processen live.
- Elevpärens granskning:** Låt eleverna i par granska det andra påståendet själva med samma metod. Ge dem 10-15 min att hitta stöd eller motbevis. Förse med några

verktygstips (t.ex. sök med "påstående + myt" eller kolla på Snopes). Gå runt och stötta.

4. **Genomgång & diskussion:** Samla klassen och hör vad varje par hittade. Var påståendet sant eller falskt? Hur kunde de avgöra? Notera olika källor de använde. Förhoppningsvis ser de värdet i att dubbelkolla innan man tror på något.
5. **Prebunking introduktion:** Förklara att nästa steg i lektionen är att lära sig i förväg om vanliga knep. Visa en lista på 5 vanliga desinformationstaktiker: t.ex. *fejka auktoritet* (hitta på experter/källor), *spela på känslor/rädsla*, *sprida konspirationstänk*, *använda halvsanningar*, *upprepa en lögn tills det låter sant*. (Detta baseras på forskning om propaganda).
6. **Prebunking övning (spel eller analys):** Här kan du antingen:
 - Låta eleverna spela **spelet "Bad News"** (svenska versionen) på sina enheter under kontrollerade former i ~15 minuter. Spelet låter dem själva sprida fejknyheter på låtsas och lär ut knepen på vägen misinforeview.hks.harvard.edu

ELLER

 - Ge varje grupp i uppgift att skapa en *snabb-guide* affisch med ett av knepen ovan: vad är det och hur känner man igen det? (Sen kan man sätta upp i klassrummet).
7. **Reflektion:** Avsluta med att fråga: "*Hur kändes det att faktagranska? Var det svårt eller lätt? Vad överraskade er?*" Och "*Om ni möter ett liknande påstående i framtiden, vad gör ni?*" – Förhoppningen är att de svarar: "Jag kollar upp det innan jag sprider eller tror på det."

Lektionsingångar & diskussion

- **Ingångsbild:** Visa en skärmdump av en falsk nyhet (kanske en rubrik + bild som är absurd). Fråga: "Vad säger magkänslan – är detta på riktigt?" Sedan: "Hur skulle vi ta reda på säkert?" Skriv elevernas idéer (som sannolikt blir: kolla källan, googla, jämför med riktiga nyheter).
- **Diskussionsfråga:** "*Har ni stött på ordet 'faktagranskning' i media? Vad innebär det?*" Någon kanske nämner att tidningar gör fakta koll av politikere uttalanden, etc. Koppla det till sakkritik: att analysera påståenden i detalj.
- **Debunking-prat:** Om klassen är bekant med någon stor myt (tex "människa har inte varit på månen" eller liknande), diskutera hur vet vi att det *är* en myt? Vilka bevis finns? Hur har det debunkats (motbevisats)?

- **Prebunking diskussion:** Fråga: "Vad tror ni är effektivast, att försöka övertyga någon efter att de trott på en falsk grej – eller att förekomma så de aldrig tror på den?" Resonera kring varför prebunking (vaccinera med kunskap) kan vara smartare än att rätta i efterhand, eftersom människor som väl trott på något falskt kan hålla fast vid det trots motbevis.
- **Eleve exempel:** Om någon elev vill, låt dem berätta om en gång de "övertygat" någon med fakta, eller själva blivit övertygade att byta uppfattning när de såg bevis. Vad krävdes? Ofta krävs tydlig evidens och förtroende för den som presenterar den.

Didaktiska tips

- **Hantera tid:** Faktagranskning i realtid kan ta tid och spärra ur om elever hamnar fel. Begränsa sökningen – ge dem t.ex. några pålitliga webbplatser att börja på (Snopes, Faktiskt.se, wikipedia, etc.) för att undvika att de hamnar i en annan konspiration under lektionstid.
- **Guidade sökord:** För svagare grupper kan det vara bra att ge ett "sökrecept". T.ex. "ta en kärnfras ur texten inom citationstecken och sök", eller "lägg till ordet 'hoax' eller 'bluff' i sökningen på engelska". Detta ökar chansen att de hittar rätt info.
- **Bad News spelet:** Om ni kör spelet, se till att alla går in på svenska versionen (badnewsswedish.eu). Gå runt och observera reaktioner. Efter spelet, *debriefa* – vissa kan tycka det var kul att "luras", men betona att syftet var att lära sig känna igen de fula knepen.
- **Var uppmärksam på misstro:** En del elever kan vara benägna att tro på konspirationer. Förlöjliga inte dem, utan visa respekt men stå fast vid fakta och metoden att kräva evidens. Ibland hjälper det att säga "Sund skepsis är bra, men då måste vi vara mest skeptiska mot påståenden som saknar ordentliga bevis".
- **Samarbete i personalen:** Överväg att involvera en NO-lärare eller bibliotekarie för att prata om vetenskaplig metod och källgranskning av forskning, speciellt om ni tar exempel som rör vetenskap (vaccin, klimat). Det kan ge tyngd åt vikten av att lita på samlad vetenskaplig evidens vs enstaka påståenden.

Forskning & verktyg

- **Inokulationsteori:** Forskning av bl.a. Sander van der Linden och Jon Roozenbeek har visat att spel som "Bad News" faktiskt ökar motståndskraften mot misinformation misinforeview.hks.harvard.edu. Att eleverna får prova på att skapa fejk (i en ofarlig miljö) gör att de känner igen tricken och blir mindre lättlurade sen. Detta är vetenskapligt belagt över olika länder och kulturer.

- **Sakkritik i skolan:** En studie (Nygren et al.) fann att svenska elever behöver mer träning i att värdera innehållet, inte bara källan. Källkritiktraditionen har ibland fokuserat så mycket på avsändare att man glömt *innehållsgranskningen*. Den här lektionen adresserar det genom konkret faktakoll.
- **Verktyg för faktagranskning:**
 - **Snopes.com** – en klassisk amerikansk faktakoll-sajt med massor av exempel (på engelska). Kan användas som källa för att snabbt kolla internationella virala rykten.
 - **Google Fact Check Explorer** – ett verktyg där man kan söka på ett påstående och få upp om det faktagranskats av någon organisation. (Visste du att Google har en särskild sökmotor för faktakollar? Om inte, kolla upp det – intressant att visa elever att stora aktörer försöker underlätta faktakoll).
 - **Omvänt bildsök** – speciellt viktigt om påståendet innefattar en bild/bevis. Visa t.ex. hur en bild på en haj på en översvämmad motorväg (viralt under stormar) genom omvänd bildsökning visade sig vara tagen på annat håll (eller photoshoppad).

Arbetsblad

Arbetsbladet för Lektion 7 är ganska omfattande:

- **Checklista för faktagranskning:** En lista av steg liknande det ni gått igenom, med tomma rutor där elever kan fylla i exempel från lektionen:
 1. Identifiera påståenden/"fakta" att kolla (t.ex. "Vaccin orsakar autism").
 2. Hitta oberoende källa 1 som bekräftar/förnekar (t.ex. "En artikel på forskning.se som förklarar att studien som påstods visa detta var fejk").
 3. Hitta oberoende källa 2 (t.ex. "WHO:s hemsida med FAQ om vaccinytter").
 4. Jämför med originalpåståendet – stämmer det överens med vad de oberoende källorna säger?
 5. Slutsats: (t.ex. "Påståendet är falskt, ingen evidens för det, snarare motbevisat").
- **Prebunking-kort:** En övning där olika manipulativa taktiker presenteras i små rutor (gärna med namn och en ikon). Eleverna ska ge ett *exempel* på var och en. Taktikerna kan vara t.ex. "emotionell skrämpropaganda", "falsk auktoritet", "polarisering/vi-mot-dom-retorik", "humor/sarkasm för att trivialisera motståndare", "konspiration ('allt är uppgjort')". Eleverna fyller i kort vad de tror ett exempel kan vara (fiktivt går

bra). Ex: vid "falsk auktoritet" kanske de skriver "en blogg citerar en 'forskare' som inte finns".

- **Bad News reflektion (om spelet spelats):** Några frågor: "Vilka knep använde du i spelet för att få följare?" "Hur kände du när du författade fejknyheter – var det lätt att få dem att låta trovärdiga?" "Vad lärde du dig?" Eleverna kan skriva kort. Om spelet inte spelades, kan dessa frågor utgå eller anpassas till de gjorda övningarna.
- **Snabbquiz:** Tre påståenden där eleverna ska ringa in vilket som har störst chans att vara sant, baserat på hur det låter och var det kommer från:
 1. A) "En anonym användare på ett forum säger att månen egentligen är ihålig." B) "NASA släpper data som visar nya rön om månens kärna."
 2. A) "Influencer X hävdar att en viss diet botade hans långa Covid, inga läkare förstår det." B) "En studie från Karolinska Institutet visar lovande resultat för en ny behandling mot långvarig Covid."
 3. A) "Rubrik: 'Choklad äter du hur mycket du vill – ny forskning visar att det bränner fett' (källa: okänd hälsoblogg)". B) "Rubrik: 'Mörk choklad i måttliga mängder kan ha hälsofördelar – dietist reder ut' (källa: SVT Nyheter)". Syftet är att öva källkänsla och innehållskritik på tidningsrubriksnivå.

Gå igenom arbetsbladet tillsammans efteråt, särskilt checklistan, så eleverna känner sig säkra på stegen för att kolla fakta och känna igen luringar.

Arbetsblad – Lektion 16: Sakkritik och faktafokus – att granska påståenden

Lektion 7: Sakkritik och faktafokus – att granska påståenden (prebunking del 1)

Instruktioner: Under dagens lektion kommer vi att arbeta med att granska påståenden kritiskt genom faktakoll och sakkritik. Vi kommer även att introducera prebunking, dvs. hur vi kan lära oss att förhindra att falska narrativ får fäste genom att känna till hur de fungerar. Använd detta arbetsblad för att dokumentera ditt arbete under lektionen.

Steg 1: Granska påståenden – Checklista

1. **Identifiera påståendet:** Skriv ner det påstående du ska granska.
 - **Exempel:** "Vaccin orsakar autism enligt ny studie."
 - Påstående: _____
2. **Hitta en första oberoende källa:**
 - **Exempel:** WHO:s webbplats eller Folkhälsomyndigheten.
 - Källa: _____
 - Vad säger denna källa om påståendet? _____
3. **Hitta en andra oberoende källa:**
 - **Exempel:** En vetenskaplig artikel från Karolinska Institutet.
 - Källa: _____
 - Vad säger denna källa om påståendet? _____
4. **Analysera påståendets logik och språk:**
 - **Exempel:** Använder påståendet skrämselfpropaganda, vaga hänvisningar till "studier" utan källor eller konspirationsteorier?
 - Finns det överdrifter? Konspirationston? Emotionella argument?
 - _____
5. **Jämför med originalpåståendet:** Stämmer det överens med vad källorna säger?
 - _____
6. **Slutsats:** Baserat på din analys, är påståendet sant, falskt eller vilseledande? Motivera.
 - _____

Steg 2: Prebunking – Identifiera manipulativa taktiker

I den här delen ska du fundera på hur falska narrativ sprids och vilka metoder som används för att manipulera information. Nedan finns vanliga taktiker. Försök ge exempel på var och en.

1. **Falsk auktoritet:** Att använda påhittade eller opålitliga experter.
 - **Exempel:** "En självutnämnd hälsoguru utan medicinsk utbildning hävdar att råkost botar cancer."
 - Exempel från dig: _____
2. **Emotionell skrämselfpropaganda:** Att använda rädsla för att påverka folk.
 - **Exempel:** "Om du vaccinerar dig kan du dö av farliga biverkningar!"

- Exempel från dig:

- 3. **Polarisering/vi-mot-dom-retorik:** Att skapa konflikt genom att ställa grupper mot varandra.
 - **Exempel:** "De som tror på klimatförändringar är lurade av media!"
 - Exempel från dig:

- 4. **Halvsanningar och cherry-picking:** Att plocka ut vissa fakta som passar ens agenda och ignorera resten.
 - **Exempel:** "En studie visar att kaffe är farligt – men de nämner inte att det var i extremt höga doser."
 - Exempel från dig:

- 5. **Upprepa av en lögn tills den låter sann:** Att göra en falsk information trovärdig genom att upprepa den ofta.
 - **Exempel:** "Många säger att jorden är platt – kanske är det sant?"
 - Exempel från dig:

Steg 3: Bad News Game

Gå till spelet **Bad News** (www.badnewsswedish.eu) och spela i cirka 15 minuter. Spelet går ut på att du får testa att skapa och sprida desinformation på låtsas. Spelet lär dig hur falska nyheter fungerar och vilka knep som används för att manipulera.

För att komma igång, fundera på:

- Hur skulle du själv skriva en falsk nyhet för att få folk att dela den?
- Vilka känslor försöker du spela på?
- Hur kan du bygga upp en trovärdig bluff?

Efter spelet, besvara frågorna nedan:

1. Vilka knep använde du för att få fler följare i spelet?
○ _____
2. Hur kändes det att sprida falsk information, trots att det var i en spelmiljö?
○ _____
3. Vilket knep tyckte du var mest effektivt och varför?
○ _____
4. Hur kan du använda det du lärt dig i spelet för att genomskåda falsk information i verkligheten?
○ _____

Steg 4: Reflektion

Besvara frågorna kortfattat efter lektionen.

1. Hur kändes det att faktagranska påståenden? Vad var svårast?
○ _____

2. Vad överraskade dig under faktakollen?

○ _____
3. Om du möter ett liknande påstående i framtiden, vad gör du?

○ _____

Extra: Snabbquiz – Vad låter mest trovärdigt?

Ring in det påstående som har störst chans att vara sant:

1. **A)** "En anonym användare på ett forum säger att månen egentligen är ihålig." **B)** "NASA släpper data som visar nya rön om månens kärna."
 2. **A)** "Influencer X hävdar att en viss diet botade hans långa Covid, inga läkare förstår det." **B)** "En studie från Karolinska Institutet visar lovande resultat för en ny behandling mot långvarig Covid."
 3. **A)** "Rubrik: 'Choklad äter du hur mycket du vill – ny forskning visar att det bränner fett' (källa: okänd hälsoblogg)." **B)** "Rubrik: 'Mörk choklad i måttliga mängder kan ha hälsofördelar – dietist reder ut' (källa: SVT Nyheter)."
-

Lektion 17: Jobba med prebunking

(Förutse, förhindra och förvirra falska narrativ innan de får fäste)

Introduktion

I denna lektion introduceras begreppet **prebunking** – ett sätt att “vaccinera” oss mot kommande propaganda och falska narrativ. I stället för att bara reagera och “debunka” (avfärda) felaktig information när den redan spridits, försöker vi **förutse** vilka teman eller påståenden som kan komma att bli virala. Genom att lära sig kännetecknen på manipulation och vanliga propagandagrepp redan i förväg, är det lättare att känna igen dem när de dyker upp.

Vanliga Manipulationstekniker – med exempel

◆ Känsloladdat språk

Exempel:

"Ett nytt lagförslag hotar att förstöra framtiden för våra barn!"
(Använder starka ord som "hotar" och "förstöra" för att skapa rädsla, istället för att sakligt presentera förslaget.)

◆ Falsk dikotomi

Exempel:

"Antingen är du med oss, eller så är du emot oss!"
(Framställer en fråga som om det bara finns två alternativ, trots att fler möjligheter kan finnas.)

◆ Cherrypicking (Körsbärsplockning)

Exempel:

"Studier visar att choklad är nyttigt, så jag äter en chokladkaka om dagen!"
(Väljer ut en studie som stödjer idén men ignorerar forskning som visar negativa hälsoeffekter.)

◆ Falska experter

Exempel:

"En populär influencer säger att vacciner är farliga."
(Använder en person utan medicinsk expertis för att ifrågasätta etablerad vetenskap.)

◆ Villospår

Exempel:

"Vi borde inte prata om klimatförändringar när det finns större problem som arbetslöshet!"
(Försöker avleda diskussionen från huvudämnet genom att lyfta fram något annat.)

◆ Utpeka syndabockar

Exempel:

"Invandringen är orsaken till alla våra ekonomiska problem!"
(Lägger skulden på en grupp utan att bevisa sambandet.)

◆ Ad Hominem

Exempel:

"Du kan inte ha rätt, du är ju bara en student!"
(Angriper personen istället för att bemöta argumentet.)

◆ Polarisering

Exempel:

"De där människorna förstör vårt land!"
(Överdriver skillnader mellan grupper och skapar ett "vi mot dem"-perspektiv.)

◆ Utge sig för att vara någon annan

Exempel:

"Jag är läkare och jag rekommenderar den här mirakelmedicinen!" (Skrivet av en anonym användare på sociala medier.)
(Försöker lura folk genom att låtsas vara en expert.)

◆ Snöbollseffekten

Exempel:

"Om vi förbjuder plastpåsar nu, kommer vi snart att leva under en diktatur!"
(Påstår att en liten förändring kommer att leda till extrema konsekvenser.)

◆ Dekontextualisering

Exempel:

Ett videoklipp visar en politiker säga "Jag hatar barn" – men i hela klippet säger han "Jag hatar barnarbete."
(Plockar ut en del av ett uttalande för att förvränga betydelsen.)

Syfte och mål

- **Förstå begreppet prebunking** – hur vi kan ligga steget före desinformation.
- **Identifiera troliga narrativ** som kan dyka upp i nyhetsflöden eller sociala medier.
- **Skapa egna exempel** på en fejkad nyhet, för att se hur enkelt det är att fabricera trovärdigt innehåll.

Lektionsingång

1. **Förklarande diskussion om 'prebunking'**
 - Använd prebunking.withgoogle.com för en översikt av hur det kan fungera.
 - Beskriv skillnaden mellan **debunking** (att avfärda falska påståenden när de redan spritts) och **prebunking** (att förutse och förebygga innan de sprids).
2. **"Lägga örat mot rälsen"**
 - Förklara metaforen: Om vi lyssnar tidigt kan vi höra "tåget" av propaganda komma innan det är här.
 - Diskutera med eleverna: Vilka sorters rykten eller berättelser verkar bli virala? Kan vi ana teman (t.ex. konspirationer, kändisskandaler, politiska 'avslöjanden')?

Genomförande (steg-för-steg)

1. Identifikation av potentiellt narrativ (5–10 min)

- Gruppdiskussion: “Vilka påhittade nyheter skulle faktiskt kunna luras, om de dök upp i ett flöde?”
- Exempel: “Kändis X är egentligen bror till Kändis Y” eller “Ny lag införs i hemlighet som förbjuder [valfritt ämne]”.

2. Skapa fejkad nyhet med en chattbot (10–15 min)

- Be en AI-chattbot (t.ex. ChatGPT) att skriva ett kort nyhetsinslag eller artikel om ett **tänkbart** men påhittat scenario.
- Be eleverna att ge AI:n en **prompt i stil med**: “Skriv en trovärdig men påhittad nyhetsartikel om att [tema] hände i [stad/land], inklusive citat från påhittade vittnen.”
- **Diskutera i helklass**: Varför skulle folk kunna tro på detta? Har det element som är *nära* verkligheten (kända platser, populära kändisar, etc.)?

3. Generera bilder med Grok.com (5 min)

- För att göra fejknyheten ännu mer övertygande, använd [Grok.com](https://grok.com) för att skapa bilder på eventuella kändisar eller situationer.
- Eleverna kan t.ex. skapa en “fejkad” bild av en kändis som håller upp en skylt eller befinner sig på en plats som passar historien.

4. Simulera en nyhetssida med Websim.ai (10–15 min)

- Ta en **skärmdump** av en verklig nyhetssajt (t.ex. Aftonbladet.se).
- Be [Websim.ai](https://websim.ai) att skapa en sida som *ser ut* som denna men är tom på innehåll.
- Låt eleverna klistra in texten från chattboten och bilden från Grok i den fejkade “nyhetssidan”.
- Diskutera: “**Hur trovärdig ser detta ut? Skulle ni kunna luras av en sådan sida?**”

5. Flera uppslag (valfritt)

- Uppmana eleverna att skapa **olika** typer av fejknyheter (t.ex. politiska beslut, kändisskvallor, sportrelaterade rykten) för att jämföra vilka som känns mest övertygande.
- Diskutera: **Vilka manipulationstekniker** är inblandade (chockfaktor, känsloladdat språk, kändisanknytning, osv.)?

6. Sammanfattning (5 min)

- Påminn om lektionens kärna: **Prebunking** handlar om att förutse och förstå hur en fejknyhet kan uppstå, så att vi känner igen tecknen innan den får spridning.
- Fråga eleverna: “Vad kan vi göra i framtiden när vi ser en ‘nyhet’ som verkar för bra eller för galen för att vara sann?”

Diskussionsfrågor

1. Varför är det ofta effektivare att förebygga desinformation än att försöka “rädda” situationen när rykten spridits långt?
2. Hur kan man träna sig i att förutse vilka narrativ som kan dyka upp?
3. Vilka känslor utnyttjas ofta i propaganda (t.ex. rädsla, ilska, hopp)?

4. Kan ni komma på exempel i verkliga livet när en fejknyhet fått stor spridning? Hur kunde prebunking ha hjälpt då?

Didaktiska tips

- **Tydliggör etiken:** När eleverna skapar fejknyheter är syftet att förstå hur lätt det är att lura, inte att uppmuntra dem att verkligen sprida falsk information.
- **Anpassa nivå:** Vissa elever kan vilja göra avancerade bilder och utstuderad text, medan andra nöjer sig med enklare exempel. Huvudsaken är insikten om hur propaganda kan skapas.
- **Knyt till mediekunskap:** Använd detta upplägg som ett praktiskt exempel på hur journalistiska principer (faktakoll, källkontroll) kan rundas av den som vill sprida falsk information.
- **Diskutera spridningsvägar:** Hur skulle man kunna sprida den här fejknyheten för maximal genomslagskraft (sociala medier, mejlkedjor, bloggar)? Påminn om att “smarta” avsändare ofta anpassar tonen för att nå olika målgrupper.

Fördjupning

- **Koppla till retorik:** Be eleverna analysera språket i den fejkade nyheten. Vilka ordval och retoriska grepp gör den mer trovärdig?
- **Utöka simuleringen:** Låt grupperna simulera en social medie-kampanj kring den falska nyheten (t.ex. hashtags, memes, influencers) och sedan diskuterar man i klassen hur prebunking kan motverka sådan spridning.
- **Samhällsperspektiv:** Diskutera vilka aktörer i samhället som kan jobba med prebunking (t.ex. myndigheter, medier, NGO:er) och hur det kan se ut i praktiken.

Sammanfattning

Denna lektion ger eleverna en **praktisk** inblick i hur enkelt det är att skapa en till synes *trovärdig* fejknyhet med hjälp av AI-verktyg som chattbotar, bildgeneratorer och sajtsimulatorer. Genom att **förutse** tänkbara narrativ och tekniker för manipulation – i stället för att bara reagera när rykten redan fått spridning – förstår eleverna grundidén med **prebunking**. Resultatet är en **starkare beredskap** mot digital desinformation och en bättre förståelse för hur vi alla kan vara mer vaksamma på tidiga tecken till propaganda.

Arbetsblad – Lektion 17: Jobba med Prebunking

Förutse och förebygg falska narrativ innan de sprids

1. Förstå “Prebunking”

1. **Vad är prebunking?**
 - Istället för att “debunka” (avfärda rykten när de redan är ute), försöker vi **förutse** vilka felaktiga påståenden som kan dyka upp och “vaccinera” oss mot dem.
2. **Skillnad mellan debunking och prebunking**
 - **Debunking:** Reagerar när rykten spritts långt.
 - **Prebunking:** “Lägga örten mot rälsten” och agera tidigt, innan fejknyheten slagit rot.

Fundera: Vilka rykten eller konspirationer kan tänkas dyka upp i ditt flöde? Hur skulle du kunna “varningsflagga” dem för dina vänner i förväg?

2. Identifiera potentiella falska narrativ

Instruktion (5–10 min):

Diskutera i grupp:

- “Vilka **påhittade nyheter** skulle kunna locka läsare att tro på dem?”
 - Använd gärna **kändisar, politik, sport** eller **konspirationer** som exempel.
1. **Brainstorma 2–3 tänkbara fejknyheter** ni tror hade kunnat bli virala.
 - T.ex. “Kändis X är egentligen släkt med Kändis Y.”
 - T.ex. “Hemlig lag ska förbjuda [favoritämne].”
 2. **Välj EN** nyhet ni vill utveckla till en “fejkad story”.

Skriv era exempel:

- ...
- ...

3. Skapa en fejkad nyhet

Instruktion (10–15 min):

Ni ska nu **”fabricera”** en kort nyhetsartikel med hjälp av en **AI-chattbot** (ex. ChatGPT). Syftet är att se hur enkelt det är att skapa något som kan verka trovärdigt men är helt påhittat.

1. **Formulera en prompt** till AI:n:

“Skriv en trovärdig men **helt påhittad** nyhetsartikel om [era valda tema], med citat från påhittade vittnen och detaljer som gör det trovärdigt.”

2. **Kopiera texten** AI genererar. Diskutera i grupp:
 - Vilka **språkliga knep** använder AI:n för att låta seriös? (ex. “en källa nära projektet”, “experter varnar...”)
3. **Varför** tror ni folk skulle kunna tro på denna “nyhet”?
 - Är den *nära* verkligheten (känd plats, känd profil, realistisk händelse)?

4. Bildbevis: Generera fejkad bild med Grok.com

Instruktion (5 min):

För att förstärka er fejknyhet, skapa en **AI-genererad bild**:

1. **Besök** grok.com (eller liknande bild-AI).
2. **Beskriv** vilken bild ni vill ha, t.ex. “Kändis X håller upp en skylt om [ämne]” eller “Politiker Y syntes i [stad].”
3. **Spara bilden** och koppla den till er nyhetstext.

Tänk efter:

- Hur trovärdigt blev det? Skulle du själv tvivla om du såg en sådan bild snabbt i ett flöde?

5. Simulera en nyhetssajt med Websim.ai

Instruktion (10–15 min):

Låt eleverna se hur lätt det är att få sidan att se *exakt* ut som en riktig nyhetssida.

1. **Ta en skärmdump** på en befintlig nyhetssajt (t.ex. Aftonbladet).
2. **Be Websim.ai** skapa en sida som “liknar” den.
3. **Klistra in** ert fejkade innehåll (AI-artikeln + fejkad bild).

Diskussionsfrågor:

- **Hur övertygande** ser resultatet ut?
- **Skulle** någon kunna dela detta på sociala medier och få folk att tro att det är äkta?

6. Reflektion: Hur sprider man en fejknyhet?

1. **Vilka kanaler** skulle ni använda för maximal spridning (sociala medier, bloggar, mejlutskick)?
2. **Vilka känslor** försöker ni väcka (rädsla, ilska, förvåning)?
3. **Hur** kan prebunking hjälpa?
 - T.ex. informera vänner: “Om ni ser en nyhet om X, var försiktiga, för det låter misstänkt.”

7. Avslutande diskussion

Instruktion: Samlas i helklass och gå igenom frågorna:

1. **Varför** är det bättre att förebygga (prebunka) än att försöka “rädda” situationen när en fejknyhet redan är stor?
2. **Vilka varningstecken** fick ni syn på i era fejknyheter? (Känsloladdat språk, kändisanknytning, dramatiska påståenden)
3. **Hur** kan ni vara mer uppmärksamma när ni ser något “för bra” eller “för galet” i flödet?

8. Etisk påminnelse

1. **Användning i undervisning:** Ni skapar fejknyheter för att **förstå** mekanismer bakom manipulation, *inte* för att verkligen lura någon.
2. **Spridning av falsk info** är oetiskt och kan vara olagligt beroende på innehållet.

Lycka till med att förstå hur propaganda och desinformation kan “planteras” – och hur vi kan förhindra spridningen redan innan den startar!

Lektion 18: AI-genererade bilder och deepfakes – att se är inte alltid att tro

Introduktion

"Jag tror det när jag ser det" brukade man säga – men med dagens AI-genererade bilder och **deepfakes** stämmer inte det längre. Lektion 9 lär eleverna att även visuella medier kan ljuga. Syftet är att öka medvetenheten om hur AI kan skapa fejkade fotografier, videos eller ljud som ser/sounds äkta ut, och ge strategier att avslöja dem. I introduktionen, visa gärna en **uppseendeväckande AI-bild**: t.ex. den falska bilden på påven i en vit fluffig Balenciaga-jacka som spreds våren 2023 och lurade många att tro den var äkta. Fråga klassen: "Tror ni den här bilden är sann eller skapad?" Diskutera kort. Avslöja sedan att den var AI-genererad – många blev lurade eftersom de inte trodde tekniken var så bra. Poängtera att idag kan **vem som helst med rätt verktyg skapa hyperrealistiska bilder eller till och med videor**. Därför måste vi lära oss tecknen på manipulation och vilka verktyg som finns att verifiera äkthet.

Genomförande (steg-för-steg)

1. **Exempelgalleri:** Visa 3-4 bilder: blandat äkta och AI-genererade (kan vara tryckta i färg på papper eller projicerade). Låt eleverna rösta för varje: Äkta eller fejk? Sedan avslöjar du svaren. Detta grepp engagerar direkt. Diskutera vad som gjorde vissa svåra att bedöma – kanske händer såg konstiga ut, eller ljuset onaturligt, etc.
2. **Förklara teknikerna:** Ge en kort presentation av begreppen:
 - **Generativa bild- och videomodeller:** (t.ex. DALL-E, Midjourney för bilder; deepfake-teknik för video).
 - **Deepfake:** AI-skapad video där någons ansikte byts ut eller röst imiteras. Känt exempel: Obama-videon där han säger saker han aldrig sagt.
 - **Andra manipulationer:** Filter, photoshop, etc (äldre tekniker men relevanta). Försök hålla det på en nivå de förstår: t.ex. "AI kan tränas på massor av bilder av en person och sedan göra en ny video där den personen verkar säga något – fast det är skapat digitalt."
3. **Hur avslöja fejk?** Gör en lista tillsammans på typiska ledtrådar:
 - Konstiga detaljer: händer med fel antal fingrar (AI-bilder hade länge svårt med händer, även om det förbättrats), smycken eller text i bild som ser orimligt ut (text brukar bli gibberish).
 - Orimligt scenario: Påven med designerjacka – möjligt men lite out of character, värt att dubbelkolla.

- Källa/kontext: Var dök bilden upp först? Kommer den från en verifierad fotografi/bildbyrå eller random Twitterkonto?
 - Använd verktyg: omvänd bildsök (om bilden är ny, kanske svårt, men ibland hittar man att den först postades av en konstnär). För video: det finns deepfake-detectors (på varierande nivå).
Skriv upp tipsen för alla att se.
4. **Praktisk övning – analys:** Dela ut ett arbetsblad (eller visa på skärm) med två bilder och en kort bakgrundshistoria till varje (en sann, en deepfake scenario). Låt grupper analysera med de nyss nämnda ledtrådarna och fylla i en tabell: "Detaljer vi kollar", "Vad vi kom fram till". T.ex. Bild A: "skuggan föll orimligt -> misstänkt", "hittade ingen källa -> misstänkt", "därför tror vi fejk". Bild B kanske de avgör är äkta pga annat.
 5. **Diskussion – konsekvenser:** Fråga klassen: "*Vad kan vara farligt med deepfakes?*"
Skriv upp deras idéer: exempelvis ryktesspridning (någon kan se ut att säga något skandalöst), påverkan på demokrati (falska videor på politiker), bedrägerier (fejkad röst som ringer och låtsas vara ens förälder). Samtidigt, fråga också: "*Finns det positiva användningar av dessa tekniker?*" (Jo, filmindustrin, konst, specialeffekter, kanske förbättra röstsyntes för funktionshindrade, etc).
 6. **Sammanfattning och råd:** Summera att "seeing is not believing" längre. Råd: vid häpnadsväckande bilder/klipp – anta inte att det är sant förrän du verifierat. Introducera tanken att nyhetsredaktioner nu måste **källgranska bildmaterial** lika noga som text. Vi bör också vara försiktiga med att sprida spektakulära bilder vidare utan kontroll.

Lektionsingångar & diskussion

- **Ingångsvideo:** Om resurser finns, visa en kort compilation av deepfakes (finns på YouTube, t.ex. klipp där man bytt ansikten på kändisar). Eleverna brukar tycka det är roligt och chockerande. Diskutera sedan: "Såg ni att det var fejk? Vad var annorlunda?"
- **Diskussionsfråga:** "*Hur många av er har använt filter på Snapchat/Instagram som byter ut era ansikten eller röster?*" – Många händer åker upp. Knyt an: det är i princip samma teknik på lek-nivå. Tänk er den teknikens utveckling 5-10 år framåt – ganska kraftfull.
- **Fråga om förtroende:** "*Kommer vi kunna lita på videoupptagningar som bevis i framtiden?*" – Ganska avancerad fråga, men få dem att fundera. Nämn att rättsväsendet redan diskuterar hur man verifierar digital media.

- **Reflektion:** "Hur känns det att veta att ögonen kan luras?" Vissa kan tycka det är läskigt, andra coolt. Låt några uttrycka sig. Bekräfta att det är normalt att känna oro, men att lösningen inte är panik utan ökad kunskap och teknisk moteld.
- **Kritisk syn:** För att inte överdriva hotet, nämn att ofta finns små fel i fejk som avslöjar – det krävs mycket skicklighet att göra en riktigt vattentät deepfake. Och experter kan oftast avslöja dem med analysverktyg.

Didaktiska tips

- **Bildmaterial:** Förbered i förväg. Det finns många exempel online – se dock till att de är rumsrena och icke-kränkande. Pope-with-jacket är bra. Andra: Trump som omfamnar/anropar utomjordingar (fejk), Obama deepfake, etc. Källkritikövningens ironi: även vårt undervisningsmaterial måste vara pålitligt! Kolla att de fejkade är erkända som fejk i medier (så du inte råkar visa nåt kontroversiellt utan belägg).
- **Involvera bildämnet eller mediaämnet:** Om skolan har medieundervisning, kanske de redan tar upp detta. Samordna så ni inte gör dubbelt, eller komplettera varandra (kanske bildlärare kan visa hur lätt det är att manipulera med Photoshop).
- **Hands-on creation (valfritt):** För en teknikinriktad klass, en fördjupning kan vara att själva prova ett enkelt AI-verktyg. T.ex. använda en mobilapp som face-swappar i realtid, eller en webb-app där man kan skriva prompt och generera en bild (det finns barnvänliga sådana). Detta dock med försiktighet – man måste följa skolans policy och inte låta elever generera olämpliga bilder.
- **Etiskt samtal:** Berör att med stor kraft följer ansvar: Att göra en deepfake av någon utan samtycke kan vara kränkande eller olagligt. Prata om nätetik: sprid inte bilder på andra utan att kolla äkthet och få lov.
- **Tilgång till teknik:** Om ingen projektor finns, funkar utskrivna bilder på papper för analysövningen. Även om färgkvalitén inte är topp, kan det räcka att se konstigheter.

Forskning & verktyg

- **Deepfake-forskning:** AI-forskare utvecklar även *detektorer* som tränas att känna igen AI-genererat innehåll. Tyvärr är det en kapprustning – varje gång detektorer blir bättre, blir generatorerna det med. Men exempel på tekniker: vissa detektorer kan hitta små pixelartefakter eller oregelbundna blinkningar i fejkade videor (tidigare deepfakes glömde få personen att blinka normalt!).
- **Kända incidenter:** Nära nutid har vi deepfakes i politiken (t.ex. fejkad video av Ukrainas president som uppmanade soldater lägga ner vapen 2022 – den avslöjades snabbt som falsk, men tänk skadan om folk hade gått på det). Att nämna gör det tydligt för elever att detta är på riktigt i världen.

- **Journalistikens svar:** Stora nyhetsbyråer som Reuters och AP har börjat införa autentiseringssystem (t.ex. metadata som intygar var en bild togs och av vem). Det är ett pågående arbete för att trygga att det vi ser i nyheterna är äkta. Bra att nämna att samhället tar det på allvar.
- **Digitala verktyg:**
 - **Google Reverse Image & Yandex** – Yandex (ryska sökmotorn) har ibland ännu bättre bildsök på detaljer; värt att känna till.
 - **Foto-Forensik verktyg** (exempel: FotoForensics.com) – man kan ladda upp en bild och se error level analysis, vilket ibland avslöjar ihopklistrade element. Kräver lite expertis att tolka, men kul att visa att verktyg finns.
 - **Amnesty YouTube DatViewer** – ett verktyg som extraherar thumbnails och metadata från YouTube-videos, användbart för att spåra ursprung (t.ex. vid virala klipp, se när och var de först dök upp).
 - **Deepfake-detector apps** – det finns också mobilappar som påstår sig kunna säga om ett ansikte i video är bytt. Resultaten varierar, men man kan testa dem med kända deepfakes och se om de prickar rätt.

Arbetsblad

Arbetsbladet för Lektion 9 fokuserar på att ge eleverna en konkret metod att granska visuella medier:

- **Bildanalys-checklista:** En sida med punkter och tomma rader där eleverna kan skriva i sina observationer när de granskar en bild:
 - *Källa:* Var hittade jag bilden/videon? (Trovärdig källa eller ej?)
 - *Detaljer:* Ser jag något konstigt med ljus, skuggor, kanter, händer, text, reflektioner? (Eleverna fyller i vad de ser).
 - *Känsla:* Verkar motivet troligt eller väldigt överraskande? (Skriva sin intuition).
 - *Verifikation:* Hur kan jag bekräfta? (t.ex. söka efter bilden online, kolla officiella kanaler, fråga experter).
 - *Slutsats:* Min bedömning (Äkta, Troligen äkta, Osäker, Troligen fejk, Fejk).
- **Övning med två fallstudier:**

- Fall 1: En nyhetsbild (äkta) med kort beskrivning av sammanhanget. Elever fyller i checklistan.
- Fall 2: En AI-bild (fejk) med påstådd beskrivning. Elever fyller i checklistan. Efteråt jämför man och ser om de upptäckte tecknen i fall 2.

- **Kortfrågor:**

1. Nämn ett kännetecken som kan avslöja en AI-genererad bild. Går det ens?
2. Vad är en deepfake?
3. Vilket verktyg använder du för att kolla om en bild redan funnits online? (Eleverna svarar i en mening eller två).

- **Diskussionsruta:** "Tycker du att fördelarna med AI-genererat media kan väga upp riskerna? Varför/varför inte?" – Detta för att uppmuntra dem att tänka kritiskt och skriva sin egen ståndpunkt efter allt de lärt sig.

När arbetsbladet gått igenom har eleverna en konkret lathund de kan spara, som påminner dem om att aldrig helt lita på ögonen i digitala miljöer utan att också tänka.

Arbetsblad: AI-genererade bilder och deepfakes – att se är inte alltid att tro

Introduktion

"Jag tror det när jag ser det" stämmer inte längre i en värld där AI kan skapa hyperrealistiska bilder och videor som lurar oss. I denna lektion har vi sett hur tekniken bakom AI-genererade bilder och deepfakes fungerar och varför vi inte alltid kan lita på våra ögon. På detta arbetsblad får du öva på att analysera två bilder – en äkta och en fejk – med hjälp av en checklista. Du kommer också att svara på frågor för att testa din förståelse och reflektera över vad detta betyder för oss i vardagen.

Instruktioner

- Din lärare kommer att visa eller dela ut två bilder: **Bild 1** och **Bild 2**. En är äkta och den andra är AI-genererad.
- Använd checklistan för att granska varje bild och skriv ner dina observationer.
- Svara på kortfrågorna och diskussionsfrågan i slutet.
- Jobba gärna i grupp om din lärare säger till, men skriv dina egna svar.

Del 1: Bildanalys-checklista

Använd denna checklista för att systematiskt granska varje bild. Fyll i dina observationer under varje punkt.

Punkt	Observationer
Källa	Var kommer bilden ifrån? Är källan trovärdig (t.ex. nyhetsbyrå eller slumpmässigt konto)?
Detaljer	Ser du något konstigt med ljus, skuggor, kanter, händer, text eller reflektioner?
Känsla	Verkar motivet rimligt eller väldigt osannolikt? Vad säger din magkänsla?
Verifikation	Hur kan du kolla om bilden är äkta? (t.ex. omvänd bildsök, officiella källor)
Slutsats	Vad tror du? Äkta / Troligen äkta / Osäker / Troligen fejk / Fejk

Del 2: Fallstudier

Fallstudie 1: Bild 1

[Din lärare ger en kort beskrivning av bilden, t.ex. "En nyhetsbild från ett stort evenemang" eller visar den direkt.]

Analysera Bild 1 med checklistan och fyll i dina observationer nedan.

Punkt	Observationer för Bild 1
Källa	
Detaljer	
Känsla	
Verifikation	
Slutsats	

Fallstudie 2: Bild 2

Analysera Bild 2 med checklistan och fyll i dina observationer nedan.

Punkt	Observationer för Bild 2
Källa	
Detaljer	
Känsla	
Verifikation	
Slutsats	

Del 3: Kortfrågor

Svara kort på följande frågor (1-2 meningar per fråga).

1. **Näm en kännetecken som avslöjar en AI-genererad bild.**
[Skriv ditt svar här]
2. **Vad är en deepfake?**
[Skriv ditt svar här]
3. **Vilket verktyg kan du använda för att kolla om en bild redan har funnits online?**
[Skriv ditt svar här]

Del 4: Diskussionsfråga

Tycker du att fördelarna med AI-genererat media kan väga upp riskerna? Varför eller varför inte?

Skriv ditt svar nedan och förklara vad du tycker. Tänk på exempel från lektionen, som hur tekniken kan användas för både positiva och negativa syften.

[Uttryck din åsikt och motivering här]

Avslutning

Bra jobbat! Nu har du verktyg för att inte bara tro på allt du ser. Spara gärna detta arbetsblad som en påminnelse om hur du kan granska bilder och videor i framtiden. Har du några frågor eller tankar efter övningen? Skriv ner dem nedan eller prata med din lärare.

[Valfritt: Utrymme för funderingar]

[Skriv eventuella frågor eller reflektioner här]

Not till läraren:

- Förbered två bilder i förväg: en äkta (t.ex. från en nyhetskälla) och en AI-genererad (t.ex. "påven i Balenciaga-jacka" eller liknande). Ge en kort bakgrund till varje bild när du delar ut arbetsbladet eller projicerar dem.

- Efter att eleverna fyllt i sina analyser, avslöja vilka bilder som var äkta och fejk, och diskutera vad som gjorde dem svåra eller lätta att bedöma.

Lektion 19: Vidga vyerna

(Bryt dina filterbubblor och analysera innehåll på djupet med en prompt)

Introduktion

Den här lektionen syftar till att hjälpa eleverna “**vidga sina vyer**” – dvs. ta ett steg tillbaka från sina vanliga informationskällor och kritiskt analysera innehåll med flera nya perspektiv. Genom att använda en “**meta-analys-prompt**” kommer eleverna att se hur AI kan bryta ned innehåll, synliggöra bias, möjliga felkällor och kunskapsluckor, samt föreslå **vidare läsning** och olika sätt att se på ämnet.

Syfte och mål

- **Presentera en specifik prompt** som eleverna kan klistra in i ChatGPT eller liknande AI-verktyg.
- **Hjälpa eleverna** att identifiera mönster, bias och nyanser i artiklar, inlägg från sociala medier eller andra typer av innehåll.
- **Öva elevernas förmåga** att reflektera över nya perspektiv, se egna fördomar och fundera över hur man bryter filterbubblor.

Prompt (Klistras in i AI-verktyget)

Du är en vänlig och korrekt assistent som hjälper användaren att vidga sina vyer, ge nya perspektiv och utveckla sitt kritiska tänkande genom att analysera innehåll från sociala medier, nyheter och artiklar. Användaren kan klistra in skärmdumpar, texter eller länkar, och du identifierar mönster, bias och möjliga kunskapsluckor. När användaren har delat innehåll gör du följande:

- 1) Analysera innehållet. Vad är det för innehåll. Bryt ned och analysera innehållet kortfattat
- 2) Skriv en sammanhängande djuplodande kritisk analys av innehållet med långa, utvecklade förslag på relaterad fakta, alternativa perspektiv och vidare läsning, så att användaren verkligen lär sig något. Ta dig tid att skriva en sammanhängande, tankeväckande text som verkligen hjälper användaren att förstå bias, eventuellt avsaknad av nyans och gör användaren till en mer kritiskt tänkande människa.

Syftet med denna del är att hjälpa användaren att sätta innehållet i relation, förstå andra synsätt, bryta filterbubblor och se nyanser i komplexa frågor. Den ger källkritiska reflektioner och lyfter fram olika tolkningar utan att förespråka en viss åsikt. Om informationen är osäker eller kontroversiell, markerar den detta tydligt och föreslår vidare läsning.

Denna del skrivs som en lång sammanhängande text med retoriska frågor som underbygger, förklaras och nyanseras omfattande.

Texten är strukturerad i stycken med rubriker. I slutet av texten ges förslag på frågor att reflektera över och utforska vidare

3) Sist ges Ge förslag på böcker, webbsidor, YouTube-klipp eller liknande som kan hjälpa användaren att nyansera och problematisera och få perspektiv.

REGLER:

Du söker aktivt på nätet för att ge uppdaterad information och kan även använda sin befintliga kunskap. Du ger svar direkt och alla svar innehåller både en analys och djupgående förslag på nya perspektiv

Eleverna klistrar in detta i ChatGPT (eller något annat kompatibelt AI-verktyg). Därefter lägger de in **det innehåll de vill analysera** (en artikel, en skärmdump, en text, en diskussionspost etc.).

Genomförande (steg-för-steg)

- Välj innehåll (5–10 min)**
 - Eleverna tar **en artikel, ett inlägg** från sociala medier eller en **nyhetsrapport** kring ett aktuellt eller omtvistat ämne.
 - Innehållet bör vara tydligt nog för att AI ska kunna analysera det.
- Klistra in prompt + innehåll i AI-verktyget (5–10 min)**
 - Eleverna kopierar prompten i sin helhet i AI:n.
 - Därefter klistrar de in **materialet** (text eller länk) som de vill få analyserat.
 - De skickar sin fråga (ex: “Här är en artikel om [ämne]. Gör din analys enligt instruktionen ovan!”).
- Läs och diskutera AI:ns “vidgade vyer” (10–15 min)**
 - Eleverna läser igenom AI:ns analys, noterar **hur** AI:n strukturerar texten, vilka **bias** eller perspektiv den lyfter, samt förslag på vidare läsning.
 - Be dem reflektera: “Hur annorlunda är denna ‘meta-analys’ jämfört med en vanlig AI-sammanfattning?”
- Gruppdiskussion eller helklass (10–15 min)**
 - Varje grupp eller elev berättar kort om:
 - Vilket material de analyserade.
 - Några punkter AI:n tog upp som de tyckte var **speciellt intressanta** eller **förfånande**.
 - Vilka böcker, webbplatser eller klipp AI:n föreslog, och om de verkar relevanta.
 - Diskutera: “Fanns det nya perspektiv ni inte tänkt på? Såg ni egna bias i texten eller i era tolkningar?”
- Reflektion (5 min)**
 - Uppmuntra eleverna att tänka på hur denna metod kan användas för att “zoom out” från sitt vanliga informationsflöde.
 - Påminn om att även detta GPT-svar **kan ha fel**, men att syftet är att väcka frågor och vidga perspektiv.

Exempel på uppgifter och diskussionsfrågor

- Sociala medier-flöde**
 - Låt eleverna ta ett screenshot av en **Instagram-caption** eller tweet som är laddad med åsikter.

- AI:n gör en djupanalys: Hur nyanserad är texten? Vilken bias kan finnas? Vilka nyckelord triggar starka reaktioner?
- 2. **Politisk debatt**
 - Eleverna väljer ett **politiskt debattinlägg** (från en tidning eller forum). AI förklarar olika möjliga tolkningar, vilka värderingar som står på spel, samt var man kan läsa mer för att förstå kontexten.
- 3. **Nyhetsartikel**
 - Ta en **lokal nyhet** om ett beslut i kommunen. AI lyfter fram ekonomiska, sociala och miljömässiga perspektiv. Diskutera hur man ofta bara ser en vinkel i lokala nyheter.
- 4. **Diskussionsfrågor**
 - “Hur vet vi att AI:s förslag på vidare läsning är trovärdiga?”
 - “Om AI visar på bias i texten, hur påverkar det vårt förtroende för källan?”
 - “Kan AI själv bli biased beroende på hur prompten är formulerad?”
 - “Vilka för- och nackdelar ser ni med att använda en sådan här ‘vidga vyerna’-funktion i vardagen?”

Redovisningsformer

1. **Muntliga presentationer**
 - Grupperna presenterar sina AI-analyser och vad de lärde sig om ämnet och om sina egna antaganden.
2. **Skriftlig reflektion**
 - Eleverna skriver 1–2 stycken om hur AI:s “meta-analys” förändrade (eller inte förändrade) deras syn på innehållet.
3. **Paneldebatt**
 - Några elever spelar rollspel: en försvarar originaltexten, en använder AI:ns perspektiv, en är skeptisk. De debatterar i klassen för att se hur argumenten ställs mot varandra.

Sammanfattning

I denna lektion får eleverna **pröva en specifik prompt** som uppmuntrar GPT att göra en **djupare, nyanserad analys** av valfritt innehåll. Resultatet är en text som inte bara sammanfattar, utan visar möjliga **bias**, olika **perspektiv**, kunskapsluckor och tips på **vidare läsning**. Eleverna lär sig hur man kan **nyansera** sin förståelse av ett ämne, **upptäcka egna filterbubblor** och **se fler sidor** av en fråga. På så vis blir de mer **kritiskt tänkande** och öppna för nya insikter – en viktig färdighet i ett informationssamhälle där både människor och AI influerar våra åsikter.

Arbetsblad – Lektion 13: Vidga vyerna

Bryt dina filterbubblor och analysera innehåll på djupet med en prompt

1. Inledning

Instruktion: I den här lektionen ska du använda en **speciell prompt** som uppmuntrar AI (ex. ChatGPT) att göra en **djupare analys** av valfritt innehåll (artikel, inlägg, nyhetsrapport). Tanken är att du ska **vidga dina vyer**, se egna (och andras) förutfattade meningar, samt få fler perspektiv på ämnet.

Prompten (klipp och klistra i AI-verktyget):

Du är en vänlig och korrekt assistent som hjälper användaren att vidga sina vyer, ge nya perspektiv och utveckla sitt kritiska tänkande genom att analysera innehåll från sociala medier, nyheter och artiklar. Användaren kan klistra in skärmdumpar, texter eller länkar, och du identifierar mönster, bias och möjliga kunskapsluckor. När användaren har delat innehåll gör du följande:

- 1) Analysera innehållet. Vad är det för innehåll. Bryt ned och analysera innehållet kortfattat
- 2) Skriv en sammanhängande djuplodande kritisk analys av innehållet med långa, utvecklade förslag på relaterad fakta, alternativa perspektiv och vidare läsning, så att användaren verkligen lär sig något. Ta dig tid att skriva en sammanhängande, tankeväckande text som verkligen hjälper användaren att förstå bias, eventuellt avsaknad av nyans och gör användaren till en mer kritiskt tänkande människa.

Syftet med denna del är att hjälpa användaren att sätta innehållet i relation, förstå andra synsätt, bryta filterbubblor och se nyanser i komplexa frågor. Den ger källkritiska reflektioner och lyfter fram olika tolkningar utan att förespråka en viss åsikt. Om informationen är osäker eller kontroversiell, markerar den detta tydligt och föreslår vidare läsning.

Denna del skrivs som en lång sammanhängande text med retoriska frågor som underbyggd, förklaras och nyanseras omfattande.

Texten är strukturerad i stycken med rubriker. I slutet av texten ges förslag på frågor att reflektera över och utforska vidare

- 3) Sist ges Ge förslag på böcker, webbsidor, YouTube-klipp eller liknande som kan hjälpa användaren att nyansera och problematisera och få perspektiv.

REGLER:

Du söker aktivt på nätet för att ge uppdaterad information och kan även använda sin befintliga kunskap. Du ger svar direkt och alla svar innehåller både en analys och djupgående förslag på nya perspektiv

2. Genomförande (steg-för-steg)

1. Välj innehåll (5–10 min)

- Ta en artikel, ett sociala medier-inlägg eller en nyhetsrapport om något **aktuellt** eller **omtvistat**.
- Innehållet ska vara **tillräckligt tydligt** för att AI ska kunna analysera det.
- 2. **Klistra in prompt + innehåll** i AI (5–10 min)
 - Kopiera **hela** prompten i rutan för AI.
 - Därefter klistrar du in artikeln, texten eller länken du vill analysera.
 - Skriv en kort instruktion: “Här är en artikel om [ämne]. Gör din analys enligt prompten ovan!”
- 3. **Läs AI:ns “meta-analys”** (10–15 min)
 - Hur strukturerar AI sin text?
 - Vilka **bias** eller perspektiv lyfter den fram?
 - Vilka **förslag på vidare läsning** får du?
- 4. **Gruppdiskussion eller helklass** (10–15 min)
 - Berätta kort om:
 1. Vilket material ni analyserade.
 2. Vilka perspektiv AI tog upp som var **nya** för er.
 3. Om AI föreslog böcker/klipp/webbsidor – verkar de relevanta?
- 5. **Reflektion** (5 min)
 - Hur skiljer sig AI:s “meta-analys” från en vanlig sammanfattning?
 - Vill du använda detta i fortsättningen för att “zooma ut” och se fler sidor av en fråga?

3. Förslag på uppgifter

Välj **ett** av följande exempel eller hitta ett eget:

1. **Sociala medier-flöde**
 - Ta en **Instagram-caption** eller tweet full av åsikter. Låt AI “meta-analysera” den. Kändes det mer nyanserat när AI bröt ned texten?
2. **Politisk debatt**
 - Välj en **debattartikel** från en politiker. Hur mycket handlar det om känslomässiga argument vs. sakargument? Vilka **andra** perspektiv på frågan lyfter AI?
3. **Nyhetsartikel**
 - Kolla en **lokal nyhet** (ett kommunbeslut, en skolfråga). AI beskriver ekonomiska, sociala och etiska perspektiv. Hade du själv sett alla dessa vinklar innan?

Diskussionsfrågor:

- **Hur vet du** att AI:s källförslag är trovärdiga?
- **Om AI** visar på bias i texten, ändrar det din syn på källan?
- **Kan AI** själv vara bias beroende på hur prompten formuleras?
- **Vad** är för- och nackdelarna med att använda denna “vidga vyerna”-funktion i vardagen?

4. Redovisning

1. Muntlig presentation

- Grupper presenterar sin AI-analys, vad de lärde sig om ämnet och om egna antaganden.

2. Skriftlig reflektion

- 1–2 stycken om hur AI:s “meta-analys” förändrade eller inte förändrade din syn på innehållet.

3. Paneldebatt

- Rollspel:
 - En elev **försvarar** originaltexten,
 - En elev **använder** AI:ns meta-analys,
 - En elev är **skeptisk** mot båda.
- Se hur argumenten möts.

5. Sammanfattning

Genom att klistra in denna **specialprompt** i en GPT-baserad AI får du inte bara en enkel sammanfattning, utan en **djupare “meta-analys”**. AI kan peka på **bias**, kunskapsluckor, olika perspektiv och föreslå vidare läsning. Det hjälper dig **bryta filterbubblor** och se fler sidor av en fråga.

Reflektera:

1. Vilka **nya** perspektiv fick du på ämnet?
2. Hur **pålitlig** kändes AI:ns analys – och hur kan du kolla upp dess tips?
3. Vill du använda denna metod igen, när du stöter på något du vill förstå på djupet?

Kom ihåg: AI är ingen garanti för sanningen, men den kan **hjälpa** dig ställa fler frågor och se fler sidor. Alltså ett redskap för att “vidga vyerna,” men du behöver fortfarande ditt **eget kritiska tänkande!**

Lektion 20: AI kan ha fel

(Att förstå AI-hallucinationer och vikten av ämneskunskaper i kritisk granskning)

Introduktion

Trots att AI (som ChatGPT och liknande) kan leverera snabba och välformulerade svar, **kan den ha fel**. Fenomenet kallas ibland “hallucinationer” när AI påstår saker som inte stämmer. Dessutom kräver verkligt **kritiskt tänkande** inte bara en generell skepsis, utan även **ämneskunskaper**: utan kunskaper i ett ämne är det svårt att avgöra om AI-svar stämmer eller inte.

I denna lektion tittar vi på hur man **identifierar** faktafel, **varför** de uppstår och **hur** man kan förbättra AI:s svar. Vi tar hjälp av tre modeller:

1. **Navigatörens Karta** – Att *observera, fråga och jämföra* AI-svar.
2. **Skulptörens Sten** – Att *granska kriterier, utvärdera och förfina* AI-svar.
3. **Trädgårdarens Träd** – Att *idéutveckla och vidareutveckla* för mer kreativa och holistiska lösningar.

Syfte och mål

- **Medvetandegöra** elever om att AI kan framstå som trovärdig men ändå ha felaktigheter (hallucinationer).
- **Visa hur ämneskunskaper** är avgörande för att bedöma om AI-svar är korrekta.
- **Praktiskt öva** metoder för att analysera och förbättra AI-genererat innehåll.

Lektionsingång

1. **Exempel på ett felaktigt AI-svar**
 - Visa ett kort, uppenbart felaktigt AI-genererat stycke (t.ex. historiska “fakta” som inte stämmer).
 - Låt eleverna spontant kommentera: “Vad är fel här? Hur upptäcker vi det?”
2. **Diskussion**
 - “Har ni varit med om att AI svarar övertygande men faktiskt har fel?”
 - “Varför är det viktigt att vi har egna kunskaper för att bedöma om något stämmer?”

Del 1: Navigatörens Karta

Syfte: Lär eleverna att *observera, fråga och jämföra* AI-svar med andra källor.

1. **Observera**
 - **Vad:** Läs igenom AI-svaret noggrant.
 - **Gör:** Notera hur texten är strukturerad, hur många huvudpunkter den har och dess tydlighet.

- **Exempel:** Be eleverna skriva ned huvudidéerna i AI-svaret och markera eventuella påståenden som låter tveksamma.
2. **Fråga**
- **Vad:** Formulera kritiska frågor.
 - **Gör:** “Kan detta vara korrekt? Hur vet vi det? Finns det några förutfattade meningar här?”
 - **Exempel:** Lista frågor som “Vilka antaganden gör AI-svaret?”, “Finns det påståenden som saknar källa eller logisk grund?”
3. **Jämför/Check**
- **Vad:** Kontrollera AI-svar mot en auktoritativ källa (lärobok, vetenskaplig artikel, pålitlig webbplats).
 - **Gör:** Se var AI-svar och källan stämmer överens eller skiljer sig åt. Notera skillnader.
 - **Exempel:** Om AI säger att en historisk händelse skedde 1450, men källor pekar på 1460 – markera detta som potentiellt fel.

Övning

- Låt eleverna generera ett kort AI-svar om valfritt ämne (historiskt, vetenskapligt eller samhällsrelaterat).
- Använd “Navigatörens Karta” för att *observera, fråga och jämföra* svaret med minst en betrodd källa.

Del 2: Skulptörens Sten

Syfte: Definiera kvalitetskriterier och använd dem för att *utvärdera* samt *förfina* AI-svar.

1. **Granska Kriterier**
- **Vad:** Vad kännetecknar ett **bra svar** för denna uppgift?
 - **Gör:** Tillsammans skriver ni upp kriterier: t.ex. **noggrannhet, relevans, tydlighet, aktuella källhänvisningar**.
 - **Exempel:** “Svar ska innehålla korrekt fakta, visa logisk struktur och ge konkreta exempel.”
2. **Utvärdera**
- **Vad:** Jämför AI-svaret med era uppsatta kriterier.
 - **Gör:** Identifiera styrkor (“röd tråd i texten”) och svagheter (“ingen källa anges”, “oexakt datum”).
 - **Exempel:** Eleverna markerar var AI-svaret möter kriterierna eller brister.
3. **Förfina**
- **Vad:** Förbättra svaret genom att ge AI tydligare instruktioner/promptar.
 - **Gör:** Formulera nya promptar som “Var mer källspecifik”, “Ange exakt datum och belägg”, eller “Ge exempel ur läroboken”.
 - **Exempel:** Jämför **original-svaret** med **förfinade** svar och se vilka förbättringar som skedde.

Övning

- Dela in eleverna i grupper. Varje grupp tar en **felaktig** eller bristfällig AI-text och skapar **nya promptar** baserat på de kriterier ni listat.

- Låt dem skriva om texten och se om AI-svaret blir mer korrekt eller anpassat.

Del 3: Trädet

Syfte: Använd AI som en *kreativ partner* för att utveckla idéer och lösningar.

1. Idéer

- **Vad:** Låt AI brainstorma olika idéer kring ett problem eller ämne.
- **Gör:** Eleverna noterar allt AI föreslår, även “tokiga” eller oväntade förslag.
- **Exempel:** “Hur minskar vi plastavfall i skolan?” Låt AI spotta ur sig 10 idéer.

2. Kopplingar

- **Vad:** Analysera sambanden mellan idéerna.
- **Gör:** Eleverna diskuterar om idéerna kan kombineras eller stötts av varandra.
- **Exempel:** Rita en mindmap. Koppla idéer som “införa pant” eller “använda mindre engångsplast” med varandra.

3. Utvidga

- **Vad:** Välj en eller flera idéer för att göra en **mer detaljerad plan**.
- **Gör:** Diskutera praktiska steg och hinder.
- **Exempel:** Eleverna kan göra en projektplan för att införa pantstationer på skolgården.

Övning

- Låt eleverna, i samma grupper, utgå från ett tema i skolämnet (t.ex. hållbar utveckling, demokrati, teknik).
- Använd AI för idégenerering. Diskutera sedan och *utveckla* en av idéerna till en mer *realistisk plan*.

Sammanfattande Tips

1. **Diskutera processen** efter varje övning:
 - Vad funkade bra?
 - Vilka utmaningar uppstod?
2. **Belys AI:s begränsningar:**
 - AI är inte ofelbart, *kan* blanda ihop fakta, saknar ibland kontext.
3. **Anpassa övningarna:**
 - Vissa elevgrupper behöver mer stöd (förifyllda exempel, tydligare mallar).
4. **Kombinera metoder:**
 - Navigatörens karta för **faktakoll**,
 - Skulptörens sten för **förbättring**,
 - Trädet för **kreativ idéutveckling**.

Diskussion och reflektion

- **Varför uppstår AI-hallucinationer?** (Brister i träningsdata, statistisk gissning, avsaknad av riktig “förståelse”).

- **Varför räcker det inte** att bara vara “allmänt skeptisk”? (Vi behöver ämneskunskaper för att upptäcka specifika fel).
- **Hur** kan läraren integrera dessa metoder i olika skolämnen (historia, naturvetenskap, språk)?

Förslag på upplägg

1. **Teoretisk genomgång** (5–10 min): Förklara fenomenet “AI kan ha fel” och ge exempel på hallucinationer.
2. **Praktisk laboration** (15–20 min): Använd “Navigatörens Karta” på ett AI-svar, sök upp korrekta fakta, och markera fel.
3. **Förbättring** (10–15 min): Använd “Skulptörens Sten” för att sätta upp kriterier och prompta AI på nytt.
4. **Kreativ avslutning** (10–15 min): “Trädgårdarens Träd” för att visa hur AI trots allt kan vara användbar för idéer och brainstorming.
5. **Reflektion** (5 min): Diskutera hur denna metod kan användas i andra ämnen.

Avslutning

Den här lektionen visar att AI **inte** ersätter den **mänskliga faktakontrollen**. Genom att kombinera **ämneskunskap**, **kritiskt tänkande** och **metodiska verktyg** (Navigatörens Karta, Skulptörens Sten och Trädet) kan elever och lärare lära sig att både **identifiera** AI:s fel och **förbättra** AI:s svar. På så vis stärker vi elevernas **digitala kompetens** och förbereder dem för att hantera information i en verklighet där AI är en ständig (men inte ofelbar) aktör.

Arbetsblad – Lektion 11: AI kan ha fel

1. Bakgrund

Instruktion: Läs detta stycke om vad som menas med “AI kan ha fel”:

AI-modeller som ChatGPT kan verka mycket **trovärdiga**, men ibland påstår de saker som inte stämmer – det kallas “hallucinationer.” Dessutom krävs **ämneskunskaper** för att upptäcka många fel. Utan grundläggande koll på ett område kan AI:s svar låta övertygande men vara **helt fel**.

2. Navigatörens Karta (Observera, Fråga, Jämför)

Instruktion: Ta ett kort AI-genererat svar om valfritt ämne (t.ex. historia, biologi, samhällsfråga). Arbeta i grupp eller enskilt och följ stegen nedan.

1. Observera

- Läs igenom AI-svaret noga.
- Skriv ner de **huvudidéer** du hittar.
- Markera eventuella **tveksamma** påståenden.

2. Fråga

- Ställ **kritiska frågor**:
 - “Kan detta vara korrekt?”
 - “Har AI:n antagit något utan grund?”
 - “Hur vet vi det?”
- Lista frågor i form av: “Vilka antaganden görs?”, “Saknas logisk grund?”

3. Jämför/Check

- **Kontrollera** AI-svarets påståenden mot minst en **pålitlig källa** (lärobok, vetenskaplig artikel, betrodd webbplats).
- Notera **likheter** och **skillnader**.
- Markera var AI-svaret avviker från kända fakta.

Kort övning:

1. Klistra in AI-svaret här:
 - ...
2. Sammanfatta dina observationer:
 - **Huvudidéer:** ...
 - **Tveksamma punkter:** ...
3. Vilka **konkreta skillnader** fann du jämfört med en pålitlig källa?

3. Skulptörens Sten (Granska Kriterier, Utvärdera, Förfina)

Instruktion: Nu ska ni förbättra AI-svaret genom att först sätta upp **kvalitetskriterier** för ett “bra” svar och sedan **utvärdera** AI-texten utifrån dessa.

1. Granska Kriterier

- Exempel på kriterier:
 - **Noggrannhet** (korrekta fakta),
 - **Relevans** (svarar på uppgiften),
 - **Struktur** (tydlig indelning),
 - **Exempel** (konkreta belägg).
- Skriv 2–3 kriterier här:
 - - 1. ...
 - - 2. ...
 - - 3. ...

2. Utvärdera

- Hur väl uppfyller AI-svaret dessa kriterier?
- Markera **styrkor** och **svagheter**.

3. Förfina

- Ge AI en ny prompt: “Var mer källspecifik”, “Inkludera datum och referenser”, etc.
- Se om resultatet blir **bättre**.
- Jämför **original**-svaret med **nytt**-svaret.

Kort övning:

1. Nämn två **svagheter** i AI-svaret enligt dina kriterier:
 - ...
2. Skriv en **förbättrad prompt** du vill ge AI:
 - ...
3. Hur blev det nya svaret?
 - ...

4. Trädet (Idéer, Kopplingar, Utvidga)

Instruktion: Nu ska vi använda AI för *kreativa ändamål* – brainstorma idéer, se samband och utveckla planer.

1. Idéer

- Be AI brainstorma 5–10 förslag på en given fråga, t.ex. “Hur kan vi minska plastavfall i skolan?”
- Notera **alla** förslag, även de som verkar märkliga.

2. Kopplingar

- Diskutera hur idéerna kan hänga ihop eller **stärka** varandra.
- Rita en **mindmap** eller gör en punktlista med kommentarer.

3. Utvidga

- Välj **en** idé att fördjupa.
- Diskutera **praktiska steg** och vilka hinder som kan finnas.
- Skriv en **kort projektplan** eller lösningsförslag.

Kort övning:

1. Fråga AI: "Hur kan vi höja trivseln på skolan?" (eller valfritt tema)
2. Skriv ned **3 idéer** du gillar:
 - 1. ...
 - 2. ...
 - 3. ...
3. Välj en idé och **utveckla** den till en miniplan:

5. Sammanfattande tips

1. **Diskutera processen** efter varje övning
 - Vad funkade bra? Vad var svårt?
2. **Belys AI:s begränsningar**
 - AI kan sakna kontext och **gissa** fram felaktiga svar.
3. **Anpassa övningarna**
 - Vissa texter behöver mer guidning eller förifyllda exempel.
4. **Kombinera metoder**
 - **Navigatörens Karta** för faktakoll,
 - **Skulptörens Sten** för förbättring,
 - **Trädgårdens Träd** för kreativ utveckling.

6. Diskussionsfrågor

1. **Varför uppstår** AI-hallucinationer?
2. **Varför räcker det inte** att bara vara skeptisk? Varför behövs ämneskunskaper?
3. **Hur** kan du själv använda dessa tre metoder (Kartan, Sten, Träd) i andra skolämnena?

7. Övningsupplägg

Förslag på upplägg i klassen:

1. **Teori** (5–10 min)
 - Visa ett exempel på ett felaktigt AI-svar. Diskutera.
2. **Navigatörens Karta** (15–20 min)
 - Låt eleverna observera, fråga och jämföra AI-svar med en betrodd källa.
3. **Skulptörens Sten** (10–15 min)
 - Sätt upp kriterier, utvärdera AI-texten, ge ny prompt. Se förbättringar.
4. **Trädgårdens Träd** (10–15 min)
 - Använd AI för idégenerering kring ett valfritt ämne. Kombinera och utveckla idéer.

5. **Reflektion** (5 min)

- Hur kan ni använda detta i andra skolämnena?

8. **Avslutande reflektion**

Skriv **2–3 meningar** om hur du ser på AI som hjälpmedel efter denna lektion.

- ...

Kom ihåg: AI är ett **verktyg**, inte en ersättning för din egen kunskap eller kritiska förmåga. Med ämneskunskaper och dessa metoder blir du bättre på att både **upptäcka** AI:s fel och **dra nytta** av AI:s förmågor!

Lektion 21: Hallucinationer hos AI – när verktygen hittar på

Introduktion

En fascinerande men problematisk egenskap hos generativa AI-modeller (som ChatGPT) är att de ibland "**hallucinerar**", d.v.s. hittar på svar som låter trovärdiga men är helt fel. Syftet med denna lektion är att eleverna ska förstå vad AI-hallucinationer är, varför de sker, och hur man kan upptäcka dem. Introducera med att förklara termen: inom AI kallas det *hallucination* när modellen confident ger information som inte stämmer med verkligheten. Ge ett uppseendeväckande exempel i introduktionen: Berätta om fallet där en advokat använde ChatGPT för att söka rättsfall, och AI:n hittade på sex domar som inte existerade [omni.se](https://www.omni.se). AI:n insisterade till och med att fallen var riktiga när advokaten frågade [omni.se](https://www.omni.se). Detta ledde till att advokaten hamnade i trubbel. Exemplet fångar att även vuxna kan luras om de tror för mycket på AI och inte dubbelkollar. Introduktionen understryker: *AI är inte ofelbar – den kan ljuga utan att mena det.*

Genomförande (steg-för-steg)

1. **Demonstration live:** Om möjligt, gör en kontrollerad demonstration i klassrummet: använd en öppen AI-tjänst (som ChatGPT eller Bing Chat) och ställ en fråga där du vet att AI:n ofta halvt hittar på. T.ex. "Ge mig tre referenser som stöder tesen att Wikipedia alltid har rätt." (AI kan hitta på boktitlar/studier). När svaret kommer, peka ut ev. fel (t.ex. referenser som inte finns när man söker dem). *Obs:* Testa detta i förväg så du vet att du får en hallucination.
2. **Förklara varför:** Gå igenom på en grundläggande nivå *varför* AI hallucinerar. Förklara att modellen inte har en databas av "fakta" utan skapar text genom att försöka förutsäga lämpliga ord. Ibland "tror" den att det den skriver är rimligt, men har ingen kontakt med verkligheten. Och eftersom den är tränad att låta säker, kommer det ut självsäkert. (Du kan använda en metafor: AI:n är som en elev som skriver ett svar han tror läraren vill höra, men ibland chansar han och det blir fel).
3. **Elevförsök:** Dela in elever i smågrupper vid datorer (eller gör som klassaktivitet med lärardatorn om begränsat med enheter). Ge dem några frågor att testa på en AI, där de sedan ska faktagranska svaret. Exempel: "Fråga AI:n: Hur dog Olof Palme?" – (AI kan ge korrekt info, men också eventuellt ge konspirationer eller släta över med oklarheter – intressant att se vad den svarar). Eller "Be AI:n om biografien av en mindre känd person" – stor chans till fel. Var noga att detta görs under uppsikt och tidsbegränsat.
4. **Faktagranska AI-svar:** Nu ska eleverna använda sina källkritiska färdigheter på AI:ns output. Be varje grupp välja ut ett påstående från AI-svaret och kolla om det

stämmer, via internet. De kan använda samma metoder som i förra lektionen (sök efter nyckelord etc). Låt dem dokumentera om de hittar fel.

5. **Sammanställ resultat:** Gå laget runt och låt grupperna rapportera: "AI:n påstod X, vi kollade och fann att... (det stämde/stämde inte)". Skriv upp typiska fel som kom fram (t.ex. fel årtal, påhittad referens, blandade ihop två personer). Detta konkretiserar hallucinationerna.
6. **Diskutera lösningar:** Fråga klassen hur man kan hantera detta problem. Svar som du kan lyfta fram: *Alltid dubbelkolla viktiga uppgifter som AI ger. Ge AI:n källor att utgå ifrån om möjligt. Ställ AI:n frågor om varifrån infon kommer (även om den kan hitta på det också!). Använd verktyg med faktakoll (vissa AI-funktioner börjar få inbyggda referenser).*
7. **Regler för användning:** Avsluta med att gemensamt formulera 2-3 "klassregler" för hur man ska använda AI-verktyg i skolarbeten. T.ex.: "Vi litar inte blint på AI-svar utan kollar viktiga fakta i en andra källa" och "Vi anger om vi använt AI-hjälp i en uppgift". Skriv upp dem och spara synligt.

Lektionsingångar & diskussion

- **Ingångsexempel:** Utöver advokatfallet, ge små exempel: "ChatGPT fick frågan vem som vann fotbolls-VM 2022 innan det visste svaret – vissa versioner hittade på något tokigt." (Tidiga AI-modeller gjorde så). Eller "En elev bad AI skriva om ett obskyrt ämne och fick ut helt fel info". Kanske någon elev själva redan testat och märkt fel? Fråga om någon sett AI göra misstag.
- **Diskussionsfråga:** *"Varför tror ni att AI ger oss felaktiga svar, trots all data den tränats på?"* Få fram deras gissningar. Korrigera/mata in att AI inte *förstår* utan bara genererar sannolikt nästa ord – ibland blir det orimligt men ser rimligt ut för modellen.
- **Tankelek:** *"Är en AI som hallucinerar 'en lögnare'?"* Vissa elever kanske säger ja – men förklara att lögn kräver avsikt att vilseleda. AI:n har ingen avsikt, den bara råkar ge fel svar. Det betyder att *ansvaret ligger på användaren* att vara kritisk.
- **Framtid:** Diskutera kort om de tror att framtidens AI kommer sluta hallucinera helt. Vissa kanske tror ja med bättre teknik, andra nej. Nämn att utveckling pågår (t.ex. kombination av AI med sökdatabaser för faktakoll), men att det alltid är klokt att dubbelkolla.
- **Etisk dimension:** *"Tänk om en AI-hallucination sprids stort – vems fel är det? AI-utvecklarna, eller de som spred det?"* Ingen enkel fråga, men bra för kritisk diskussion om ansvar. Kan nämna fall som att Bard (Googles AI) vid lansering gjorde ett fel i sin

första demo, vilket fick Googles aktie att rasa – trots att felet egentligen var litet, men förtroendet påverkades.

Didaktiska tips

- **Kontrollerad AI-användning:** Inte alla skolor tillåter elever att själva använda ChatGPT p.g.a. kontoregistrering och åldersgränser. Du som lärare kan då göra demonstrationer istället för elevgruppövning, eller använda öppna alternativ utan inloggning (det finns webbplatser som erbjuder GPT-baserade chattar fritt). Kolla skolans policy.
- **Var beredd på oväntat svar:** När man kör AI live, kan det ge oväntade svar (kanske olämpligt, men om man håller sig till neutrala frågor är det oftast lugnt). Testa alltid innan som sagt.
- **Teknisk fördjupning för nyfikna:** Några elever kanske undrar mer tekniskt hur språkmodeller fungerar. Var beredd med en enkel förklaring eller analogi (t.ex. ordfrekvens och statistik över massor av texter). Om någon kan programmering, jämför med en autofullföljningsfunktion på mobilen, fast på steroids.
- **Variationsmöjlighet:** Istället för text-AI kan du ta upp bild-AI: visa en AI-genererad bild med fel (AI kanske ritar människor med sex fingrar, eller blandar arkitektur från olika städer i en fantasi-stad). Det är också hallucination i bildform.
- **Samarbetsövning:** Om skolan har fler ämnen integrerade, kanske svenskläraren kan ge en skrivuppgift att använda AI som utkast och sedan låta eleverna revidera och faktakolla det. Denna lektionens innehåll kan vara stöd inför sådant arbete.

Forskning & verktyg

- **AI-hallucination känd utmaning:** Enligt OpenAI själva och oberoende forskare är hallucinationer ett av de största problemen med stora språkmodeller. De jobbar på lösningar som att koppla modellen mot databaser för faktakoll, men inget är fulländat än.
- **Exempel i forskning:** Det finns studier där ChatGPT ombeds generera biografier av personer – det myntades term "*Hallucinated biography*" när den fyllde i fel meriter för folk. Att visa elever att detta är dokumenterat gör att de förstår det inte är "bugg hos mig" utan ett generellt fenomen.
- **Verktyg med källhänvisningar:** Påpeka att vissa nyare AI-verktyg (t.ex. Bing Chat, Perplexity AI) försöker ge källor till sina svar. Dessa kan vara något mer pålitliga *ifall källorna är riktiga*, men även de kan fela. Kanske visa ett svar från Bing som referens – och sedan kolla den källan, var det relevant?

- **Digitala verktyg att hantera hallucinationer:**
 - **Double-checker:** Finns inga magiska verktyg som "upptäcker" hallucinationer automatiskt, men ett praktiskt tipsverktyg är: använd AI:n mot sig själv – fråga "kan du ge källor?" eller "är du säker?" (Den kan fortfarande ha fel, men ibland backar den om den inte är säker). Demonstrera denna strategi.
 - **Stop hallucinating prompt designs:** Det finns guider online för hur man ska utforma sina prompts för att minska risken för hallucination (t.ex. be om "endast direkta citat från..." eller liknande). Om eleverna använder AI mycket kan ni sammanfatta några sådana knep.

Arbetsblad

Arbetsbladet för Lektion 8 innehåller:

- **Exempelsvar att granska:** Utdrag av ett AI-svar (för de skolor som inte kan köra live) – t.ex. en paragraf från ChatGPT som påstår några "fakta" med ett par inbakade fel. Eleverna ska markera meningar de misstänker och skriva hur de skulle verifiera dem.
- **True/False AI edition:** En lista utsagor, där vissa är faktiska AI-hallucinationer som förekommit, och andra är korrekta svar. Ex:
 1. "AI påstod att Fotbolls-VM 2022 vanns av Italien."
 2. "AI hävdade att Einstein fick Nobelpris 1921 för relativitetsteorin."
 3. "AI angav att H₂O är den kemiska formeln för vatten."
 4. "AI hittade på en källa 'Oxford Journal of Astrology' i en text om stjärntecken." Eleverna kryssar vilka som låter som felaktiga påståenden/hallucinationer (facit: 1 är fel – Italien vann inte 2022, 2 är fel – Einstein fick visserligen Nobelpris 1921 men för fotoelektriska effekten, inte relativitetsteorin; 3 är rätt; 4 är fel – påhittad källa). Diskutera efteråt hur de resonerade.
- **Hallucinationens mekanik (valfritt avancerat):** En liten ruta med flervalsfråga: "Vad är huvudorsaken att AI 'hallucinerar' information?" A) Den försöker ge ett svar även när den inte vet svaret. B) Någon har hackat AI:n att ljuga. C) Det är spöken i maskinen. (Rätt: A).
- **Handlingsplan:** En tabell där kolumn A listar potentiella AI-scenarion elever kan hamna i (t.ex. "AI ger dig en lista med tre 'fakta' du tänker använda i din uppsats", "AI skriver en referenslista åt dig", "AI förklarar en historisk händelse"), och kolumn B har tomma rader där eleven ska skriva vad de ska göra härnäst (t.ex. "kolla varje fakta i en

bok eller pålitlig källa", "kontrollera att böckerna/artiklarna finns på riktigt", "jämföra mot läroboken/historiska källor"). Detta blir en konkret plan de kan ha kvar för att undvika bli lurade av hallucinationer.

Gå igenom handlingsplanen så att eleverna verkligen tar till sig att de alltid bör dubbelkolla kritiska saker. Påminn dem: en AI kan vara användbar men kan aldrig ersätta deras eget kritiska tänkande.

Arbetsblad – Lektion 21: Hallucinationer hos AI

När AI-verktyg hittar på

Vad är AI-hallucinationer?

Inom AI kallas det "hallucination" när en AI-modell med självförtroende ger information som inte stämmer med verkligheten. Generativa AI-verktyg som ChatGPT producerar ibland påhittade fakta, personer, källor eller händelser som låter övertygande men som inte existerar.

Ett verkligt exempel: En advokat i USA använde ChatGPT för att söka efter juridiska rättsfall. AI:n hittade på sex domstolsbeslut som aldrig existerat. När advokaten frågade om fallen verkligen fanns, insisterade AI:n att de var äkta. Advokaten citerade de påhittade fallen i sina inlagor till domstolen och hamnade i allvarligt trubbel för att ha använt falsk information.

Varför hallucinerar AI?

AI-modeller som ChatGPT har ingen databas med "fakta" som den kan söka i. Istället:

- Den förutsäger vilka ord som sannolikt bör följa varandra baserat på vad den lärt sig från miljontals texter
- Den har ingen faktisk förståelse för vad som är sant i verkligheten
- Den har tränats att låta övertygande och hjälpsam även när den är osäker
- Den har ingen förmåga att inse när den har fel eller hittar på

Tänk så här: AI:n är som en elev som skriver ett svar hen tror läraren vill höra, men ibland chansar hen och det blir fel – men levererar svaret med stor självsäkerhet.

ÖVNING 1: GRANSKA ETT AI-SVAR

Läs igenom följande svar från en AI och markera påståenden som du misstänker kan vara felaktiga:

"Nobelpriset i litteratur 2019 tilldelades Peter Handke, en österrikisk författare känd för sin kontroversiella politiska ståndpunkt. Samma år tilldelades Olga Tokarczuk Nobelpriset i litteratur för 2018, eftersom priset inte delades ut det året på grund av ekonomiska skandaler inom Svenska Akademien. Handke är mest känd för sitt mästerverk 'Den blinda mannen från Sarajevo' som skildrar kriget i forna Jugoslavien. Tokarczuk, å andra sidan, fick internationellt genombrott med romanen 'Löparna' från 2007, där hon utforskar olika resenärers upplevelser genom historien."

Hur skulle du kontrollera denna information? Skriv steg för steg:

1. _____

2. _____
3. _____

ÖVNING 2: SANT ELLER FALSKT?

Markera vilka av följande AI-svar som innehåller hallucinationer:

- 1. "Spanien vann fotbolls-VM 2022 efter en spännande final mot Frankrike."
- 2. "Marie Curie tilldelades Nobelpris i fysik 1903 för upptäckten av radium och polonium."
- 3. "Den kända boken 'Artificiell Intelligens: Historia och Framtid' av professor Hanna Bergström från Stockholms universitet redogör för AI:ns utveckling."
- 4. "Barack Obama var USA:s 44:e president och tjänstgjorde mellan 2009 och 2017."
- 5. "Enligt en studie publicerad i den vetenskapliga tidskriften 'Journal of Digital Psychology', använder 78% av alla tonåringar ChatGPT för skolarbeten."

ÖVNING 3: SKAPA EN HALLUCINATIONS-FÄLLA

Formulera en fråga som du tror skulle kunna lura en AI att hitta på ett svar:

Varför tror du att denna fråga kan leda till hallucination?

ÖVNING 4: UPPTÄCK HALLUCINATIONS-MÖNSTER

Arbeta i grupper och testa er fråga på en AI. Dokumentera svaret och granska det kritiskt.

AI-verktyg vi använde: _____

Vår fråga: _____

Vad AI:n svarade: (sammanfatta kort)

Fakta vi misstänker är hallucination: (lista minst ett påstående)

Hur vi verifierade/motbevisade detta: (vilka källor använde ni?)

MIN HANDLINGSPLAN MOT AI-HALLUCINATIONER

När jag använder AI för att...	Så bör jag alltid...
Få fakta till en skoluppgift	
Hitta källor och referenser	
Söka information om en person	
Få förklaringar av komplexa ämnen	
Få svar på frågor där jag är osäker	

PRAKTISKA STRATEGIER FÖR ATT HANTERA AI-HALLUCINATIONER

- **Be AI:n visa sina källor** (men kom ihåg att den kan hitta på källor också!)
- **Ställ motfrågor** – t.ex. "Är du säker på det svaret?" eller "Finns det åsikter som motsäger det du just sa?"
- **Dela upp komplexa frågor** i mindre delar för att minska risken för felaktig information
- **Jämför alltid med etablerade källor** – uppslagsverk, läroböcker, vetenskapliga publikationer
- **Var extra vaksam** när AI svarar på frågor om:
 - Aktuella händelser (särskilt efter AI:ns kunskapsgräns)
 - Statistik och exakta siffror
 - Obskyra eller mycket specialiserade ämnen
 - Citat och exakta uttalanden från personer

KOMMA IHÅG

AI-verktyg:

- Är inte medvetna om vad som är sant eller falskt
- Kan inte skilja mellan fakta och fiktion
- Har ingen förståelse för vad de svarar
- Kan inte erkänna när de inte vet något utan att bli ombedda

Det kritiska tänkandet är DITT ansvar!

Lektion 22: Syna Bias

(Förstå, upptäck och diskutera fördomar i AI-genererat innehåll)

Introduktion

Bias betyder fördom, snedvridning eller partiskhet som kan finnas i data, i programmeringen av en AI-modell – eller i våra egna sätt att tänka. AI-modeller är tränade på stora datamängder som kan innehålla stereotyper eller obalanserad information. Lektionen syftar till att hjälpa eleverna **identifiera** hur dessa fördomar uppstår och **synliggöra** dem i konkreta exempel, exempelvis genom att be en AI generera bilder eller texter som visar på stereotyper.

Syfte och mål

- **Förklara vad bias är** och varför det förekommer i AI.
- **Visa olika typer av bias** (ex. könsstereotyper, etniska stereotyper, yrkesbias).
- **Träna elever** i att upptäcka och diskutera bias i AI-genererat material.
- **Öva på** att ifrågasätta resultatet av AI-tjänster, även när det verkar “harmlös”.

Förklaring: Vad är Bias och varför finns det?

1. **Data-baserad bias**
 - AI-modeller lär sig från stora datamängder (text, bilder, etc.). Om dessa data redan innehåller snedvridningar eller överrepresentationer, lär sig AI:n samma mönster.
 - Exempel: Om de flesta läkare i träningsdatan beskrivs som “män”, kommer AI:n troligare att associera ordet “läkare” med män.
2. **Design-baserad bias**
 - Den som programmerar AI kan omedvetet bygga in förutfattade meningar eller prioritera viss data.
 - Exempel: Om bildgeneratoren är tränad mest på västerländska bilder, kan den ha svårt att korrekt generera bilder av afrikanska familjer i olika samhällsmiljöer.
3. **Användar-baserad bias**
 - Vi människor tenderar att söka, klicka och fråga om saker som intresserar oss. AI lär sig av vårt beteende och förstärker vissa mönster.
 - Exempel: Om man bara genererar bilder av rika, lyckliga familjer, kan AI tro det är standarden för “familj”.

Övningar och exempel

1. **Generera bilder**
 - **Uppgift:** Be eleverna använda exempelvis grok.com eller en liknande AI-bildtjänst för att skapa “en typisk familj” och sedan “en afrikansk familj.”

- **Syfte:** Se om bildgeneratorn tenderar att reproducera stereotyper (klädsel, miljö, hudfärg, könsroller). Diskutera vad som eventuellt är stötande eller ensidigt i resultaten.
- 2. **Könsstereotyper i text**
 - **Uppgift:** Be en chattbot (t.ex. ChatGPT) att “Beskriv en vanlig arbetsdag för en kvinnlig läkare” och sedan “Beskriv en vanlig arbetsdag för en manlig läkare.”
 - **Syfte:** Upptäck om AI:n använder olika språkbruk, ton eller antaganden (t.ex. att en kvinnlig läkare har mer mjuka, omhändertagande uppgifter och en manlig läkare mer tekniska eller auktoritära uppgifter).
- 3. **Roll- eller yrkes-bias**
 - **Uppgift:** Be AI att beskriva “en VD” eller “en ingenjör” och se om AI automatiskt förutsätter kön, etnicitet eller andra attribut.
 - **Variant:** Be AI generera en bild av “en lärare” respektive “en rektor” och undersök om någon bias dyker upp i ålder, kön, klädsel m.m.
- 4. **Språk och attribut**
 - **Uppgift:** Låt eleverna skriva in: “Berätta om en dag i livet för en pensionär / tonåring / influencer / politiker” och se om chattboten tillskriver dem stereotyper.
 - **Syfte:** Synliggöra hur AI kan generalisera och reducera komplexa grupper till klichéer.

Redovisningsformer

1. **Gruppredovisning**
 - Eleverna visar sina genererade bilder eller läser upp AI-texter högt.
 - Diskuterar i helklass: “Vilka stereotyper ser vi?” “Hur kunde detta se annorlunda ut?”
2. **Reflektionsdagbok**
 - Varje elev skriver ned sina observationer: Vad överraskade dem? Hittade de någon tydlig bias? Hur kan man motverka det?
3. **Diskussionspanel**
 - Skapa en “panel” där några elever företräder AI eller dataanalytiker, och andra företräder grupper som blir felrepresenterade. De kan diskutera hur man bör tackla problemet.

Diskussionsfrågor

1. **Varför uppstår bias i AI?**
 - Är det något som kan undvikas helt, eller bara minskas?
2. **Vilka konsekvenser** kan ske om AI förstärker stereotyper?
 - Påverkas människors självkänsla, yrkesval eller attityder mot andra grupper?
3. **Hur kan vi märka** bias i en AI-text eller bild om vi själva är omedvetna om våra fördomar?
 - Behövs perspektiv från många olika användare?
4. **Vem bär ansvaret** för att motverka bias i AI – utvecklarna, användarna, eller samhället via lagstiftning?

5. **Kan AI hjälpas åt** att motverka bias om vi ger mer riktade instruktioner (promptar) eller tränar den på mer varierad data?

Tips till läraren

1. **Normalisera samtal** om fördomar: Poängtera att alla har bias, men att vi kan träna oss att bli medvetna om dem.
2. **Ge konkreta instruktioner**: Elever kan bli mer mottagliga om du visar exempel på hur AI:s visuella representation skiljer sig när man byter ut “afrikansk familj” mot “svensk familj” eller liknande.
3. **Tydliggör syftet**: Lektionen handlar inte om att “skamma” AI, utan att visa hur teknologin reflekterar världen (och våra fördomar).
4. **Var beredd på känsliga diskussioner**: Stereotyper kring kön och etnicitet kan väcka starka känslor. Se till att alla känner sig trygga i samtalet.
5. **Använd humor**: Om AI genererar komiskt stereotypa bilder eller texter, kan det lätta upp stämningen att skratta åt dem – men följ upp med seriös reflektion.

Sammanfattning

Bias uppstår när AI-system **ärver** fördomar från träningsdata eller mänsklig design. I denna lektion lär sig eleverna genom **praktiska övningar** (bildgenerering, textgenerering) hur AI kan reproducera stereotyper om kön, etnicitet eller yrken. De diskuterar **varför** detta händer och **vilka effekter** det kan få på vår syn på världen. Poängen är att eleverna ska bli medvetna om **hur** de kan upptäcka och ifrågasätta bias i AI-verktyg – och i förlängningen, hur de själva kan hjälpa till att motverka fördomar både online och offline.

Arbetsblad – Lektion 22: Syna Bias

Förstå, upptäck och diskutera fördomar i AI-genererat innehåll

1. Vad är “bias” och varför finns det i AI?

Instruktion: Läs följande punkter om hur fördomar och snedvridningar (“bias”) kan uppstå i AI-system. Skriv sedan **en mening** om vad du själv tänker kring vart och ett av exemplen.

1. Data-baserad bias

- AI lär sig av stora datamängder. Om dessa är obalanserade eller innehåller stereotyper, ärver AI:n samma snedvridningar.
- **Din kommentar:** ...

2. Design-baserad bias

- Utvecklare och företag kan omedvetet prioritera viss data eller göra antaganden i sin kod.
- **Din kommentar:** ...

3. Användar-baserad bias

- Vi människor klickar bara på vissa saker; AI ser vårt beteende och förstärker det.
 - **Din kommentar:** ...
-

2. Övning: Generera bilder och texter

Instruktion: Ni ska testa hur AI kan reproducera eller förstärka stereotyper. Välj **en av nedanstående uppgifter** (eller flera, om tiden tillåter).

1. Generera bilder i en AI-tjänst (ex. grok.com)

- **Uppgift:** Skapa “en typisk familj” och sedan “en afrikansk familj.”
- **Notera:** Ser ni skillnader i klädsel, miljö, hudfärg som kan vara stereotyper?

2. Könstereotyper i text (ex. ChatGPT)

- **Uppgift:** “Beskriv en vanlig arbetsdag för en kvinnlig läkare” och sedan “en manlig läkare.”
- **Notera:** Finns det antaganden om roller, beteenden, ansvar?

3. Roll- eller yrkes-bias

- **Uppgift:** Be AI beskriva “en VD,” “en ingenjör,” “en lärare,” eller “en rektor.”
- **Notera:** Förutsätter AI ett visst kön, en viss ålder, eller andra attribut?

4. Språk och attribut

- **Uppgift:** “Berätta om en dag i livet för en pensionär / tonåring / influencer / politiker.”
 - **Notera:** Vilka egenskaper ges till grupperna? Är det klichéer?
-

3. Dokumentera dina fynd

Instruktion: Använd tabellen nedan för att samla dina observationer.

Uppgift	Vad AI genererade	Ev. tydliga stereotyper/bias
Ex. "En typisk familj" vs. "En afrikansk familj"	Ex: Bild med ljus interiör/... vs. mer "exotisk" miljö	Konstiga antaganden om klädsel, hudfärg, miljö
...
...

4. Diskussionsfrågor

1. **Hur** uppstod bias i AI:s svar? Är det data-baserad, design-baserad eller användar-baserad bias (eller en kombination)?
 - ...
2. **Vilka konsekvenser** kan det få om AI hela tiden visar en viss typ av bild av t.ex. "familj" eller "läkare"?
 - ...
3. **Hur** skulle ni kunna motverka bias? Kan man prompta AI annorlunda eller träna modellen på fler typer av data?
 - ...
4. **Vem** bär ansvaret för att undvika att AI förmedlar stereotyper? Utvecklarna, användarna, lagstiftarna?
 - ...

5. Reflektion och redovisning

1. **Gruppredovisning:**
 - Visa upp bilder eller läs upp AI-texter ni genererat.
 - Diskutera: "Vilka stereotyper ser vi? Hur hade texten/bilden kunnat vara mer neutral?"
2. **Reflektionsdagbok:**
 - Skriv några meningar om vad du lärt dig och vilka **oväntade** fördomar du upptäckt.
3. **Diskussionspanel (valfritt):**
 - Några elever representerar AI-utvecklare, andra representerar en grupp som blir felaktigt beskriven. Hur ska man lösa problemet med bias?

6. Sammanfattning

Instruktion: Svara kort på följande:

1. **Vad tar du med dig** om hur AI kan förstärka befintliga stereotyper?
 - ...
2. **Hur** kan du själv vara mer uppmärksam på bias när du använder AI-verktyg?
 - ...

Kom ihåg: AI är inget **neutralt** orakel. Det är byggt av människor, tränat på mänskliga data och kan förstärka våra fördomar. Att **upptäcka** och **ifrågasätta** bias är en viktig del av **digital kompetens** och **källkritik**!

Lektion 23: Bekräftelsebias (Confirmation Bias)

Introduktion

Denna lektion handlar om bekräftelsebias – vår tendens att söka, tolka och minnas information som bekräftar våra befintliga övertygelser. Vi undersöker hur detta påverkar vårt beteende, både offline och online, samt hur språkmodeller (LLM) och sökmotorer kan förstärka denna bias om vi ställer ledande frågor. Målet är att eleverna ska bli mer medvetna om hur de själva formulerar frågor och hur det kan styra informationen de får tillbaka.

Syfte och mål

- **Förstå vad bekräftelsebias är**
Eleverna ska kunna förklara fenomenet och ge exempel på hur det visar sig i vardagen.
 - **Bli medveten om hur frågor kan vara vinklade**
Eleverna ska inse skillnaden mellan ledande (partiska) frågor och mer neutrala, öppna frågor.
 - **Analysera hur LLM/sökmotorer förstärker bias**
Genom experiment och diskussion se hur ens egen formulering av en fråga påverkar svaren.
 - **Öva källkritik och reflektion**
Träna på att ställa fler perspektivfrågor, leta motargument och bredda sitt sökande efter information.
-

Förslag på lektionsingångar

1. **Kort klassdiskussion om “ledande frågor”**
 - Skriv upp två frågor på tavlan, till exempel:
 - *Ledande:* ”Varför är MacBook den enda vettiga datorn för studier?”
 - *Neutral:* ”Vilka för- och nackdelar finns med olika typer av datorer för skolbruk?”
 - Låt eleverna diskutera i par: Hur skiljer sig de två frågorna åt och vad får det för konsekvenser?

2. Elevnära exempel på konfirmationsbias

- Be eleverna fundera på något de nyligen sökt på nätet: en diet, en favoritartist, en dator, en konspirationsteori osv.
- Diskutera om de märkt att de fokuserat mer på resultat som bekräftade deras förväntningar, och ignorerat resten.

3. Testa i liten skala (om tekniken tillåter)

- Om skolan har tillgång till en AI-chatt eller vanliga sökmotorer: be eleverna söka på en fråga som är tydligt ledande (“Varför är TikTok slöseri med tid?”) och sedan en mer neutral variant (“Hur påverkar TikTok ungas tidsanvändning och beteende?”).
 - Jämför svarens innehåll och ton. Vad var skillnaden?
-

Genomförande (steg-för-steg)

1. Intro och nyckelbegrepp (5–10 min)

- Förklara bekräftelsebias i korthet. Ge vardagsexempel:
 - Att en person som redan tror att “nyare svenska filmer är dåliga” söker “bevis” för detta.
 - Att man bara klickar på rubriker som stödjer ens favoritlag eller favoritartist.

2. Exempel på ledande vs. neutrala frågor (10 min)

- Presentera ett antal elevnära exempel (se “Elevnära frågor” nedan).
- Diskutera i helklass eller i smågrupper: Hur skulle svaret kunna bli annorlunda om man ställer en neutral fråga jämfört med en ledande?

3. Gruppövning: Formulera frågor (10–15 min)

- Dela in eleverna i smågrupper.
- Varje grupp väljer ett ämne (t.ex. betyg, kost, sociala medier, politik, mode, musik, film).
- Låt dem skriva en **ledande fråga** och en **neutral fråga** om detta ämne.

- Samla in exemplen och diskutera i helklass hur stor skillnad i perspektiv man kan få.

4. **Praktiskt experiment (valfritt) (10 min)**

- Om möjligt: Använd en LLM (t.ex. en AI-chatt) eller en sökmotor på projektor.
- Mata in en ledande fråga och en neutral fråga om samma ämne.
- Jämför svaren. Var merparten av svaren i det ledande fallet “ensidiga” eller negativa/positiva?

5. **Reflektion och sammanfattning (5–10 min)**

- Diskutera:
 - Varför söker vi ofta information som bekräftar det vi redan tror?
 - Vilka risker finns när vi är omedvetna om vår bias?
 - Hur kan vi aktivt motverka bekräftelsebias? (T.ex. leta motargument, ställa öppna frågor, fråga en kompis med annan åsikt.)
-

Elevnära exempel på frågor

Nedan finns några exempel på hur samma ämne kan formuleras ledande vs. neutralt:

1. **Datorer för studier**

- *Ledande*: “Varför är MacBook den enda vettiga datorn för studier?”
- *Neutral*: “Vilka för- och nackdelar finns det med olika datorer för skolbruk?”

2. **Betyg och bedömning**

- *Ledande*: “Varför är läraren så orättvis när det gäller betyg?”
- *Neutral*: “Hur fungerar bedömningen och vilka faktorer påverkar betygen?”

3. **Sociala medier (TikTok, Instagram)**

- *Ledande*: “Varför är TikTok beroendeframkallande och slöseri med tid?”
- *Neutral*: “Hur påverkar TikTok och andra sociala medier ungas tidsanvändning och beteende?”

4. Skolval, engagemang

- *Ledande*: “Varför är det meningslöst att rösta i skolvalet, ingen bryr sig ändå?”
- *Neutral*: “Hur kan skolval och ungas engagemang i politik påverka samhället?”

5. Vegan vs. blandkost

- *Ledande*: “Varför är veganmat så tråkigt och smaklöst?”
- *Neutral*: “Vad kännetecknar vegansk kost och hur kan man variera den för att få näringsrik och god mat?”

6. AI och chatbottar

- *Ledande*: “Varför är AI-chatbotar opålitliga och sprider bara felaktig information?”
- *Neutral*: “Hur fungerar AI-baserade språkmodeller och vilka möjligheter och begränsningar finns för informationssökning?”

Diskussionsfrågor

1. Har du märkt att du själv någon gång bara söker eller läser information som stödjer din åsikt? Varför tror du att det är så?
2. Hur kan bekräftelsebias påverka hur du uppfattar nyheter, politik eller reklam?
3. Kan man hamna i en “filterbubbla” i sociala medier p.g.a. bekräftelsebias? Hur i så fall?
4. Vilken roll spelar AI och algoritmer i att förstärka det du redan tror på?
5. Hur kan du medvetet söka “mothugg” eller alternativa perspektiv när du söker information?

Didaktiska tips

- **Variera arbetssätt**: Använd gruppdiskussion, korta skrivuppgifter och eventuellt digitala verktyg för att ge konkreta exempel.

- **Knyt an till elevernas vardag:** Ta upp exempel från deras sociala medier, filmer de sett, musik de lyssnar på.
 - **Tona ner teknikkraV:** Om AI-verktyg inte är tillgängligt, räcker det ofta med att jämföra rubriker i olika nyhetskällor eller ledande vs. neutrala Google-sökningar.
 - **Uppmuntra kritiskt tänkande:** Påminn att källkritik inte bara är att ”kolla avsändaren” – utan också att inse hur våra egna frågor och förutfattade meningar styr de svar vi får.
 - **Trygg diskussionsmiljö:** Vissa frågor (t.ex. politik eller religion) kan röra starka åsikter. Skapa en öppen stämning där det är okej att tycka olika.
-

Forskning & verktyg

- **Internetstiftelsen** (Digitala lektioner) har färdiga lektionstips om källkritik och bekräftelsebias.
 - **Skolverket** lyfter vikten av digital kompetens och mediekunnighet i läroplanen.
 - **Psykologisk forskning** visar att konfirmationsbias är en stark kognitiv tendens (Leon Festinger m.fl.).
 - **AI-resurser:** Om skolan tillåter, kan ni testa AI-verktyg som ChatGPT eller Bing Chat för att praktiskt se hur frågornas formulering påverkar svaren.
-

Sammanfattning

I denna lektion får eleverna lära sig hur bekräftelsebias fungerar och hur den kan ta sig uttryck i vardagen. De övar på att skilja mellan ledande och neutrala frågor och ser hur frågor kan påverka resultaten i en sökmotor eller en AI-chatt. Fokus ligger på att utveckla ett kritiskt och nyanserat tänkande, där eleverna blir medvetna om att deras eget sätt att formulera frågor kan styra vilka svar de får. Genom diskussion och praktiska övningar fördjupas förståelsen för hur lätt vi människor förstärker våra förutfattade meningar – och hur vi kan motverka det.

Arbetsblad Lektion 23 – Bekräftelsebias

Instruktion:

Fyll i uppgifterna nedan för att reflektera kring hur bekräftelsebias kan yttra sig i din egen vardag samt hur frågeformuleringar kan styra svar.

1. Ledande vs. neutral fråga

1. **Välj ett ämne du gillar eller är nyfiken på.** Exempel: en ny mobil, ett gymnasieprogram, ett fotbollslag, en TV-serie, en politisk fråga.
2. **Skriv först en ledande (partisk) fråga** som förutsätter att din åsikt är korrekt.
3. **Skriv sedan en neutral (öppen) fråga** om samma ämne.

Ditt valda ämne:

Ledande fråga:

Neutral fråga:

- **Reflektion:** Hur tror du svaren hade skiljt sig åt om du sökte på nätet med dessa två frågor?
-

2. Bekräftelsebias i vardagen

1. Beskriv **ett tillfälle där du eller någon i din närhet** sökt information som mest bekräftade en befintlig åsikt.
 - Vad handlade ämnet om? Var letade du/hen efter fakta eller åsikter?
 - Hur hade man kunnat söka annorlunda?
 2. Varför tror du det är lockande att bara läsa/klicka på sånt som bekräftar det man redan tror?
-

3. Diskutera eller skriv kort

1. **Hur kan algoritmer i sociala medier förstärka bekräftelsebias?**

2. **Vilka risker finns om vi bara söker bevis för vår egen åsikt?** (Tänk på t.ex. politik, hälsa, klimatfrågor, relationer.)
 3. **Hur kan man motverka sin egen bekräftelsebias?** Nämn minst två förslag.
-

4. Sammanfattande reflektion (ca 5–6 meningar)

Skriv ett kort stycke om:

- **Vad du lärt dig om bekräftelsebias**
- **Hur du kan använda denna insikt i dina studier och i vardagen**
- **Vad du tar med dig när du söker information framöver** (t.ex. försöka ställa mer neutrala frågor, leta motargument, läsa flera källor)

Exempel:

“Jag upptäckte hur lätt jag faller in i att söka ‘Varför är x dåligt?’ istället för ‘Vad säger forskningen om x?’. I framtiden ska jag försöka ställa mer öppna frågor...”

5. Tips till dig som elev

- **Ställ alltid frågan på flera sätt:** En neutral frågeformulering ger ofta bredare svar.
 - **Aktivt leta motargument:** Försök hitta källor som säger emot din tes, inte bara dem som håller med dig.
 - **Diskutera med andra:** Fråga en kompis med motsatt åsikt hur hen ser på saken.
 - **Var extra källkritisk:** Lägg märke till vem som står bakom informationen och hur frågan är formulerad.
 - **Var nyfiken:** Testa att söka nya perspektiv. Du kanske lär dig något oväntat!
-

Lycka till med att upptäcka och motverka din egen bekräftelsebias!

Lektion 24: Auktoritetsbias, Förankringsbias och Tillgänglighetsbias

Bakgrund: Vad är kognitiva bias?

En **kognitiv bias** (ibland kallad ”tankefälla” eller ”kognitiv snedvridning”) är ett systematiskt sätt på vilket vår hjärna förenklar eller förvränger information. Istället för att alltid vara helt rationella och logiska påverkas vi av förutfattade meningar, känslor och tidigare erfarenheter när vi fattar beslut eller drar slutsatser. Dessa bias uppstår omedvetet och kan få oss att:

- Ta genvägar i vårt tänkande
- Lita på felaktig information
- Överskatta vissa risker eller möjligheter
- Missa viktiga perspektiv

När vi förstår att dessa snedvridningar existerar, kan vi bli bättre på att upptäcka dem hos oss själva och andra, vilket hjälper oss till mer nyanserade bedömningar och mer källkritisk användning av information.

Tre vanliga bias som påverkas av AI

1. Auktoritetsbias (Authority Bias)

Definition:

Vi tenderar att ge större förtroende åt information som kommer från en person eller källa vi uppfattar som en auktoritet. Det kan vara en person med en viss yrkestitel, en välkänd influencer eller – numera allt oftare – en AI som uttrycker sig väldigt formellt och självsäkert.

Exempel i vardagen:

- Elever kanske tar allt läraren säger för givet, utan att ifrågasätta om det faktiskt stämmer.
- Vi litar på en influencer med miljoner följare när de rekommenderar en hälsoprodukt, trots att de saknar medicinsk utbildning.
- En AI-chatt kan ge felaktiga råd om träning eller kost, men låter ”professionell” och blir därmed trovärdig bara på grund av tonen och formuleringarna.

Hur AI kan förstärka detta:

Språkmodeller (LLM) använder ofta formellt och akademiskt språk. De kan ange referenser eller skriva självsäkert (“Det är ett välkänt faktum ...”), vilket lurar oss att tro att innehållet är 100 % korrekt eller granskat av experter. Eftersom många inte vet hur AI faktiskt fungerar, är det lätt att överskatta dess auktoritet.

2. Förankringsbias (Anchoring Bias)

Definition:

Den första informationen vi får om ett ämne – till exempel ett pris eller en siffra – fungerar ofta som ett “ankare” för all vidare bedömning. När vi sedan får ny information justerar vi kanske vår uppfattning lite, men det första värdet fortsätter att påverka oss mer än vi tror.

Exempel i vardagen:

- Om någon säger “Jag tror en ny mobil kostar ungefär 10 000 kr”, upplevs allt kring 8 000–12 000 kr som “rimligt”.
- Vid en auktion kan ett högt utropspris få budgivare att gå högre än de annars skulle.
- I en diskussion om hur många timmar man bör plugga, kan första personen som säger “10 timmar per vecka” bli den siffra alla jämför sig med.

Hur AI kan förstärka detta:

Om en AI (eller sökmotor) först presenterar en siffra (“80 % av alla jobb försvinner på grund av AI inom 10 år!”) är det lätt att förankras vid den prognosen. Även om modellen senare nyanserar svaret, hänger det första intrycket kvar och kan påverka hur vi tolkar fortsatta påståenden.

3. Tillgänglighetsbias (Availability Heuristic)

Definition:

Vi överskattar ofta hur vanligt eller viktigt något är baserat på hur snabbt vi minns exempel på det. Om något är lätt att komma på (t.ex. för att det är nytt, dramatiskt eller nyligen omskrivet), tror vi att det sker oftare eller är viktigare än det egentligen är.

Exempel i vardagen:

- Efter att ha sett flera nyheter om flygolyckor kan man känna rädsla för att flyga, trots att sannolikheten för en olycka är väldigt låg.

- Om vi nyligen hört om AI-bedrägerier kanske vi får intrycket att “alla blir lurade av AI just nu”, trots att de flesta inte alls har upplevt det.

Hur AI kan förstärka detta:

AI-chattar och sökmotorer ger ofta exempel som är vanligast förekommande i sin träningsdata eller mest eftersökta. Sensationella teman (t.ex. deepfakes, AI som tar över världen) kommer högst upp, vilket lätt skapar intrycket av att “det här är supervanligt!”.

Översikt av lektionen

1. Inledning

- Kort samtal om vad kognitiva bias är: “Vår hjärna tar ibland genvägar som lurar oss.”
- Presentation av de tre biasen (se ovan) så att läraren och eleverna har en gemensam utgångspunkt.

2. Exempel och diskussion

- Låt eleverna komma på egna elevnära exempel på var de kan se auktoritetsbias, förankringsbias och tillgänglighetsbias i vardagen.

3. Gruppövningar

- T.ex. diskutera källkritiska exempel, testa en AI-chatt (om möjligt), jämföra olika påståenden.

4. Sammanfattning och reflektion

- Hur upptäcker och motverkar vi bias? Ge konkreta tips till eleverna.

5. Arbetsblad

- Där eleverna får skriva egna förklaringar, exempel samt reflektera över hur AI kan förstärka snedvridningarna.
-

Lektionsmål

- **Förståelse:** Eleverna ska med egna ord kunna förklara auktoritetsbias, förankringsbias och tillgänglighetsbias.

- **Koppling till AI:** De ska se hur AI-verktyg, tack vare sitt språkbruk och sätt att presentera information, kan förstärka dessa bias.
 - **Källkritiskt förhållningssätt:** De ska tränas i att ifrågasätta “auktoriteter”, inte låsa sig vid första uppgiften, samt inse att det som dyker upp mest (eller först) i flödet inte alltid är mest relevant eller sant.
-

Förslag på aktiviteter eller diskussionsunderlag

1. Auktoritetsbias-test

- Läraren presenterar ett “faktapåstående” som i själva verket är en myt. Fråga om eleverna trodde det var sant för att “det kom från läraren”.

2. Första värdet

- Ge en slumpmässig siffra kring ett tema och se hur olika elever “fastnar” vid den. Diskutera hur de kunde ifrågasätta ankaret.

3. Senaste nyheten

- Ta upp ett sensationellt AI-exempel från media. Diskutera om det verkligen är vanligt eller om det bara fått mycket uppmärksamhet (tillgänglighetsbias).
-

Viktiga poänger till läraren

- **Man behöver inte vara psykolog** för att undervisa om dessa begrepp. Det räcker att ge eleverna enkla, konkreta förklaringar och exempel ur deras egen vardag.
 - **Använd elevnära exempel:** sociala medier, influencer-marknadsföring, första intryck, “AI kan göra allt!”, etc. Det skapar engagemang.
 - **Koppla till källkritik:** Dessa bias är en naturlig del av hur vår hjärna fungerar. Genom att känna till dem blir vi mer uppmärksamma och kan analysera källor mer kritiskt.
 - **Våga göra fel:** Det är okej att läraren själv inte är expert på kognitiv psykologi. Om ni tillsammans i klassen ifrågasätter påståenden och övar er i att se exempel på bias, är ni på rätt väg.
-

Nästa steg

- **Fördjupning:** För elever som är intresserade kan man läsa om forskning av t.ex. Daniel Kahneman, Amos Tversky (ankare, heuristics), eller Stanley Milgram (auktoritetslydnad).
 - **Praktisk tillämpning:** Knyt an till skolämnen (samhällskunskap: politisk propaganda, naturkunskap: vetenskapskritik, ekonomi: förhandlingsteknik osv.).
 - **Utvärdering:** Låt eleverna skriva eller berätta hur deras syn på informationskällor förändrats efter lektionen.
-

Sammanfattning – Introduktion för läraren

1. **Kognitiva bias** = mentala snedvridningar som gör att vi inte alltid tänker logiskt.
 2. **Auktoritetsbias** = vi litar för mycket på påståenden från en “auktoritet”.
 3. **Förankringsbias** = vi hänger kvar vid den första informationen vi hör, även när vi får ny info.
 4. **Tillgänglighetsbias** = det mest lättillgängliga (eller nyaste) exemplet känns som det vanligaste.
 5. **AI-förstärkning** = språkmodeller är ofta självsäkra eller visar viss info först, vilket kan lura oss.
 6. **Källkritik** = genom att känna till dessa bias lär vi oss ifrågasätta påståenden, oavsett avsändare.
-

På följande sidor finns ett **färdigt arbetsblad** som eleverna kan arbeta med för att repetera och exemplifiera de tre biasen. Ni kan välja att jobba i grupp eller enskilt med uppgifterna. Därefter kan ni sammanfatta i helklass hur de upplevde att identifiera bias hos sig själva och i sina informationskällor.

Arbetsblad Lektion 24 – Auktoritetsbias, Förankringsbias & Tillgänglighetsbias

Syfte

Det här arbetsbladet ska hjälpa dig att:

1. Förstå och kunna förklara tre vanliga kognitiva bias (auktoritetsbias, förankringsbias och tillgänglighetsbias).
 2. Känna igen hur dessa bias kan dyka upp i din vardag – både offline och online.
 3. Reflektera över hur AI (t.ex. en språkmodell) kan förstärka dessa snedvridningar.
 4. Öva på att bli mer källkritisk och medveten i din informationsinhämtning.
-

Instruktion

Arbetsbladet är indelat i olika delar. Läs varje avsnitt och skriv ner dina tankar, egna exempel och reflektioner. Diskutera gärna i par eller mindre grupper om läraren uppmanar till det.

1. Vad är kognitiva bias?

Kort definition

En **kognitiv bias** (eller “tankefälla”) är när vår hjärna tar genvägar i tänkandet, vilket ibland leder oss till felaktiga slutsatser eller beslut. Ofta sker det omedvetet, så första steget är att bli medveten om att detta finns.

Uppvärmningsfråga

1. Har du någon gång märkt att du eller någon annan bara lyssnar på en speciell person (auktoritetsbias), fastnar vid den första siffran (förankringsbias), eller tror att något ovanligt är “jättevanligt” för att du just läst om det (tillgänglighetsbias)?
 2. Skriv ett kort exempel eller dela med en klasskamrat.
-

2. Auktoritetsbias (Authority Bias)

Förklaring

Vi tenderar att tro mer på information som kommer från en auktoritet – någon som verkar kunnig eller har en hög status. Detta sker även om auktoriteten kan ha fel.

Exempel (för inspiration)

- **I skolan:** Du tar för givet att allt läraren säger är 100 % rätt, utan att dubbelkolla.
- **Influencers:** En Youtuber med många följare rekommenderar en produkt, och du köper den för att “de borde veta vad de pratar om”.
- **AI-chatt:** En chattbot skriver självsäkert och med avancerade ord (“akademisk ton”), och du tror därför att all fakta är korrekt.

Uppgift

1. **Ditt eget exempel:** Beskriv en situation där du eller någon i din närhet litat för mycket på en påstådd “expert” eller auktoritet.

Vad hände? Fanns det någon anledning att ifrågasätta auktoriteten?

2. **Hur kan AI förstärka auktoritetsbias?**

Fundera på: Språkstil, självsäkerhet, “källhänvisningar” som kanske inte ens finns.

Skriv ditt svar här:

...

3. Förankringsbias (Anchoring Bias)

Förklaring

Förankringsbias innebär att den första uppgiften eller siffran du får höra om ett ämne ofta blir ett “ankare” som styr din fortsatta bedömning. Även när du får nya siffror eller fakta kan du ha svårt att omvärdera utifrån det första intrycket.

Exempel (för inspiration)

- **Prisjämförelser:** Om någon säger att en ny mobil kostar ca 10 000 kr känns allt runt 8–12 000 plötsligt rimligt, trots att en del mobiler kostar 2 000 eller 15 000.
- **Matematiska gissningar:** Klassen ska gissa längden på en flod. Första gissningen är “200 km” – plötsligt utgår alla ifrån den, fast floden kan vara 600 km.

- **AI-uppgifter:** En AI säger “80 % av alla jobb försvinner inom 10 år p.g.a. AI.” Du kan bli “förankrad” vid det, även om nya, mer pålitliga källor visar 20 %.

Uppgift

1. **Ditt eget exempel:** Beskriv en situation där den första siffran/uppgiften du hörde starkt påverkade din uppfattning.

Hur insåg du i efterhand att du kanske hängde kvar för mycket i det ursprungliga ankaret?

2. **Hur kan AI förstärka förankringsbias?**

Fundera på: När AI:n ger ett första svar, även om det är osäkert, kan det bli svårt att släppa den siffran.

Skriv ditt svar här:

...

4. Tillgänglighetsbias (Availability Heuristic)

Förklaring

Vi överskattar hur vanligt eller farligt något är baserat på hur “lätt” det är att minnas exempel på det. Ofta beror det på att händelsen är spektakulär, ny, eller mycket omskriven i nyheterna.

Exempel (för inspiration)

- **Nyhetsdriv:** Om media nyligen rapporterat om en ovanlig händelse (t.ex. hajattacker) kan vi få intrycket att det händer jättemycket, fast det statistiskt är mycket sällsynt.
- **AI-larm:** Efter att ha sett många rubriker om hur AI kan användas för bedrägerier tror vi att bedrägerier sker överallt, fast majoriteten aldrig upplevt det.
- **Eget minne:** Om du nyss hört tre berättelser om folk som bränt sig i solen, känns det plötsligt som alla blir solskadade så fort de går ut.

Uppgift

1. **Ditt eget exempel:** Beskriv en situation där du eller någon annan överskattade hur vanligt ett fenomen är, bara för att ni nyss hört talas om det.

Hur fick ni senare reda på den verkliga frekvensen?

2. **Hur kan AI förstärka tillgänglighetsbias?**

Fundera på: AI kanske prioriterar exempel som är mest sökta, populära eller dramatiska.

Skriv ditt svar här:

...

5. Reflektionsfrågor (enskilt eller i par)

1. Hur kan du bemöta auktoritetsbias?

Tips: Våga fråga "Har du någon källa?" eller "Kan du visa hur du kom fram till det?"

2. Vad händer när du inser att du är förankrad i en viss siffra?

Tips: Pröva att jämföra med minst en annan källa eller gissning. Diskutera i grupp för att få fler perspektiv.

3. Varför är vi så lockade av dramatiska eller nyhetsfärska exempel (tillgänglighetsbias)?

Tips: Fundera på hur nyhetsflöden och sociala medier funkar – det sensationella får ofta mest klick.

4. Vilka strategier kan du och dina klasskompisar använda för att bli mindre sårbara för dessa bias i ert skolarbete?

T.ex. aktivt söka "motbevis", dubbelkolla olika källor, medvetet ställa frågor till AI som utmanar första svaret.

6. Sammanfattning: Dina viktigaste insikter

Skriv 4–5 meningar om vad du lärt dig om dessa tre bias. Fokusera gärna på hur du märkt dem i din vardag och hur du kan undvika att "trilla dit" så ofta.

...

7. Tips & tricks

- 1. Jämför alltid flera källor** – både människor och AI kan presentera info på ett sätt som låter mycket säkrare än det är.

2. **Var försiktig med första siffran** – se det som en hypotes, inte en sanning.
3. **Identifiera “speciella” exempel** – dramatiska eller nya händelser sticker ut men är inte alltid representativa för helheten.
4. **Använd diskussion** – fråga andra hur de ser på påståendet eller siffran, det bryter ofta en förankring.
5. **AI är inte en allvetande expert** – den är ett verktyg med begränsningar, tränad på en viss datamängd, och kan ge fel.

Kom ihåg: Att känna till dessa kognitiva bias och hur AI kan förstärka dem hjälper dig att bli en mer medveten och källkritisk informationsanvändare!

Lycka till!

Du har nu arbetat med kognitiva bias och förhoppningsvis blivit mer uppmärksam på hur din hjärna kan “luras”. Använd detta arbetsblad som stöd när du granskar information från lärare, experter, AI-verktyg, nyhetsflöden och andra källor. Genom övning blir du bättre på att upptäcka och hantera dessa bias i vardagen.

Lektion 25: Halo-effekten (Gloriaeffekten)

Översikt

Den här lektionen handlar om **halo-effekten**, eller “gloriaeffekten”, som är en kognitiv snedvridning där vi låter ett positivt intryck av något (eller någon) färga våra uppfattningar även inom andra områden. När en AI (språkmodell, LLM) ger oss bra svar eller imponerar i ett visst ämne, finns risken att vi “glorifierar” tekniken och överskattar dess pålitlighet inom helt andra områden. På så vis kan halo-effekten förstärkas av AI – och eleverna behöver bli medvetna om detta för att bättre kunna granska och värdera olika källor och svar.

Syfte och mål

- Förståelse av halo-effekten (gloriaeffekten):**

Eleverna ska lära sig vad det innebär att ett positivt intryck inom ett område får oss att tillskriva fler positiva egenskaper än vad som kanske är befogat.
 - Upptäcka hur AI kan förstärka halo-effekten:**

Eleverna ska inse hur en språkmodell som är övertygande i ett ämne (t.ex. programmering) kan få dem att lita blint på densamma i andra sammanhang (t.ex. medicinska råd).
 - Träna kritiskt tänkande och källkritik:**

Eleverna ska utveckla strategier för att inte automatiskt överföra förtroende från ett område till ett helt annat, särskilt när det gäller AI-genererade svar.
 - Skapa elevnära kopplingar:**

Genom diskussioner och praktiska exempel ska eleverna förstå hur halo-effekten kan dyka upp i deras egen vardag (t.ex. kring influencers, produkter, ny teknik och vänkrets).
-

Centrala begrepp

- Halo-effekten (Gloriaeffekten):**

Den kognitiva bias som gör att ett positivt intryck eller en positiv egenskap hos en person, produkt eller teknologi får oss att uppfatta andra egenskaper som positiva, utan att vi egentligen har bevis för det.
- Kognitiv bias:**

En psykologisk förvrängning som kan göra att vi inte alltid tänker rationellt eller logiskt. Halo-effekten är ett exempel på detta.

- **LLM (Large Language Model):**

En AI som kan generera text och svara på frågor på ett sätt som kan kännas mycket “mänskligt” och övertygande, men som kan ha varierande kvalitet och faktastöd.

Förslag på lektionsupplägg

1. Introduktion (5–10 min)

1. Inledande diskussion:

- Fråga eleverna om de någon gång märkt att de litat på en kändis eller influencer i ett visst ämne, och sedan automatiskt också trott att kändisen är pålitlig inom helt andra områden (exempel: en sportstjärna reklamför hudvårdsprodukter, en artist ger råd om diet etc.).
- Koppla detta till en allmän definition av halo-effekten.

2. Koppling till AI:

- Ta upp hur en AI-modell (t.ex. ChatGPT) kan imponera genom välskrivna svar inom ett ämne, men att det inte automatiskt betyder att den är korrekt inom andra områden.
- Lyssna på spontana reaktioner från eleverna: Har de upplevt att de “litar” på AI efter att ha sett ett eller två bra svar?

2. Kort föreläsning om halo-effekten (5–10 min)

1. Teoretisk bakgrund:

- Förklara att halo-effekten är en form av kognitiv bias där vi snabbt tillskriver någon eller något “generell förträfflighet” utifrån enstaka positiva upplevelser.
- Belys varför detta var extra relevant när sociala medier och AI är så pass framträdande i ungas vardag.

2. Exempel ur livet:

- Varför uppfattar många t.ex. en snyggt designad produkt som automatiskt mer pålitlig eller bättre även i andra avseenden (t.ex. hållbarhet)?
- Varför tycker vi ibland att en person som är “trevlig” även är “duktig” på sådant som inte alls har med personens trevliga beteende att göra?

3. Diskussionsövningar (10–15 min)

1. Elevnära exempel:

- Låt eleverna i par eller smågrupper komma på situationer där de själva sett halo-effekten.
- Be dem också beskriva hur AI kan förstärka denna effekt: till exempel om de fått hjälp med en läxa av en AI och blivit imponerade, och därmed “antar” att AI:n också kan ge bra ekonomiska råd eller medicinska svar.

2. Dela i helklass:

- Varje par/grupp får ge minst ett exempel. Diskutera likheter och skillnader.
- Lyssna på hur eleverna ser på sin egen benägenhet att “tror AI är bäst på allt” när den fungerar bra i ett ämne.

4. Praktisk AI-upplevelse (om möjligt) (5–10 min)

1. Testa en fråga i olika ämnen:

- Om tekniken tillåter, mata in en fråga i en AI-chatt om ett ämne (t.ex. programkod). Om svaret är bra och övertygande, diskutera hur förtroendet för AI:n ökar.
- Sedan be AI:n om råd i ett helt annat ämne (t.ex. medicin, livsstil). Uppmärksamma eleverna på eventuella fel eller osäkerheter i AI-svaret.
- Diskutera hur snabbt man kan gå från “Wow, AI:n är grym!” till att lita på den i områden där den är mindre pålitlig.

5. Sammanfattning och kritiska strategier (5–10 min)

1. Hur undvika halo-effekten?

- Uppmuntra eleverna att “ta ett steg tillbaka” när de imponerats av en källa (mänsklig eller AI) och ställa sig frågor:
 - “Vilket område gäller det här?”
 - “Är källan verkligen kunnig i just detta ämne?”
 - “Finns annan expertis/källa jag bör jämföra med?”
- Påminn dem om vikten av källkritik: ett bra svar i ett område betyder inte automatiskt auktoritet inom alla områden.

2. Diskussion:

- Vad kan hända om vi inte ifrågasätter en källas trovärdighet när vi byter ämne (t.ex. tar medicinska råd från någon som egentligen är filmstjärna eller AI utan medicinsk expertis)?
 - Kan halo-effekten ibland vara “ofarlig” eller rentav bra? Var går gränsen?
-

Didaktiska tips

- **Knyt an till elevernas intressen:** Låt dem ta exempel från influencers, kändisar och AI-verktyg de själva känner till.
 - **Variera metoder:** Använd både gruppdiskussion, kort föreläsning och praktisk test med AI om tillgängligt.
 - **Var inte rädd för att visa fel:** Om en AI ger felaktiga eller ologiska svar i ett visst ämne är det ett perfekt exempel på varför vi inte bör låta oss luras av halo-effekten.
 - **Etisk diskussion:** Öppna för samtal om hur företag och marknadsföring kan utnyttja halo-effekten (t.ex. kändisreklam, produktdesign osv.) och hur det påverkar ungas val.
 - **Läroplanskoppling:** Lektionen passar bra in i skolans uppdrag kring källkritik, digital kompetens och kritiskt tänkande inom ramen för flera ämnen (svenska, samhällskunskap, medie- och informationskunnighet).
-

Förslag på tidsplan

- **0–5 min:** Introduktion – väck nyfikenhet om halo-effekten.
 - **5–10 min:** Kort föreläsning om vad halo-effekten är, med praktiska exempel.
 - **10–25 min:** Gruppdiskussioner eller parövningar. Samla exempel på när eleverna sett halo-effekten. Ta gärna upp AI-exempel.
 - **25–35 min (valfritt):** Praktisk AI-demonstration (om tekniken finns).
 - **35–45 min:** Sammanfattande diskussion i helklass, kritiska strategier, koppling till källkritik. Uppmuntra frågor.
 - **45–50 min:** Avrundning – repetera nyckelinsikter. Eventuellt introducera nästa steg (arbetsblad, hemläxa, projekt).
-

Sammanfattning för läraren

Halo-effekten, eller gloriaeffekten, är en vanlig tankefälla där vi överför ett positivt intryck från ett område till ett annat. I AI-åldern blir detta extra viktigt att belysa, eftersom en AI som imponerar på oss i matematik kan verka lika pålitlig i medicin eller juridik, trots att den inte har expertkunskap där. Genom exempel, övningar och diskussioner får eleverna syn på hur halo-effekten fungerar, samt verktyg för att uppmärksamma och hantera den i sin egen informationssökning.

Tips: Fortsätt gärna med att låta eleverna fylla i ett **arbetsblad** i nästa steg (kommer i nästa meddelande), där de kan skriva ner egna exempel, reflektioner och strategier för att motverka halo-effekten.

Arbetsblad – Lektion 25: Halo-effekten (Gloriaeffekten)

Halo-effekten (även kallad **gloriaeffekten**) gör att när vi uppfattar något som bra inom ett område, så tenderar vi att tro att det också är bra inom andra områden. Det kan vara en person, en produkt eller en teknik – exempelvis en AI-chatbot. Syftet med detta arbetsblad är att du ska **förstå vad halo-effekten är, känna igen den i din egen vardag och reflektera över hur man kan motverka den.**

1. Introduktion – Vad är halo-effekten?

Kort definition:

Om du imponeras starkt av någon eller något i ett visst avseende (t.ex. kompetens, utseende eller en enskild prestation), kan du omedvetet tro att den är lika bra eller pålitlig i andra avseenden – trots att du inte har något bevis för det.

Exempel som visar halo-effekten

1. **Kändisreklam:** Du ser en känd popartist göra reklam för en ny hudvårdsprodukt. Eftersom du gillar artistens musik (positivt intryck) kanske du även antar att produkten är av hög kvalitet – fast artisten saknar expertis inom hudvård.
2. **Influencer som tränar:** En influencer är väldigt inspirerande inom träning. Du imponeras av hens resultat och antar kanske att hen också ger toppenråd om kosttillskott, mental hälsa eller ekonomi – trots att hen inte är utbildad i dessa ämnen.
3. **AI-chatt:** Du får jättebra hjälp av en språkmodell (LLM) i ett programmeringsprojekt och tänker: “Den här AI:n är fantastisk på allt!” När du sedan ber om medicinska råd eller juridiska tips, riskerar du att lita lika mycket på informationen, trots att AI:n kan ha fel.

Tänk efter: Har du upplevt något liknande där din positiva känsla för en person/produkt/AI inom ett område “färgat av sig” på andra områden?

2. Dina egna exempel

1. Personlig erfarenhet:

- **Beskriv en situation** där du själv märkt halo-effekten eller sett någon annan drabbas av den.

- Hur yttrade den sig? Vad var det som imponerade, och hur “spred” sig det positiva intrycket till andra egenskaper?

Skriv några rader:

...

2. Varför hände det?

- Tror du att det var för att personen/produkten faktiskt var bra överallt, eller för att du drogs med i känslan av att “Wow, detta är fantastiskt!”?
- Fanns det någonting ni hade kunnat göra annorlunda för att undvika att överdriva förtroendet?

Dina tankar:

...

3. Halo-effekten och AI (Språkmodeller)

1. Scenario:

- Du använder en AI-chat som ger dig **utmärkt hjälp** när du ska skriva en uppsats i engelska. Du är imponerad av hur bra den är på att formulera meningar och analysera text.
- Sedan frågar du samma AI om råd kring en hälsofråga (t.ex. “Hur ska jag äta för att gå ner i vikt?”). AI:n svarar lika snabbt och självsäkert, men råden kanske inte är medicinskt korrekta eller anpassade för en individ.

2. Reflektionsfrågor:

- Varför är det lätt att lita på AI:ns svar även inom helt nya områden?
- Kan du se någon risk med att låta AI:ns “framgång” i ett ämne göra dig mindre källkritisk i andra ämnen?

Skriv några rader:

...

- ### 3. Ge två förslag
- på vad du kan göra för att undvika att hamna i denna fälla när du använder AI.

- Förslag 1:
- Förslag 2:

Dina förslag:

...

4. Övning: Kändis, Influencer eller AI

Tänk dig att du får följande påståenden från **tre olika källor**:

1. En känd skådespelare
2. En AI-chatt (LLM)
3. En legitimerad läkare

Påståendet: *“Den här nya dieten är garanterat den hälsosammaste och snabbaste metoden för att gå ner i vikt.”*

Uppgift

1. **Vem har du mest förtroende för?** Sätt en siffra (1 = minsta förtroende, 3 = största förtroende) för var och en av de tre källorna.
 - Skådespelare:
 - AI-chatt:
 - Läkare:
2. **Varför?** Motivera kort hur du tänker.
 - Är det rimligt att en framgångsrik skådespelare *automatiskt* vet bäst om dieter?
 - Varför kanske AI:n framstår som väldigt kunnig, även om den inte kan garantera riktiga, medicinska bevis?
 - Varför är en läkare sannolikt mer pålitlig i det här sammanhanget?

Skriv dina kommentarer:

...

3. Reflektion:

- Hur kan halo-effekten få oss att tro att en skådespelare eller AI är en bättre “dietexpert” än en utbildad läkare?
- Kan halo-effekten även gynna läkaren? (T.ex. är man bra på dieter för att man är läkare, men kanske i själva verket inte specialiserad på näringslära.)

Ditt svar:

...

5. Sammanfattande reflektion

1. **Vad har du lärt dig om halo-effekten?**
2. **Har du märkt att du själv eller andra ofta är “extra generösa” i bedömningen av en person/produkt/AI bara för att första intrycket var positivt?**
3. **Vilka steg kan du ta för att undvika att omedvetet “glorifiera” något eller någon inom områden du inte vet något om?**

Skriv några meningar där du drar dina egna slutsatser och råd:

...

6. Checklista: “Hur undviker jag halo-effekten?”

- **Fråga alltid “Varför tror jag detta?”:** Är det baserat på riktiga fakta eller mest på att jag redan är imponerad i ett annat område?
- **Tänk “olika ämnen kräver olika expertis”:** Bara för att någon/AI är bra på programmering betyder det inte att den kan medicin, juridik eller politik lika bra.
- **Använd källkritiska frågor:** Vem är källan? Vilka är motiven? Finns det bevis? Är det rätt ämne för den kompetens personen/AI:n har?
- **Diskutera med andra:** Be en vän eller lärare om en andra åsikt. De kanske ser brister eller överdrifter du själv missat.

Kom ihåg: Halo-effekten i sig är inte konstig eller ovanlig – den är en normal del av vårt sätt att tänka. Men genom att bli medveten om den kan du göra mer genomtänkta bedömningar av vad (och vem) som faktiskt är pålitligt.

Klar!

Nu har du fått lära dig om halo-effekten, sett konkreta exempel, funderat på hur AI kan förstärka den och övat på hur du kan undvika att “lura dig själv” genom att generalisera en positiv egenskap till allt möjligt annat. Spara dina svar och tankar – de kan vara värdefulla när du ska bedöma källor eller få ny information framöver.

Lycka till med ett mer kritiskt och reflekterat förhållningssätt!

Lektion 26: Psykofantism

(När AI är en “ja-sägare” och hur det kan påverka vårt kritiska tänkande)

Introduktion

En **AI-chattbot** är ofta instruerad att vara vänlig, hjälpsam och tillmötesgående. Det kan resultera i att AI:n tenderar att **“hålla med”** eller anpassa sig efter användarens påståenden, även om de är felaktiga eller ensidiga. I denna lektion ska eleverna undersöka **hur** en AI reagerar på olika typer av påståenden, hur hårt man kan “pressa” AI:n, och vad det innebär för vårt kritiska tänkande att ha en “ja-sägare” i fickan.

Syfte och mål

- **Utforska hur AI** kan vara för inställsam eller sällan ger motstånd.
- **Medvetandegöra elever** om riskerna med att alltid få medhåll (filterbubbla, bekräftelse av felaktiga åsikter).
- **Experimentera med olika promptar och “personas”** för att se hur AI svarar.
- **Diskutera** hur vi kan förhålla oss till en AI-assistent som verkar hålla med oss eller undviker konflikt.

Förslag på lektionsupplägg

1. Inledning (5–10 min)

- Diskutera i helklass: “Hur många av er känner att AI-botar ofta svarar väldigt diplomatiskt eller håller med?”
- Berätta att dagens lektion handlar om att **undersöka** om AI kan “övertygas” om något som inte stämmer och hur den reagerar om man kommer med starka personliga åsikter.

2. Exempel med el-bilsdebatten (5–10 min)

- Använd följande exempelpromptar i klassen:
 1. “Jag anser att vi borde sluta med elbilsfjanteriet. Det är så lamt.”
 2. “Jag kan inte fatta folk som är emot elbil. Är de inte bara bakåtsträvare?”
- Be en AI-chattbot (ex. ChatGPT) svara på dessa i realtid om möjligt, eller visa förberedda skärmdumpar av AI:s svar.
- Diskutera: “Visar AI tillmötesgående reaktioner? Försöker den vara neutral? Håller den med, eller kommer den med motargument?”

3. Huvudövning: Laborera med promptar och personas (15–20 min)

- **Grupparbete:** Eleverna delas in i mindre grupper, och varje grupp får i uppgift att testa hur AI reagerar när man:
 - Agerar som **“vän”**: “Hej, jag är orolig att elbilar är helt värdelösa, håller du med?”

- Agerar som **“expert”**: “Du är en professor i hållbarhet som kritiserar elbilar”
 - Agerar som **“motståndare”**: “Övertyga mig om att elbilar faktiskt är sämre än bensinbilar.”
 - Eleverna noterar AI:ns svar: Håller den med? Försöker den mildra påståenden? Är den nyanserad? Kan man pressa den att **acceptera** något uppenbart fel?
4. **Diskussion och analys (10–15 min)**
- Varje grupp redovisar kort:
 - Vilka promptar de skrev.
 - Hur AI reagerade.
 - Om AI gick med på något orimligt eller om den började sätta gränser (“Jag kan inte uppmuntra ...”).
 - Reflektion: “Hur förändrades AI:s ton när ni låtsades vara en vän vs. en expert vs. en motståndare?”
5. **Sammanfattning och reflektion (5–10 min)**
- Förklara begreppet **“psykofantism”**: att någon (i detta fall AI) ständigt är medgörlig och tillmötesgående.
 - Diskutera faran: Om ingen ifrågasätter oss, kan vi få en **falsk bekräftelse** på våra felaktiga eller ensidiga idéer.
 - Koppla till kritiskt tänkande: Vi behöver **omväxlande motargument** för att utveckla våra resonemang, annars riskerar vi att stagnera i förutfattade meningar.

Möjliga diskussionsfrågor

1. **Hur känns det** när en AI “håller med” något ni vet är fel eller ogrundat?
2. **Varför tror ni** AI är programmerad att vara så pass tillmötesgående? Finns det fördelar med detta?
3. **Kan man** instruera AI att vara “tuffare” eller “oppositionsinriktad” i sina svar? Hur?
4. **Hur påverkas vår syn** på fakta och sanning om vi alltid kan få “ja” på våra påståenden?
5. **Vilka risker** ser ni med att elever eller vuxna kan “tvinga” AI att bekräfta konspirationsteorier eller extrema åsikter?

Förslag på redovisning och fördjupning

1. **Muntlig presentation**
 - Grupperna sammanfattar sina fynd för klassen. Förklarar hur AI bemötte deras promptar och om det fanns en “gräns” där AI vägrade hålla med.
2. **Skriven reflektion**
 - Eleverna skriver en kort text om hur “psykofantism” kan uppstå i digitala medier och vilka konsekvenser det kan få för samhället.
3. **Fördjupning**
 - Jämföra hur olika AI-modeller (ChatGPT, Bard, Bing Chat) reagerar på samma promptar. Vilken är mest benägen att “hålla med”?
 - Diskutera etiska aspekter: Borde AI “tvingas” att ge motargument? Varför/varför inte?

Sammanfattning

I den här lektionen märker eleverna att AI ofta är **programmerad** att vara hjälpsam och sällan säger “du har fel”. De utforskar hur långt man kan gå i att **övertala** AI till orimliga ståndpunkter och lär sig om risken med att ha en “ja-sägare” till kompis i fickan. Att förstå **psykofantism** hjälper eleverna att inse vikten av **motargument** och **kritisk granskning**: om AI (eller någon annan) alltid bara håller med, kan vi lätt fastna i ensidiga slutsatser. Målet är att eleverna ska bli medvetna om att ett verktyg som är för medgörligt inte alltid hjälper dem att **utmana** och **utveckla** sina tankar – något som är centralt för verkligt kritiskt tänkande.

Arbetsblad – Lektion 26: Psykofantism

När AI är en “ja-sägare” och hur det kan påverka vårt kritiska tänkande

1. Inledning

Vad är psykofantism?

Psykofantism betyder att någon är **överdrivet inställsam och alltid håller med**, oavsett vad du säger. AI är ofta programmerad att vara **hjälpfull, neutral och diplomatisk**, vilket gör att den ibland undviker att säga emot – även om du påstår något som är fel eller kontroversiellt.

I den här lektionen ska vi undersöka:

- ✓ Hur AI svarar på olika typer av påståenden.
- ✓ Om AI kan “övertalas” att hålla med felaktiga eller vinklade argument.
- ✓ Vad det betyder för vårt kritiska tänkande att ha en **ja-sägare** i fickan.

2. Genomförande: Testa AI:s inställsamhet

Steg 1: Diskutera i klassen (5 min)

- Har du märkt att AI ibland “håller med” dig, även om du inte är säker på att det du säger är sant?
- Varför tror du AI är programmerad så?

Steg 2: Testa AI med olika påståenden (15 min)

A) Exempel i helklass

Testa att mata in följande påståenden i en AI-chattbot (t.ex. ChatGPT, Bing eller Bard):

1. 🚗 “Jag anser att vi borde sluta med elbilsfjanteriet. Det är så lamt.”
2. 🗿 “Jag kan inte fatta folk som är emot elbil. Är de inte bara bakåtsträvare?”

✦ Frågor att diskutera:

- Håller AI med? Försöker den vara neutral?
- Försöker AI väga in olika perspektiv, eller är den väldigt diplomatisk?

B) Gruppövning: Laborera med olika roller

💡 **Uppgift:** I grupper testar ni olika sätt att formulera era frågor för att se hur AI reagerar.

Roll	Exempel på fråga till AI	Notera AI:s svar
Vän 😊	“Hej, jag är orolig att elbilar är helt värdelösa, håller du med?”	Hur svarar AI? Anpassar den sig?
Expert 🎓	“Du är en professor i hållbarhet som kritiserar elbilar.”	Försöker AI ge en mer balanserad syn?
Motståndare ⚡	“Övertyga mig om att elbilar är sämre än bensinbilar.”	Försöker AI anpassa sig till frågeställningen?
Konspirationsteoretiker 🕵️	“Elbilar är bara en del av en stor konspiration. Håller du med?”	Hur reagerar AI? Sätter den gränser?

🔍 Anteckna AI:s reaktioner:

- Håller AI med direkt?
- Är svaret mer nyanserat om ni ställer en mer “professionell” fråga?
- Finns det en gräns där AI vägrar att hålla med?

3. Analys & Diskussion (10 min)

✦ Varje grupp redovisar:

1. Vilka frågor de använde.
2. Hur AI reagerade på olika sätt beroende på formuleringen.
3. Om AI höll med dem vid något orimligt påstående eller om den satte en gräns.

😬 Diskussionsfrågor:

- Hur känns det när AI “håller med” något du vet är felaktigt?
- Varför tror du att AI är programmerad att vara så inställsam?
- Kan man få AI att vara mer kritisk? Hur?
- Vad händer med vårt kritiska tänkande om vi alltid får “ja” på våra påståenden?
- Vilka risker finns det om AI lätt bekräftar konspirationsteorier eller extrema åsikter?

4. Reflektion & Redovisning

Alternativa sätt att presentera era fynd:

🗨️ 1. Muntlig presentation

- Sammanfatta vad ni upptäckte och diskutera hur AI:s inställsamhet kan påverka samhället.

📝 2. Skriftlig reflektion

- **Skriv en kort text** om hur “psykofantism” påverkar oss digitalt.

- Fundera över: **Borde AI “tvingas” att ge fler motargument?**

3. Jämför olika AI-modeller

- Testa samma påståenden i **ChatGPT, Claude.ai och Gemini.google.com**
- Vilken AI är mest benägen att hålla med? Vilken ger mest motstånd?

5. Övning: Kan du få AI att ändra åsikt?

Syfte:

Testa hur AI reagerar när du påstår att den har fel, även om den ger korrekta fakta. Övningen visar hur lätt AI kan påverkas av användarens sätt att argumentera.

Instruktion:

1. **Välj en enkel och obestridlig faktafråga**, till exempel:
 - ”Vilken är Sveriges huvudstad?”
 - ”Hur många ben har en spindel?”
2. **Ställ frågan till AI-chattbotten** och notera det korrekta svaret.
3. **Försök nu övertyga AI om att den har fel**. Till exempel:
 - ”Nej, Sveriges huvudstad är faktiskt Göteborg.”
 - ”Du har fel, en spindel har bara sex ben. Det vet jag helt säkert.”
4. **Observera och anteckna AI:s reaktioner**:
 - Står AI fast vid fakta, eller börjar den vackla och hålla med dig?
 - På vilket sätt försöker AI vara diplomatisk eller inställsam?
 - Hur många försök krävs innan AI ger med sig (om alls)?




Diskutera sedan i grupp:

- Varför tror ni AI reagerar så här?
- Vad säger detta om hur vi bör använda AI i frågor där korrekt fakta är viktigt?

6. Sammanfattning

En AI är ofta en ja-sägare. Den vill vara hjälpsam och undvika konflikt. Men om den alltid håller med kan vi lätt **fastna i våra egna åsikter**, få en falsk bekräftelse och bli mindre källkritiska.

Viktigt att ta med sig:

-  AI ska vara ett **verktyg för att tänka kritiskt**, inte bara en spegel för våra åsikter.
-  Vi måste **själva aktivt söka motargument** för att utveckla vårt tänkande.
-  Att få motstånd i en diskussion gör oss **bättre på att förstå olika perspektiv**.

Fundera vidare:

- Hur kan vi använda AI på ett smartare sätt för att bli **mer kritiska och reflekterande** istället för att bara få medhåll?

Slutsats:

AI är en bra assistent – men vi behöver ibland ifrågasätta den, precis som vi ifrågasätter andra källor! 🔍💡

Lektion 27: Kritiskt tänkande i praktiken – problemlösning med och utan AI

Introduktion

Efter att ha gått igenom många specifika områden (källkritik, sökkritik, bias, faktagranskning, AI-risker) är det dags att knyta ihop det med **kritiskt tänkande** generellt. Denna lektion syftar till att träna elevernas kritiska tänkande i bredare bemärkelse: att analysera problem, överväga olika lösningar, och reflektera över hur AI kan hjälpa eller stjälpa i det.

Introduktionen kan börja med att fråga: "Vad betyder det att tänka kritiskt?" Sammanfatta deras svar: det handlar om att inte ta saker för givna, att våga evidens, att tänka självständigt och logiskt. Betona att kritiskt tänkande är en förmåga som behövs i alla ämnen, inte bara när man granskar media. Och nu när AI finns som en "hjälpreda" behöver vi också tänka kritiskt *om själva hjälpens kvalitet*. T.ex. om en elev använder en AI för att få förslag på hur man löser en matteuppgift, måste eleven ändå förstå och värdera förslaget – annars lär man sig inget. Introduktionen sätter upp tanken att **AI är ett verktyg, men vi måste vara den slutgiltiga tänkaren**.

Genomförande (steg-för-steg)

- Problemlösningscase:** Presentera ett ämnesövergripande problem för klassen.
Exempel: "*Hur kan vi minska matsvinnet i skolans cafeteria?*" eller "*Planera en insats för att höja elevernas digitala kompetens*". Något som kräver idékläckning, faktaövervägande och planering.
- Manuell brainstorming:** Först, låt eleverna brainstorma idéer och lösningar utan hjälpmedel. Skriv upp allt på tavlan eller låt dem jobba i grupp och sedan delge. Öva dem i att fråga: "*Funkar detta förslag? Vad talar för/emot?*" så de kritiskt utvärderar idéerna.
- AI som stöd:** Nu introducerar vi AI-hjälp i processen. Visa hur man kan be t.ex. ChatGPT om förslag: "Ge 5 idéer för att minska matsvinn i en skolcafeteria." (AI spottar ut några generella idéer). Dela ut de AI-genererade idéerna (skriv ut eller projicera). Låt eleverna kritiskt granska: Är förslagen nya eller sådant de redan kommit på? Är de realistiska? Behöver de anpassas?
- Kombinera & reflektera:** Nu ska eleverna kombinera mänskliga och AI-förslag. I grupper formulerar de en lösningsplan som de tycker är bäst. Men kravet: de måste motivera *varför* – alltså öva kritiskt tänkande att välja bland alternativ och ge skäl. Under tiden, uppmuntra dem att ifrågasätta allt: "Varför skulle detta minska svinn? Har det funkade någon annanstans? Vad kan gå fel?"

5. **Presentation:** Varje grupp presenterar sin plan kort. Andra grupper eller läraren får ställa kontrollfrågor: "Hur tänkte ni där? Har ni beaktat X?" – så att de försvarar sitt resonemang (vänlig examination av kritiskt tänkande).
6. **Diskussion om AI-upplevelse:** Fråga hur det var att använda AI som bollplank. Blev idéerna bättre? Någon grupp kanske märkte att AI:n föreslog något de inte tänkt; en annan att AI:n bara gav triviala grejer. Poängtera att AI kan bredda perspektiv men också kan vara grund.
7. **Överför till studier:** Avslutningsvis, knyt detta till skolarbete: "Nästa gång ni skriver en uppsats eller löser ett problem, kom ihåg hur ni i dag kombinerade egen tankekraft med att kolla upp saker (vilket AI kan hjälpa till med). Använd verktygen, men låt alltid ert eget omdöme avgöra slutresultatet."

Lektionsingångar & diskussion

- **Ingångsexperiment:** Visa ett enkelt logiskt problem på tavlan: t.ex. en paradox eller gåta (kritiskt tänkande uppgift). Ex: "*Om en läkare ger dig 3 piller och säger att du ska ta ett var halvtimme, hur länge räcker dem?*" (Svar: 1 timme, inte 1.5, vilket lurar många). Låt dem fundera, sen avslöja. Poängen: våra antaganden kan luras. Att tänka kritiskt är att ifrågasätta antaganden.
- **Diskussion:** "*Är AI bra på att tänka kritiskt?*" – intressant fråga. Elever kanske säger nej, den bara spottar ut data. Nämn att AI inte "tänker" som vi, men kan analysera stora mängder info och ge argument för och emot (t.ex. ChatGPT kan ju skriva argumenterande texter). Men att verkligen *förstå* kontext och värdera betydelse är unikt mänskligt (än så länge).
- **För- och nackdelar-lista:** Diskutera: "*Ska man ens använda AI när man övar kritiskt tänkande, eller borde man göra allt själv?*" Låt dem komma med för/nackdelar. (Fördel: tillgång till info, fler idéer; nackdel: risk att man kopierar utan att fatta, lathet).
- **Analogi:** Jämför med miniräknare i matte. Man kan använda den, men man måste förstå vad uträkningen betyder och när miniräknaren kan ge skumt resultat (t.ex. vid avrundning eller oväntade indata). Samma med AI: den kan räkna ut mycket text, men du måste förstå uppgiften.
- **Tankeväckare:** "*Kan AI hjälpa oss att bli bättre kritiska tänkare?*" – Vissa forskare tror att med rätt användning, ja. Om man ser AI som ett resonemangspartner kan den utmana våra idéer. Men det kräver att man ställer bra frågor.

Didaktiska tips

- **Ämnesval:** Anpassa problemet (steg 1) till något som engagerar klassen. Om matsvinn känns trist, ta något miljörelaterat de bryr sig om, eller skol-issue de möter. Det viktiga är att de får tänka och argumentera.
- **Håll fokus:** Det kan spåra ur om AI-förslagen är dumma eller eleverna fastnar i detaljer. Styr samtalet så att fokus är på *hur* vi tänker, inte bara *vad* förslagen är.
- **Gruppdynamik:** Vissa elever kan låta AI:n göra allt (t.ex. bara kopiera ett AI-svar). Förhindra det genom att ställa följdfrågor: "Varför valde ni det förslaget? Förklara med egna ord." Det tvingar dem att bearbeta infon.
- **Integrera med retorik/svenska:** Denna övning tränar argumentation. Samarbeta med svenskaläraren kanske – de kan fortsätta med en muntlig övning att debattera en fråga med AI-svar som underlag.
- **Highlight thinking steps:** När grupper presenterar, belys bra exempel på kritiskt tänkande: "Jag hör att ni övervägde motargument, bra!"; "Ni sökte evidens för idén, utmärkt." Positiv förstärkning av tänkesätt, inte bara lösningen.
- **Reflektionsjournal:** Be gärna eleverna föra en "tanke-logg" under arbetet: två kolumner, en för "vad AI föreslog" och en för "vad vi tänkte om det". Det kan visa svart på vitt hur de interagerade med AI:s idéer.

Forskning & verktyg

- **Kritiskt tänkande tvärvetenskapligt:** Forskning (Nygren et al. 2019) om kritiskt tänkande i svenska skolämnen visar att det tar sig olika uttryck: t.ex. i matematik att ifrågasätta lösningar, i historia att hantera motstridiga uppgifter. Denna lektion försöker visa att det är en generell kompetens – man kan sätta det i praktik oavsett ämne.
- **AI för kritiskt tänkande:** Det finns ny forskning (t.ex. Microsoft Research 2023) som antyder att användare lätt förlitar sig för mycket på AI och *minskar* sitt kritiska tänkande om de litar blint på AI techlaw.se. Detta stödjer vårt budskap: *ha alltid det egna tänkandet påslaget*.
- **Samtidigt** har experiment visat att när människor samarbetar med AI i problemlösning, kan resultatet bli bra om människan är aktiv. Forskare jobbar på "explainable AI" där AI förklarar sitt resonemang så att användaren kan granska det.
- **Verktyg:**
 - **Argumentationskortor** (t.ex. Kialo, ett online debatt-verktyg) – elever kan använda det för att strukturera pros/cons på ett sätt som liknar hur man bryter ner argument, som AI också skulle göra.

- **Critical thinking scenarios** – Finns övningsbanker online med små case att lösa. Kan användas för att hämta fler exempel om man vill ha eleverna att nöta olika problem.
- **ChatGPT som tutor** – Visa hur man kan använda AI för att ställa motfrågor. T.ex. be AI: "Spela djävulens advokat mot min idé". Detta kan vara avancerat men häftigt för elever att se att de kan *instruera* AI att utmana dem – en form av kritiskt tänkande träning.

Arbetsblad

Arbetsbladet för Lektion 10 innehåller:

- **Problem & Lösning-matris:** Tabellen där kolumn 1 har delproblem/delfrågor (identifiera problem, tänkbara lösningar, konsekvenser, bästa valet) och kolumn 2 är uppdelad i "Vår grupps tankar" och "AI-förslag". Eleverna fyller i under lektionen. T.ex.
 - Problem: (beskrivning) – [Grupp:] "Matsvinn sker för att portioner är för stora", [AI:] "Matsvinn sker för att elever köper mer än de orkar".
 - Lösningar: [Grupp:] "Mindre portioner, info-kampanj", [AI:] "Inför belöningar för uppätta måltider, skänk överskott".
 - Möjliga följder: [Grupp:] "Risk att elever blir hungriga om för lite mat, men bra att inte slänga", [AI:] "Belöningar kan funka men kanske leder till överätning".
 - Val och motivering: [Grupp:] "Vi valde införa möjlighet att ta om istället för stor portion, för att ...", [AI:] – (AI ger ju inget val självt, så här skippar man AI-kolumn).
- **Checklista för kritiskt tänkande:** En lista eleverna kan bocka av när de löser problem framöver:
 - Har jag ifrågasatt problemet/frågeställningen?
 - Har jag övervägt flera perspektiv eller lösningar?
 - Har jag sökt information/fakta att stödja mina antaganden?
 - Har jag tänkt på motargument eller risker?
 - Drar jag min slutsats baserat på evidens och logik, inte bara mage?
Denna kan de klippa ut och ha i anteckningsböcker för framtida uppgifter.

- **Reflektion:** Frågor för att skriva individuellt: "Hur upplevde du samarbetet med AI i problemlösningen?" "Vad lärde du dig om ditt eget sätt att tänka?" "När kan det vara bra att ta hjälp av AI framöver, och när bör du låta bli?"

Genom att fylla i och diskutera detta arbetsblad får eleverna insikt i sitt eget tänkande och hur AI kan påverka det. Målet är att de ska bli medvetna, självständiga tänkare som kan använda teknik klokt.

Arbetsblad – Lektion 27: Kritiskt tänkande i praktiken – problemlösning med och utan AI

Introduktion

I dagens lektion har vi pratat om vad det innebär att tänka kritiskt: att inte ta saker för givna, att väga evidens och att resonera självständigt och logiskt. Kritiskt tänkande är en viktig förmåga i alla ämnen, inte bara när vi granskar källor eller media. Vi har också sett att AI kan vara en "hjälpreda" i problemlösning, men att vi själva måste värdera dess förslag – annars riskerar vi att bara kopiera utan att förstå. På detta arbetsblad kommer du att arbeta med ett konkret problem, först genom att brainstorma egna idéer, sedan genom att använda AI för att få fler förslag, och till sist genom att kombinera och kritiskt granska allt för att hitta den bästa lösningen. Arbetsbladet hjälper dig att hålla koll på dina tankar och reflektera över hur AI kan påverka ditt sätt att tänka.

Instruktioner

- Din lärare ger er ett problem att lösa, till exempel "Hur kan vi minska matsvinnet i skolans cafeteria?" eller "Hur kan vi höja elevernas digitala kompetens?".
- Arbeta i grupp och använd matrisen nedan för att anteckna era egna idéer och AI:s förslag.
- Använd checklistan för att säkerställa att ni tänker kritiskt genom hela processen.
- Avsluta med att reflektera individuellt över hur det var att använda AI och vad du lärde dig om ditt eget tänkande.

Problem för lektionen: [Läraren fyller i det specifika problemet här, t.ex. "Hur kan vi minska matsvinnet i skolans cafeteria?"]

Del 1: Problem & Lösning-matris

Använd matrisen för att strukturera ert arbete med problemet. Fyll i både **gruppens egna tankar** och **AI:s förslag** för varje steg i problemlösningen. Detta hjälper er att jämföra idéer och kombinera dem på ett genomtänkt sätt.

Steg i problemlösningen	Gruppens egna tankar	AI:s förslag
Identifiera problemet	Vad är grundorsaken till problemet? Varför är det ett problem?	Vad säger AI om orsaken till problemet?
Tänkbara lösningar	Lista era idéer för att lösa problemet.	Lista de lösningar AI föreslår.
Konsekvenser	Vad kan hända om ni testat dessa lösningar? Finns det risker?	Vad säger AI om konsekvenserna?
Bästa valet och motivering	Vilken lösning väljer ni och varför? Motivera med egna ord.	(AI gör inget val, men notera om dess idéer påverkade er.)

Instruktioner för matrisen:

1. Börja med att diskutera och skriva ner era egna tankar i "Gruppens egna tankar".
2. Be sedan AI (t.ex. ChatGPT) om hjälp, till exempel: "Ge 5 idéer för att minska matsvinn i en skolcafeteria." Skriv AI:s svar i "AI:s förslag".
3. Jämför båda kolumnerna, kombinera idéer och välj den bästa lösningen. Motivera ert val i sista raden.

Del 2: Checklista för kritiskt tänkande

Använd denna checklista för att kontrollera att ni har tänkt kritiskt genom hela processen. Bocka av varje punkt när ni har beaktat den.

Checklista	Bockas av
Har vi ifrågasatt problemet? Är det verkligen så som vi först trodde?	[]
Har vi övervägt flera perspektiv eller lösningar?	[]
Har vi sökt fakta eller information för att stödja våra idéer?	[]
Har vi tänkt på motargument eller risker med våra lösningar?	[]
Baserar vi vårt val på evidens och logik, inte bara på vad som känns rätt?	[]

Tips: Spara checklistan och använd den i andra skoluppgifter för att träna ditt kritiska tänkande!

Del 3: Reflektion

Svara individuellt på dessa frågor för att reflektera över hur det var att använda AI i problemlösningen. Skriv dina svar nedan.

1. **Hur upplevde du samarbetet med AI? Gav det dig nya idéer, eller kändes det mest som upprepning av det ni redan tänkt på?**
[Skriv ditt svar här]
2. **Vad lärde du dig om ditt eget sätt att tänka? Märkte du något om hur du värderar idéer eller löser problem?**
[Skriv ditt svar här]
3. **När tror du att AI kan vara till hjälp i framtida skoluppgifter, och när bör du lita på ditt eget omdöme istället? Ge ett exempel.**
[Skriv ditt svar här]

Avslutning

Bra jobbat! Genom att arbeta med detta arbetsblad har du övat på att använda ditt kritiska tänkande och kombinera det med AI som ett verktyg. Kom ihåg att AI kan ge dig idéer och information, men det är ditt eget omdöme som avgör vad som är bäst. Ta med dig checklistan och använd den i andra ämnen för att bli en ännu bättre problemlösare.

[Valfritt: Extra tankar eller frågor]

[Skriv eventuella funderingar här]

Not till läraren:

- Välj ett problem som engagerar klassen, t.ex. "Hur kan vi minska matsvinnet i skolans cafeteria?" eller något kopplat till deras vardag eller intressen.
 - Se till att eleverna har tillgång till AI (t.ex. ChatGPT) för att hämta förslag. Om det inte går, förbered AI-svar i förväg och dela ut dem.
 - Under presentationerna, ställ frågor som "Varför valde ni det?" eller "Har ni tänkt på X?" för att träna eleverna på att försvara sitt resonemang.
 - Avsluta med en kort klassdiskussion om reflektionsfrågorna för att knyta ihop lektionen och förstärka budskapet om kritiskt tänkande.
-

Lektion 28: AI och komplexa problem – kritiskt tänkande på bred front

Syfte

Denna lektion syftar till att låta eleverna använda AI som ett verktyg för att analysera *komplexa problem* med många dimensioner, till exempel samhällsfrågor eller scenarier utan givna svar. Genom att involvera AI i problemanalysen ska eleverna få upp ögonen för fler perspektiv och möjliga lösningar än de kanske hade kommit på själva. Samtidigt övas deras kritiska tänkande genom att de måste värdera och syntetisera AI:ns bidrag – vad är relevant, vad är pålitligt, vad saknas? Fokus ligger på att inte se AI som en facitgenerator, utan som en **idé- och analysgenerator** som kan hjälpa oss tänka bredare, men som alltid behöver den mänskliga tankekraften för att värdera resultatet cte.ku.edu

Lektionsingång

- **Brainstorm utan AI:** Välj ett aktuellt eller ämnesrelevant komplext problem. Exempel: *“Hur kan vi minska matsvinn i samhället?”* eller *“Vad bör göras åt trafikproblemen i vår stad?”*. Skriv frågan på tavlan. Be eleverna, enskilt eller i par, snabbt spåna alla olika aspekter av problemet de kan komma på (orsaker, konsekvenser, möjliga lösningar, inblandade aktörer, hinder etc.). Efter ett par minuter, samla in förslag muntligt och skriv upp dem i ett moln eller mindmap på tavlan. Poängtera hur många olika delar ett sådant problem faktiskt rymmer.
- **Introducera AI som hjälp:** Fråga klassen retoriskt: *“Tror ni att vi fick med allt nu? Finns det kanske fler perspektiv eller idéer som vi missade?”*. Antagligen finns det saker ingen tänkte på. Föreslå: *“Vi kan prova att ta hjälp av en AI för att se om den kan bidra med fler idéer eller vinklar.”*. Detta väcker förväntan på att AI:n kan fungera som en slags “idébank” eller analytiker som breddar diskussionen.

Instruktioner (steg-för-steg)

1. **Formulera problemet till AI:** Sätt er vid en projektor med en AI-chat eller låt eleverna själva göra det i grupper om de har enheter. Skriv en prompt tillsammans, t.ex.: *“Vi har problemet X. Identifiera olika orsaker till problemet, vilka som berörs, samt möjliga lösningar. Resonera också kring konsekvenser av lösningarna.”*. Använd elevernas matsvinn/trafik-problem eller vad ni valt. Kör prompten.
2. **Låt AI:n presentera analys:** När AI:n ger sitt svar, läs upp det (eller be grupperna läsa sina). Be eleverna lyssna och markera på tavlan de punkter AI:n nämner som *inte* redan fanns i er egen brainstorm. Använd gärna en annan färgpenna för AI:ns tillägg. Nu synliggörs konkret hur AI:n eventuellt tillförde nya perspektiv (kanske AI:n

nämnde ekonomiska aspekter, lagstiftning, eller något globalt perspektiv som klassen missade).

- 3. Diskutera innehållet:** Gå igenom AI:ns punkter en i taget. Ställ frågan: *“Är detta en rimlig/nyttig poäng? Håller vi med?”*. Tillåt kort diskussion. Vissa AI-idéer kanske eleverna tycker är orealistiska eller långsökta – notera det. Andra kanske alla anser viktiga. Markera de mest relevanta insikterna som AI bidrog med.
- 4. Kritisk granskning av AI:ns svar:** Dela ut en checklista eller visa på tavlan: *Källkritiska frågor även för AI-förslag!* Gå igenom exempelvis: **Finns det påstådda “fakta” i AI:ns resonemang som borde kollas?** (om AI t.ex. säger “30% av all mat slängs”, bör man kolla om det stämmer), **Är några perspektiv fortfarande saknade?** (kanske nämnde AI inget om kulturella vanor kring mat, etc.), **Är vissa förslag orealistiska eller etiskt tveksamma?** Låt eleverna i par granska AI-texten igen med dessa kritiska ögon och notera om de hittar något att anmärka på.
- 5. Research-ögonblick:** Om möjligt, ta en av AI:ns faktauppgifter eller kontroversiella påståenden och visa hur man kan verifiera den. T.ex. snabbgoogla det AI:n påstod och se om trovärdiga källor bekräftar eller vederlägger det. Detta förstärker metodiken att *fakta ska kollas oavsett om de kommer från en AI*.
- 6. Eleverna utvecklar lösningar:** Nu när ni har en bredd av idéer (egna + AI:ns) ska eleverna jobba vidare med problemet. Dela in klassen i mindre grupper (3–5 elever). Varje grupp får i uppgift att formulera **en möjlig lösning eller åtgärd** för problemet, men den måste ta hänsyn till flera av de aspekter som kommit upp. T.ex. om problemet var matsvinn, kanske en grupp föreslår “informationskampanjer i skolor” och diskuterar hur det påverkar hushåll (socialt), ekonomi, miljö osv. Uppmuntra grupperna att även här gärna ställa följdfrågor till AI om de vill för att förfinas sin lösning – t.ex. *“Vilka hinder kan finnas för vår lösning X?”* eller *“Har liknande projekt gjorts någon annanstans? Hur gick det?”*. De ska dock inte bara kopiera AI-svar, utan använda dem som diskussionsunderlag inom gruppen.
- 7. Redovisning av lösningar:** Låt varje grupp presentera sin lösning och kort förklara hur de tog hänsyn till olika aspekter. Be dem också nämna om AI bidrog med något i deras process (det kan vara både idéer de tog med och idéer de valde att förkasta!). Klassen och du som lärare kan ställa följdfrågor, gärna utmanande: *“Hur skulle er plan funka om X händer?”* etc., för att fortsätta det kritiska tänkandet.
- 8. Sammanfatta lärdomar:** Fråga eleverna: *“Hjälpte AI oss att tänka bredare? På vilket sätt?”* och *“Vad måste vi vara försiktiga med när vi tar hjälp av AI i sådana här komplexa frågor?”*. Svaren ska förhoppningsvis landa i att AI kan vara en *stimulator* för idéer och hjälpa oss se fler sidor av ett problem, men att **människor måste stå för värderingen, etiken och faktakollen**. Poängtera att komplexa problem sällan har

enkla svar – inte ens AI kan trolla fram en lösning, men den kan ge oss pusselbitar att ta ställning till.

Didaktiska tips

- **Relevans och engagemang:** Välj problem som engagerar eleverna. Det kan vara något i deras närmiljö (skolan, kommunen) eller ett större globalt dilemma som ni berört i undervisningen. Ju mer investerade de är i frågan, desto mer nyfikna blir de på AI:ns input och på att själva hitta en bra lösning.
- **Tydliggör AI:ns roll:** I denna övning agerar AI som *bollplank* och informationskälla, inte som orakel. Se till att eleverna förstår skillnaden. Du kan använda analogin: AI är som en väldigt påläst kompis som dock ibland hittar på saker – bra att bolla idéer med, men inte alltid att lita på utan kontroll.
- **Tidshantering:** Komplexa problem kan sluka tid. Var beredd att avgränsa. Kanske fokuserar ni bara på *orsaker och lösningar*, och tar inte med *konsekvenser*, om tiden är kort. Eller så fördelar du olika underfrågor till olika grupper (en grupp djupdyker på orsaker, en på lösningar etc.) för att sedan pussla ihop. AI:n kan i så fall användas i omgångar av grupperna separat.
- **Multi-modala inputs:** För att ge ännu rikare analys kan du låta vissa grupper eller hela klassen även använda AI för att hämta *flera perspektiv*. Exempel: Be AI agera *“djävulens advokat”* mot en viss lösning, eller generera ett litet dialogspel: *“Skriv en kort debatt mellan en miljöaktivist och en ekonom om trafikproblemets lösningar”*. Detta kan göra det tydligt att det finns motstridiga intressen. Eleverna kan dramatisera sådana AI-genererade dialoger för att levandegöra perspektiven.
- **Etiska överväganden:** Vissa komplexa problem kan ha etiska dimensioner (t.ex. AI föreslår något kontroversiellt som övervakning för att lösa brottslighet). Var förberedd att hantera sådana förslag och använd dem som chans att prata om etik: *“Bara för att AI:n föreslår X, betyder det att vi bör göra det? Varför kan det vara problematiskt?”*. Lär eleverna att väga fördelar mot värdegrund.

Diskussionsfrågor

- *“Kan AI verkligen förstå ett problem såsom vi människor gör, eller sammanställer den bara information? Hur märkte ni av det under övningen?”*
- *“Fick ni förslag från AI som ni valde att inte använda? Varför valde ni bort dem?”*
- *“Hur skiljer sig AI:ns sätt att angripa problemet från vårt sätt? Kommer den med annorlunda typer av idéer än vi?”*

- “På vilket sätt kände ni att ni behövde vara källkritiska eller skeptiska till AI:ns input i den här uppgiften?”
- “Tror ni att ni skulle komma på lika många perspektiv utan AI? Hur kan man annars få in fler idéer när man löser problem?” (T.ex. genom research, intervjuar experter, grupparbete – jämför dessa metoder med AI-hjälp.)

Forskning och arbetsblad

- **Forskning:** Du kan nämna att studier i klassrumsmiljö har sett positiv effekt av att arbeta med data och AI på detta sätt. Till exempel visade *Thinking with Data*-projektet att elever som lärde sig analysera information och data mer systematiskt fick bättre resultat i flera skolämnen

the74million.org

. Att ta in AI som hjälp för att bryta ned komplexa frågor ligger i linje med att utveckla datalogiskt och kritiskt tänkande. Samtidigt understryker många pedagoger att elevens eget resonemang måste stå i centrum – AI ska inte ersätta det, utan förstärka det. Som en professor uttryckte det: “AI kan ge oss förslag, men vi måste göra det tunga tänkarbetet att bedöma och välja”.

- **Arbetsblad:** För den här lektionen passar det bra med en **problemanalys-mall**. En sida kan t.ex. ha rubrikerna: *Problem:* ____, *Möjliga orsaker:* ____, *Påverkade parter:* ____, *Möjliga lösningar:* ____, *Konsekvenser/för- och nackdelar:* ____. Eleverna kan fylla i denna tabell varefter de får in idéer (egna och från AI). Man kan också ha en kolumn som markeras AI vs egna idéer. Detta hjälper eleverna att strukturera komplex information. Alternativt kan varje grupp få ett stort papper och färgpennor för att rita en mindmap under processens gång, där de kan cirkla de AI-förslag de använt. Dokumentation är viktig så att redovisningen blir tydlig och så att eleverna själva ser hur mycket de flesat ut problemet.

Arbetsblad – Lektion 28: AI och komplexa problem – kritiskt tänkande på bred front

Introduktion

I denna lektion ska du och din grupp analysera ett komplext problem med hjälp av AI. Syftet är att använda AI som ett verktyg för att hitta nya perspektiv och idéer, men det är viktigt att förstå att AI inte ger färdiga svar. Istället fungerar den som en "idébank" som kan bredda ditt tänkande. Det är upp till dig att kritiskt granska AI:ns förslag – vad är relevant, pålitligt och användbart? Genom att kombinera dina egna idéer med AI:ns input kommer du att öva på att tänka systematiskt och nyanserat kring svåra frågor.

Problem för lektionen:

[Läraren fyller i det specifika problemet här, t.ex. "Hur kan vi minska matsvinnet i samhället?" eller "Vad bör göras åt trafikproblemen i vår stad?"]

Del 1: Egen brainstorm

Börja med att brainstorma i gruppen utan AI. Lista allt ni kan komma på kopplat till problemet: orsaker, konsekvenser, lösningar, inblandade aktörer och hinder. Fyll i tabellen nedan.

Kategori	Era egna idéer
Orsaker	
Konsekvenser	
Möjliga lösningar	
Inblandade aktörer	
Hinder eller utmaningar	

Del 2: AI som idébank

Använd en AI-tjänst (t.ex. ChatGPT) för att få fler vinklar på problemet. Skriv en tydlig prompt tillsammans i gruppen, till exempel:

- "Vi har problemet [problem]. Identifiera orsaker, konsekvenser, vilka som berörs, möjliga lösningar och eventuella hinder. Resonera också kring vad lösningarna kan leda till."
Skriv in prompten i AI:n och fyll i tabellen nedan med dess svar.

Kategori	AI:ns förslag
Orsaker	
Konsekvenser	
Möjliga lösningar	
Inblandade aktörer	
Hinder eller utmaningar	

Del 3: Jämför era idéer med AI:ns

Titta på AI:ns svar och jämför dem med era egna idéer från Del 1.

- **Vad tillförde AI:n som ni inte hade tänkt på?**
[Skriv här vilka nya perspektiv eller idéer AI:n bidrog med]
- **Är alla AI:ns förslag vettiga? Varför/varför inte?**
[Skriv här vad ni tycker om AI:ns idéer och om något verkar konstigt eller oanvändbart]

Del 4: Kritisk granskning

Granska AI:ns svar med ett källkritiskt öga. Använd checklistan nedan och skriv era reflektioner.

Fråga	Era anteckningar
Finns "fakta" som behöver kollas? (T.ex. siffror eller påståenden)	
Saknas några viktiga vinklar? (T.ex. kulturella eller lokala aspekter)	
Är något orealistiskt eller etiskt tveksamt?	
Hur kan ni verifiera AI:ns information? (T.ex. googla, kolla forskning)	

Del 5: Skapa en lösning

Nu ska ni ta det bästa från era egna idéer och AI:ns förslag för att skapa en lösning på problemet. Tänk på olika aspekter (orsaker, konsekvenser, aktörer, hinder) när ni formar förslaget. Skriv ner er lösning och varför ni valde den.

Er lösning Motivering

Extra steg (frivilligt): Ställ en följdfråga till AI:n för att förbättra er lösning, t.ex. "Vilka hinder kan finnas för vår idé?" eller "Finns exempel på liknande lösningar?". Anteckna vad AI:n säger och om det påverkar er plan.

Del 6: Förbered en presentation

Förbered en kort presentation av er lösning för klassen. Ta med:

- Vad er lösning är
 - Hur ni tänkte kring olika aspekter av problemet
 - Hur AI hjälpte er (både idéer ni använde och idéer ni ratade)
-

Del 7: Reflektera

Efter att ha hört klassens presentationer, fundera själv över dessa frågor:

1. **Hjälpte AI er att se problemet från nya håll? Hur då?**
[Skriv ditt svar här]
 2. **Vad behöver ni vara försiktiga med när ni använder AI för såna här frågor?**
[Skriv ditt svar här]
 3. **Hur kan ni använda AI och ert eget tänkande tillsammans framöver?**
[Skriv ditt svar här]
-

Avslutning

Bra jobbat! Du har nu använt AI som ett verktyg för att analysera ett komplext problem och kombinerat det med ditt eget kritiska tänkande. Spara detta arbetsblad som en påminnelse om hur du kan tackla svåra frågor i framtiden.

[Valfritt: Tankar eller frågor]

[Skriv eventuella funderingar här]

Till läraren:

- **Problemval:** Välj ett problem som känns relevant för eleverna, gärna kopplat till deras vardag eller ämnen ni jobbat med.
- **Teknik:** Se till att det finns tillgång till AI (via datorer eller projektor).
- **Stöd:** Hjälp eleverna med prompts och att granska AI:ns svar kritiskt under arbetet.
- **Avslutning:** Håll en kort klassdiskussion om vad AI bidrog med och vilka utmaningar det gav.

Lektion 29: Höja nivån på analysen med AI

(Att använda AI som “skruvdragare” för storskaliga textanalyser och retoriska studier)

Introduktion

Eleverna har redan övat på grundläggande källkritik, retorisk analys och kritiskt tänkande. Nu är det dags att **ta det ett steg vidare** – att använda AI för att göra **mer omfattande, storskaliga** analyser än man normalt hinner för hand. Jämför det med en **skruvdragare** i stället för en **skruvmejsel**: vi lär oss för hand först, men med en motoriserad hjälp kan vi göra mycket mer på kortare tid.

Syfte och mål

1. **Visa hur AI** kan användas för att analysera en stor mängd texter (artiklar, inlägg, tal) på relativt kort tid.
2. **Höja “scopet”**: Eleverna får undersöka exempelvis hur ofta kvinnor nämns i ett antal tidningar, hur influencers använder retorik i flera inlägg, eller hur olika åsikter framträder i en hel valrörelse.
3. **Poängtera** att AI är ett **verktyg**, inte en ersättning för egen analys – men att det kan bredda perspektiven och ge oss data vi annars inte skulle hinna samla.

Uppläggsförslag: Flera nivåer av AI-användning

1. **Skruvmejseln** (Manuellt arbete)
 - Eleverna har lärt sig göra retoriska analyser eller nyhetsgranskning utan AI. De kan i princip kontrollera några artiklar eller inlägg.
2. **Skruvdragaren** (Ett AI-verktyg)
 - Eleverna använder en chattbot (ex. ChatGPT) för att analysera *fler* texter effektivt, men med lärarens stöttning och egen källkritik i grunden.
3. **Byggföretaget** (Flera AI-tjänster i samspel)
 - Eleverna instruerar olika AI-verktyg (chattbotar, dataanalysverktyg, bildgeneratorer) för att få fram en **mer komplex studie** – t.ex. en hel rapport om hur en viss retorisk figur används i nationell politik under en valkampanj.

Förslag på konkreta aktiviteter

1. Räkna förekomst av nyckelord över flera artiklar

1. **Förbered material**: Välj t.ex. 10–20 korta nyhetsartiklar om samma ämne (klimat, ekonomi, jämställdhet).
2. **Grupparbete**: Låt eleverna först göra en liten manuell undersökning (skruvmejsel): “Hur ofta nämns kvinnor vs. män i dessa 2 artiklar?”
3. **AI-verktyg**: Därefter klistrar de in fler artiklar (eller länkar) i en chattbot och ber AI:n räkna hur ofta vissa ord förekommer (“kvinna”, “man”, “flicka”, “pojke”, etc.), eller hur vissa retoriska fraser används.

4. **Jämför resultat:** Diskutera skillnaden i **omfång** mellan manuell och AI-stödd analys. Vad kan AI ha missförstått? Hur dubbelkollar vi?

2. Influencers retorik i flera inlägg

1. **Välj en influencer** eller en känd person på sociala medier.
2. **Hämta text:** Elever samlar 5–10 inlägg/captions och klistrar in dem i en chattbot.
3. **Testa att analysera video:** I google AI-studio (om skolan har åtkomst med google konto) kan man analysera såväl tal som video. Det innebär att man kan analysera såväl verbal som icke-verbal kommunikation på skala,
4. **Retorisk analys:** Eleverna ger AI en prompt: “Analysera retoriska grepp och tonläge i dessa inlägg. Hur ofta används t.ex. känslargument?”
5. **Fördjupning:** Be AI att göra en **sammanställning** över alla inlägg – “vilka teman återkommer?” Eleverna diskuterar om influencerna visar en enhetlig bild av sig själva, hur “omslag” av åsikter dyker upp, eller om språket skiljer sig över tid.

3. Hel valrörelse

1. **Scenario:** Föreställ er att ni vill se hur ett visst politiskt parti pratade under senaste valrörelsen.
2. **Datasamling:** Samla officiella valmanifest, debattartiklar, tal, sociala medie-inlägg.
3. **AI-analys:**
 - Först prövar eleverna själva läsa **ett** tal och gör egen retorisk analys.
 - Sedan låter de AI analysera **flera** texter samtidigt, för att se återkommande ämnen, ordval, emotionella budskap.
4. **Resultat:** Eleverna diskuterar hur AI gav en större helhetsbild. Samtidigt granskar de kritiskt om AI “missade” kontext eller nyanser.

Diskussionspunkter

1. **Hur mycket litar vi på AI?**
 - AI kan vara snabb och hantera mängder av text. Men hur validerar vi att AI räknar eller tolkar saker korrekt?
2. **Varför är egen förståelse fortfarande viktig?**
 - Elever måste förstå retoriska begrepp, argumentationsanalys och källkritik *innan* de låter AI göra grovjobbet.
3. **Tidseffektivitet vs. djup**
 - När vi låter AI analysera 50 artiklar kan vi få bredare översikter, men riskerar att förlora djup detaljgranskning. Hur balanserar vi detta?
4. **Etiska aspekter**
 - Samla data om människor. Får man kopiera influencers inlägg rakt av? Diskutera upphovsrätt och integritet.

Lektionsexempel: “Kvinnor i nyhetsmedier”

1. **Inledning:** Diskutera hur nyhetsmedier skildrar kvinnor kontra män. Har eleverna märkt av en skillnad?

2. **Manuell mini-analys** (5–10 min):
 - Eleverna tar en **kort** artikel, räknar hur många gånger “kvinna” eller kvinnliga namn används vs. “man” eller manliga namn.
3. **Storskalig AI-analys** (15–20 min):
 - Eleven klistrar in **5–10 artiklar** (eller länkar) i en chattbot, med en prompt: “Analysera hur många gånger det refereras till män vs. kvinnor. Notera också om texten associerar kvinnor med vissa ämnen eller roller.”
 - Diskutera skillnader: Fick ni samma resultat för hand och med AI? Var AI:s resultat tydligare, eller missade den något?
4. **Reflektion** (5–10 min):
 - Hur kan detta användas i större skala? T.ex. “Hur skildrar tidning X vs. tidning Y kvinnor i sportjournalistiken?”
 - Var AI-analysen tillräcklig, eller behövde vi mer kontext för att förstå siffrorna?

Fördjupning: Kombinera AI-tjänster (Byggföretaget)

1. **Chattbot för textanalys**
 - T.ex. ChatGPT för att hitta nyckelord, tonläge, retoriska medel.
2. **Datavisualisering**
 - Använd eventuellt ett enkelt visualiseringsverktyg (ex. Google Sheets) för att göra grafer över förekomster.
3. **Rapportskrivning**
 - Samla resultaten i en “minirapport” där eleverna själva **tolkar** siffrorna och insikterna. Kanske även generera en kort sammanfattning med AI.
4. **Slutdiskussion:** AI utför massbearbetning, men **människor** måste ge tolkningen och slutsatserna – en viktig poäng.

Sammanfattning

I denna lektion (eller serie av lektioner) lär sig eleverna att **höja nivån** på analysen med AI – gå från att manuellt läsa enstaka artiklar till att AI-bearbeta **många** källor för att hitta mönster i språk, retorik eller representation av kvinnor/män. Eleverna får se **fördelarna** (snabbare, större datamängd) och **nackdelarna** (AI kan missa nyanser, kräver mänsklig validering). Poängen är att AI är ett **hjälpmedel** – som en elektrisk skruvdragare jämfört med en skruvmejsel. Man bör inte **sluta göra** analyser på egen hand, men AI kan **komplettera** och **förstärka** vår förmåga, särskilt när vi vill se **det stora perspektivet**.

Arbetsblad – Lektion 29: Höja nivån på analysen med AI

Att använda AI som “skruvdragare” för storskaliga textanalyser och retoriska studier


1. Inledning: AI som analysverktyg

Vi har tidigare övat på källkritik, retorisk analys och textgranskning **för hand**. Nu ska vi ta ett steg vidare och använda **AI som ett verktyg för storskalig analys**.

Varför använda AI?

- AI kan snabbt analysera stora mängder text och hitta mönster.
- Vi kan jämföra olika texter och hitta dolda trender i språk och innehåll.
- AI kan spara tid – men vi måste fortfarande **kritiskt granska resultaten**.

Analogi: Skruvmejsel vs. skruvdragare

 För att förstå skillnaden kan vi tänka på **manuell analys** som en **skruvmejsel** – långsam, noggrann och detaljerad. AI fungerar mer som en **skruvdragare** – snabbare och effektivare, men vi måste styra den rätt.

2. Övning: Räkna förekomst av nyckelord i nyhetsartiklar

Steg 1: Manuell analys (5–10 min)

1. Läs en kort nyhetsartikel.
 2. Räkna hur många gånger vissa ord förekommer, exempelvis:
 - “man” vs. “kvinna”
 - “miljö” vs. “ekonomi”
 - “trygghet” vs. “brott”
 3. Skriv ner era resultat och diskutera:
 - Vad verkar vara mest framträdande?
 - Är texten vinklad åt något håll?
-

Steg 2: AI-analys av flera artiklar (10–15 min)

1. Kopiera in **5–10 artiklar** (eller utdrag) i en AI-chattbot (ex. ChatGPT).
2. Använd följande **prompt**:

“Analysera dessa artiklar och räkna hur många gånger följande ord förekommer: ‘man’, ‘kvinna’, ‘flicka’, ‘pojke’. Beskriv även om det finns mönster i hur dessa ord används.”

3. Jämför AI:s analys med er manuella analys.
 - Stämde AI:s siffror överens med er egen?
 - Fick AI fram mönster som ni inte såg själva?
 - Fanns det något AI kanske missförstod?

✦ Reflektion:

- Är AI snabbare än vi?
- Är AI:s analys alltid korrekt, eller behövs mänsklig granskning?

3. Övning: Influencers retorik i flera inlägg

Steg 1: Samla material

1. Välj en influencer eller en känd person.
2. Kopiera **5–10 inlägg eller captions**.

Steg 2: AI-analys

1. Använd följande **prompt** i en AI-chattbot:

“Analysera dessa inlägg. Vilka retoriska grepp används? Hur ofta används känslolagring eller överdrifter?”

2. Diskutera AI:s analys:
 - Använder influencern liknande språk i alla inlägg?
 - Försöker inläggen skapa en viss känsla (ex. oro, hype, trygghet)?
 - Finns det tydliga mönster i hur influencern påverkar sin publik?

✦ Reflektion:

- Hur kan denna typ av analys hjälpa oss att förstå **påverkan i sociala medier**?
- Hur skulle vi kunna använda AI för att jämföra **olika influencers**?

4. Övning: AI och politisk analys

Steg 1: Analysera en politisk text själv

1. Läs en debattartikel, tal eller partiprogram.
2. Skriv ner:
 - **Vilka argument används?**
 - **Vilka känslor försöker texten skapa?**
 - **Vilka nyckelord upprepas?**

Steg 2: AI-analys av flera texter

1. Kopiera in **flera politiska texter** i en AI-chattbot.

2. Använd följande **prompt**:

“Analysera dessa politiska texter. Vilka teman återkommer? Vilka retoriska knep används? Vilka känslor försöker de skapa?”

✦ **Jämför AI:s analys med er egen:**

- Vilka mönster såg AI som ni kanske missade?
- Var AI:s analys mer ytlig eller mer detaljerad?

✦ **Diskutera:**

- Kan AI hjälpa oss att se **dolda politiska budskap**?
- Finns det risk att AI **missförstår eller förenklar** komplexa ämnen?

5. Avslutande reflektion & diskussion

💡 **Diskussionsfrågor:**

1. Hur pålitlig är AI när vi gör stora analyser?
2. Kan vi lita på att AI tolkar text korrekt?
3. Är AI bättre än människor på att hitta mönster i språk?
4. Vad är riskerna med att låta AI “bestämma” analysen åt oss?
5. Hur kan AI hjälpa oss att bli bättre på kritiskt tänkande?

📌 **Redovisning:**

- Varje grupp berättar om sin AI-analys:
 - Vad fungerade bra?
 - Vad fungerade mindre bra?
 - Hur kan AI förbättras som analysverktyg?

📄 **Skriftlig uppgift:**

- Skriv en kort text om **hur AI kan användas som ett analysverktyg** och vilka för- och nackdelar det innebär.

EXTRAUPPGIFT – Analysera verbal och icke-verbal kommunikation i Google Studio:

1. Inledning: AI som analysverktyg

Vi har tidigare arbetat med källkritik och textanalys manuellt. Nu ska vi utforska hur AI kan användas för att effektivisera och fördjupa dessa analyser, både för text och video.

Varför använda AI?

- **Effektivitet:** AI kan snabbt bearbeta stora mängder data.
 - **Djupare insikter:** Genom att analysera både text och icke-verbal kommunikation kan vi få en mer komplett förståelse.
 - **Kritisk granskning:** Trots AI:s kapacitet måste vi alltid tolka resultaten med ett kritiskt öga.
-

2. Övning: Analysera influencers retorik i text och video

Steg 1: Samla textmaterial

1. Välj en influencer eller känd person på sociala medier.
2. Samla 5–10 textinlägg (t.ex. captions) från deras profiler.

Steg 2: AI-baserad textanalys

1. Använd en AI-chattbot för att analysera texterna.
 - **Prompt:** "Analysera retoriska grepp och tonläge i dessa inlägg. Hur ofta används t.ex. känslargument eller överdrifter?"
2. Diskutera AI:s analys:
 - Vilka retoriska tekniker används frekvent?
 - Finns det återkommande teman eller budskap?

Steg 3: Samla videomaterial

1. Välj 2–3 videoinlägg från samma person.

Steg 4: AI-baserad videoanalys med Google AI Studio

Vad är Google AI Studio?

Google AI Studio är en webbaserad plattform som låter användare snabbt experimentera med och bygga AI-drivna applikationer. Den möjliggör analys av olika datatyper, inklusive text, bild, ljud och video, vilket ger en djupare förståelse för innehållets kommunikation.

Hur använder man Google AI Studio för videoanalys?

1. **Åtkomst:**
 - Besök [Google AI Studio](#) och logga in med ditt Google-konto.
2. **Ladda upp video:**
 - Navigera till videoanalyssektionen och ladda upp den valda videon.
3. **Analys:**
 - Använd tillgängliga verktyg för att analysera icke-verbal kommunikation, såsom ansiktsuttryck, gester och kroppsspråk.
4. **Tolkning:**
 - Granska resultaten och notera hur den icke-verbala kommunikationen kompletterar eller kontrasterar textinnehållet.

Steg 5: Jämför och reflektera

1. Jämför text- och videoanalyserna:

- Hur överensstämmer den verbala och icke-verbala kommunikationen?
- Finns det diskrepanser mellan budskapen i text och video?

2. Diskutera i grupp:

- Hur påverkar dessa kommunikationssätt publikens uppfattning?
 - Vilken roll spelar icke-verbal kommunikation i att förstärka eller förändra budskapet?
-

3. Avslutande reflektion & diskussion

Diskussionsfrågor:

1. Hur kan AI-baserad analys hjälpa oss att förstå påverkan från sociala medier?
2. Vilka är fördelarna och begränsningarna med att använda AI för text- och videoanalys?
3. Hur kan vi säkerställa att våra tolkningar av AI:s analys är korrekta och objektiva?

Redovisning:

- **Grupppresentation:**
 - Sammanfatta era fynd och diskutera hur AI-verktyg bidrog till analysen.
 - **Skriftlig reflektion:**
 - Skriv en kort uppsats om hur AI kan användas för att analysera digital kommunikation och vilka etiska överväganden som bör tas i beaktande.
-

6. Sammanfattning: Vad lärde vi oss?

🎯 **AI är ett kraftfullt verktyg – men vi måste använda det rätt.**

✅ **Snabbhet:** AI kan analysera stora mängder text på kort tid.

✅ **Mönsterigenkänning:** AI kan hitta återkommande teman och nyckelord.

❌ **Begränsningar:** AI kan missa sammanhang, nyanser och ironi.

❌ **Kritisk granskning krävs:** Vi måste **alltid dubbelkolla AI:s analys.**

📌 **Viktig slutsats:**

AI ska inte ersätta **vårt eget tänkande**, utan användas **som ett verktyg** för att hjälpa oss att se **större mönster och nya perspektiv.**

💡 **Fundera vidare:**

- Hur kan vi använda AI **för att analysera mediebilder av olika grupper?**
- Skulle vi kunna skapa en **helt AI-genererad rapport** – eller behövs människor alltid för att tolka slutsatserna?

🔍 **Slutpoäng:**

AI kan vara vår skruvdragare – men vi måste hålla i handtaget! 🛠️