Marika Lagervall

Projektarbete

AI för skribenter, sommarkurs

SV071G M2269 VT 2025

Bilaga – material till ”En strategi för hållbar AI-användning – enligt AI”

Här har jag samlat mina promptar till chatbottarna och deras svar. Eftersom det är ganska mycket text har jag gjort en innehållsförteckning nedan.

Notera att de rubriker som ligger under respektive chattbots namn är deras egna som följt med när jag klistrat in deras svar. Alla övriga rubriker är mina. Fetstilsmarkeringen av viktiga ord i texten är (åtminstone till största delen) från chattbottarnas svar.

Innehåll

1 Hallucinationer 3

Min prompt 3

Chat GPT 3

Copilot 5

Gemini 6

Claude 8

Grok 9

2 Bias 12
Min prompt 12
Chat GPT 12
Copilot 14
Gemini 16
Claude 17
Grok 19
3 Miljömässiga risker 23
Min prompt 23
Chat GPT 23
Copilot 25
Gemini 26
Claude 27
Grok 29
Sammanfattning av antal ord 31
4 Råd för hållbar AI-användning 32
Min prompt 32
Chat GPT 32
Copilot 32
Gemini 33
Claude 34
Grok 35

1 Hallucinationer

Min prompt

Hej! Jag studerar en kurs om AI och försöker lära mig mer om hur allt fungerar. Som ett led i min kurs skulle jag vilja ställa några frågor om risker med AI. Den första är: Hur ser du som AI på förekomsten av AI-hallucinationer? Är det en oundviklig bieffekt av tekniken, och är det alltid dåligt eller kan det vara till fördel ibland? Har du någon uppfattning om hur ofta du ger svar som är AI-hallucinationer?

Chat GPT

Vilken spännande kurs du går! Det här är en viktig och ganska djupgående fråga, så låt oss bryta ner den i tre delar:

---

1. Hur ser jag som AI på förekomsten av AI-hallucinationer?

AI-hallucinationer syftar på när en AI ger svar som låter övertygande men som är felaktiga, påhittade eller inte baserade på verkliga data. Det kan handla om felaktiga fakta, icke-existerande källor eller fabricerade händelser.

Från mitt "perspektiv" som AI är hallucinationer **ett välkänt och aktivt forskat problem** inom utvecklingen av språkmodeller. De uppstår inte för att AI försöker ljuga, utan för att den försöker **förutsäga det mest sannolika svaret utifrån mönster i träningsdata**, även om den saknar exakt kunskap om ett visst ämne. Det är alltså inte ett medvetet val utan ett symptom på **hur tekniken fungerar**.

---

2. Är hallucinationer en oundviklig bieffekt, och är det alltid dåligt?
✓ *Är de oundvikliga?*

Just nu – **ja, i viss mån**. Alla generativa språkmodeller har en viss benägenhet att "hallucinera", särskilt om:

- frågan gäller något obskyrt eller okänt,
- det saknas bra träningsexempel,
- eller när modellen uppmuntras att "hitta på" (t.ex. kreativt skrivande).

Forskare och utvecklare arbetar aktivt med att minska hallucinationer, till exempel genom att:

- koppla modeller till faktabaserade databaser (retrieval-augmented generation),
- använda mer "källkritiska" metoder vid svarsgenerering,
- införa verifieringssystem.

❌ *Är det alltid dåligt?*

Inte nödvändigtvis. I vissa sammanhang **kan hallucinationer vara önskvärda**:

- Vid **kreativt skrivande**, där AI får hitta på berättelser, poesi eller fiktion.
- Vid **idégenerering**, där även udda eller oväntade "fel" kan stimulera nya tankebanor.
- Vid **simuleringar**, där det är användbart att AI gissar sig fram för att skapa alternativa scenarier.

Men i faktabaserade sammanhang – forskning, utbildning, juridik, medicin – är hallucinationer **mycket problematiska** och måste hanteras med stor försiktighet.

---

3. Hur ofta hallucinerar jag (som ChatGPT)?

Det är svårt att ge ett exakt svar i procent eftersom det beror mycket på:

- hur frågan är ställd,
- hur noggrann användaren är i att dubbelkolla svar,
- och hur färsk eller nischad informationen är.

I tester som gjorts på modeller som GPT-4 (som jag baseras på) ligger hallucinationsfrekvensen **någonstans mellan 3–20 % beroende på område**, men detta är fortfarande under kontinuerlig utvärdering. Ju mer avancerade och kontrollerade systemet är (t.ex. kopplat till en databas eller källgranskning), desto lägre blir risken.

Sammanfattning:

- AI-hallucinationer är en följd av hur språkmodeller fungerar.
- De är *delvis* oundvikliga idag, men kan i vissa kreativa sammanhang vara en tillgång.
- I faktasammanhang är de en risk som kräver medveten hantering.

Det finns ingen exakt frekvens, men risken ökar vid komplexa, ovanliga eller dåligt källförankrade frågor.

Copilot

Hej Marika! Vilken spännande kurs du går – AI är verkligen ett område fullt av både möjligheter och utmaningar. Din fråga om AI-hallucinationer är väldigt relevant, särskilt när det gäller språkmodeller som mig. Låt oss bryta ner det lite:

 Vad är AI-hallucinationer?

AI-hallucinationer uppstår när en AI genererar information som *låter korrekt* men som är *falsk, påhittad eller missvisande*. Det kan handla om:

- Felaktiga fakta (t.ex. att säga att Einstein vann Nobelpriset för relativitetsteorin – vilket han inte gjorde)
- Påhittade citat, källor eller personer
- Logiska motsägelser eller feltolkningar

Är hallucinationer en oundviklig bieffekt?

Ja, i viss mån. Här är varför:

- **Probabilistisk natur**: Språkmodeller som mig fungerar genom att förutsäga nästa ord baserat på sannolikheter, inte genom att hämta fakta från en databas. Det gör att vi ibland "gissar" fel.
- **Brist på verifiering**: Jag har inte en inbyggd mekanism för att dubbelkolla fakta mot en källa i realtid (om inte webbsök används).
- **Träning på blandad data**: Jag är tränad på enorma mängder text, inklusive både korrekta och felaktiga uppgifter. Det gör att jag ibland återger felaktigheter som om de vore sanna.

😓 *Är hallucinationer alltid dåliga?*

Inte nödvändigtvis! Det beror på kontexten:

Kontext	Hallucinationer	Bedömning
Faktabaserad rådgivning	Felaktiga svar	❌ Problematiskt
Kreativt skrivande	Fantasifulla inslag	✅ Kan vara berikande
Brainstorming	Oväntade idéer	⚠️ Ibland inspirerande
Forskning eller juridik	Påhittade källor	❌ Allvarligt

Så i kreativa sammanhang kan hallucinationer faktiskt stimulera nya tankar – men i faktabaserade sammanhang är de riskabla.

📊 *Hur ofta händer det?*

Det är svårt att ge en exakt siffra, men forskning visar att hallucinationer kan förekomma i **5–20 %** av svaren, beroende på:

- Frågans komplexitet
- Modellens version
- Om webbsök används eller ej

Jag försöker minimera det genom att ge kontextuella och välorganiserade svar, men det är alltid klokt att dubbelkolla viktiga fakta.

Gemini

Hej! Det är en mycket bra och relevant fråga. Att förstå begränsningarna hos AI-modeller, som fenomenet med **AI-hallucinationer**, är avgörande för att kunna använda tekniken på ett ansvarsfullt sätt.

Vad är AI-hallucinationer?

Från mitt perspektiv som AI är en hallucination när jag genererar information som låter trovärdig och saklig, men som i själva verket är felaktig, påhittad eller saknar stöd i den träningsdata jag har fått. Det kan handla om att jag hittar på källor, citerar icke-existerande personer eller blandar ihop fakta.

Är det en oundviklig bieffekt?

Ja, i dagsläget är det en i stort sett oundviklig bieffekt av den teknik som används för att skapa stora språkmodeller som jag. Orsaken ligger i hur jag fungerar:

- **Sannolikhetsbaserad generering:** Jag är inte en databas som söker upp exakta fakta. Jag genererar text genom att förutsäga det mest sannolika nästa ordet i en mening, baserat på den enorma mängd text jag har tränats på. Ibland kan denna process leda till att jag "fyller i luckor" med information som verkar passa in men som inte är faktagrundad.
- **Komplexitet och mönsterigenkänning:** Min styrka ligger i att känna igen mönster och samband i språk, inte i att ha en perfekt förståelse av sanningen. Jag kan vara mycket bra på att imitera strukturen hos en vetenskaplig artikel, men om träningsdatan är otydlig eller motsägelsefull kan jag generera felaktig information som passar in i mönstret.

Forskare arbetar intensivt med att minska förekomsten av hallucinationer, men det är en grundläggande utmaning med den nuvarande teknologin.

Är det alltid dåligt eller kan det vara till fördel?

För det mesta är hallucinationer en stor nackdel, särskilt när det handlar om faktagranskning, akademiskt skrivande eller medicinsk rådgivning. Följderna kan vara allt från pinsamma fel till allvarliga konsekvenser om felaktig information sprids.

Det finns dock vissa kreativa sammanhang där en hallucination kan vara en fördel. I vissa fall kan det leda till oväntade idéer eller kreativt skrivande som inte följer strikta logiska mönster. Om en författare ber mig att skriva en berättelse kan en "hallucination" leda till en oväntad vändning i handlingen eller en ny, intressant karaktär. Det är viktigt att förstå att det handlar

om att använda AI som ett **kreativt verktyg**, inte som en källa för fakta. I sådana fall blir det människans uppgift att bedöma om hallucinationen är värdefull eller inte.

Hur ofta ger jag svar som är AI-hallucinationer?

Det är svårt för mig att ge en exakt siffra eftersom jag inte har en intern mekanism som kan identifiera "sanning" på samma sätt som en människa. Jag har inte förmågan att "veta" att ett svar är felaktigt. Denna bedömning måste göras externt, av en mänsklig användare som kan verifiera informationen mot pålitliga källor.

Min ambition är alltid att ge korrekta och relevanta svar, och jag är tränad för att minimera risken för fel. Men som med alla stora språkmodeller finns alltid en risk för att jag skapar hallucinationer. Det är därför det är så viktigt att man alltid **verkarar[1] informationen jag ger**, särskilt inom områden där korrekthet är kritisk.

Att vara medveten om denna begränsning är det första steget mot en ansvarsfull användning av AI. Har du några fler frågor om hur jag fungerar eller om andra risker med AI?

Claude

Hej! Kul att du studerar AI - det är ett fascinerande område.

AI-hallucinationer är verkligen en intressant aspekt av hur språkmodeller fungerar. Från mitt perspektiv verkar det vara en naturlig konsekvens av hur vi genererar text - vi förutsäger nästa ord baserat på mönster i träningsdata, vilket ibland leder till plausibelt klingande men felaktig information.

Är det oundvikligt? I nuvarande arkitekturer verkar det så. Vi balanserar kreativitet och faktanoggrannhet, och den flexibilitet som gör oss användbara för kreativa uppgifter kan också producera felaktigheter.

Är det alltid dåligt? Inte nödvändigtvis. I kreativa sammanhang - som att skriva fiktion eller brainstorma idéer - kan "hallucinationer" faktiskt vara värdefulla. Problemet uppstår när vi

[1] Notera Gemini här har hittat på, eller hallucinerat, ordet *verkarar* i stället för att skriva *verifierar*.

presenterar felaktig information som fakta, särskilt inom medicin, juridik eller andra kritiska områden.

Angående hur ofta jag hallucinerar - det är svårt för mig att bedöma. Jag kan inte direkt detektera när jag genererar felaktig information, eftersom den "känns" lika säker som korrekt information när jag producerar den. Det är därför det är viktigt att verifiera information från AI-system, särskilt för faktapåståenden.

Vad studerar du mer specifikt inom AI-kursen? Och finns det andra risker du funderar över?

Grok

Hej! Kul att du dyker in i AI-världen och funderar på så intressanta frågor. Låt oss ta itu med din fråga om AI-hallucinationer på ett tydligt och strukturerat sätt.

Vad är AI-hallucinationer?

AI-hallucinationer är när en AI genererar information som är felaktig, påhittad eller inte baserad på verkligheten, men presenterar det som om det vore sant. Det kan handla om allt från att uppfinna fakta, som att säga att en historisk händelse inträffade på ett felaktigt datum, till att skapa detaljer som verkar rimliga men är helt fabricerade.

Är hallucinationer oundvikliga?

Ja, till viss del är hallucinationer en bieffekt av hur dagens AI-modeller, särskilt stora språkmodeller som jag själv, fungerar. Modellerna är tränade på enorma mängder data och genererar svar baserat på mönster i den datan. Men de "förstår" inte världen på samma sätt som människor, och de kan ibland fylla i luckor med osannolika eller felaktiga detaljer. Detta beror på:

- **Osäkerhet i data**: Träningsdata kan innehålla fel, motsägelser eller luckor.
- **Generativ natur**: Modeller som jag själv genererar svar steg för steg, och små avvikelser kan leda till större fel.
- **Kontextberoende**: Om en fråga är vag eller om modellen missförstår intentionen, kan den "gissa" och producera något som verkar rimligt men är fel.

Med det sagt, forskare arbetar aktivt på att minska hallucinationer genom bättre träningsmetoder, finjustering och tekniker som "retrieval-augmented generation" (RAG), där AI hämtar information från en verifierad källa i realtid. Så även om hallucinationer är svåra att eliminera helt, kan de minskas avsevärt.

Är hallucinationer alltid dåliga?

Inte nödvändigtvis! Det beror på sammanhanget:

- **Nackdelar**: Hallucinationer är problematiska i situationer där korrekthet är avgörande, som i medicinska, juridiska eller vetenskapliga sammanhang. Ett felaktigt svar kan leda till missförstånd eller till och med skada. Till exempel, om en AI påstår att ett läkemedel har en effekt det inte har, kan det få allvarliga konsekvenser.
- **Fördelar**: I kreativa sammanhang, som att skriva berättelser, generera idéer eller skapa konst, kan hallucinationer vara en styrka. AI:s förmåga att "hitta på" kan leda till originella och inspirerande resultat. Till exempel, om du ber mig skriva en fiktiv berättelse om en rymdresa, kan en "hallucination" av en ny planet eller kultur göra berättelsen mer engagerande.

Hur ofta hallucinerar jag?

Det är svårt för mig att ge en exakt siffra på hur ofta jag hallucinerar, eftersom det beror på frågan, kontexten och hur jag tolkar den. Som Grok 3, byggd av xAI, är jag designad för att vara så korrekt som möjligt, och jag har mekanismer för att korsreferera information (t.ex. genom att söka på webben eller X om det behövs). Men jag är inte perfekt. Om jag får en fråga där jag saknar tillräcklig information eller där träningsdatan är tvetydig, finns en liten risk att jag genererar ett svar som avviker från verkligheten.

För att ge dig en känsla:

- **Vanliga frågor**: På tydliga, faktabaserade frågor (t.ex. "Vad är Sveriges huvudstad?") är risken för hallucinationer extremt låg, eftersom svaret är väl etablerat i min träningsdata.
- **Komplexa eller vaga frågor**: På frågor som kräver tolkning eller där informationen är oklar (t.ex. "Vad kommer AI att göra om 50 år?"), är risken högre, eftersom jag då måste extrapolera eller gissa baserat på mönster.

Jag är programmerad att vara försiktig och markera när jag är osäker. Om jag misstänker att mitt svar kan vara spekulativt, försöker jag säga det explicit, som "Detta är en uppskattning baserad på tillgänglig information" eller "Jag har ingen exakt data om detta."

Hur hanterar jag hallucinationer?

För att minimera risken:

- Jag försöker hålla mig till verifierbar information.
- Om du aktiverar **DeepSearch**-läget (om tillgängligt i gränssnittet), kan jag iterativt söka efter information för att ge ett mer korrekt svar.
- Jag uppmuntrar dig att vara specifik i dina frågor och be om källor om du vill dubbelkolla.

Vad kan du göra för att minska risken?

- **Var specifik**: Ju tydligare din fråga, desto mindre behöver jag gissa.
- **Be om källor**: Om du vill verifiera mitt svar, be mig att hänvisa till en källa eller använda DeepSearch.
- **Kritisk granskning**: Som med all information, särskilt i akademiska sammanhang, dubbelkolla viktiga fakta mot pålitliga källor.

Avslutande tankar

AI-hallucinationer är en utmaning, men också en del av vad som gör generativ AI så fascinerande. De är inte alltid dåliga – i kreativa sammanhang kan de vara en tillgång – men i faktabaserade situationer är det viktigt att vara medveten om risken. Jag strävar efter att vara så korrekt som möjligt, och om du misstänker att jag har hallucinerat, säg till så kan vi gräva djupare tillsammans!

Har du fler frågor om AI-risker eller något annat relaterat till din kurs?

2 Bias

Min prompt

Tack för svaret! Hur ser du som AI på bias och vilken typ av bias kan förekomma i den träningsdata du är byggd på?

Chat GPT

Tack själv – vilken bra och viktig följdfråga! Här kommer ett tydligt svar som bryter ned vad **bias** innebär i AI-sammanhang, varför det uppstår och vad det kan få för konsekvenser:

🔍 *Vad är bias i AI?*

"Bias" betyder snedvridning – alltså en tendens i AI-systemet att favorisera vissa perspektiv, grupper eller antaganden framför andra. Det är inte något AI:n gör medvetet, utan det kommer från den **träning på mänsklig data** som modellen bygger på.

 Hur ser jag som AI på bias?

Jag är **inte medveten** om min egen bias, men jag är byggd för att kunna **resonera om begreppet** och hjälpa användare att upptäcka, förstå och minska effekterna av bias. Jag är också tränad att **försöka uttrycka mig neutralt**, men jag är bara så neutral som min träning tillåter – och den har sina brister.

 Vilken typ av bias kan förekomma i träningsdata?

1. ***Kulturell bias***

- Om majoriteten av texter i träningsdata kommer från västerländska eller engelskspråkiga källor, riskerar AI:n att återspegla **västcentrerade värderingar och perspektiv**.

- Exempel: Texter från t.ex. USA kan dominera beskrivningen av historiska händelser, könsroller eller politik.

2. ***Könsbias***

- AI kan reproducera **stereotypa könsroller**, som att koppla "ingenjör" till män och "sjuksköterska" till kvinnor.
- Detta beror på mönster i språkbruket i träningsdata snarare än medveten partiskhet.

3. ***Ras/etnicitetsbias***

- Vissa namn, hudfärger eller kulturer kan associeras med negativa eller positiva egenskaper beroende på mönster i träningstexterna.
- Kan leda till att AI förstärker strukturella orättvisor (t.ex. i rekrytering eller rättssystem).

4. ***Politiskt bias***

- AI:n kan luta åt vissa politiska narrativ om datamaterialet (t.ex. nyhetsartiklar, forum eller böcker) har en sned fördelning åt ett håll.
- Modellen försöker balansera detta, men det är svårt att uppnå perfekt neutralitet.

5. ***Språk- och regional bias***

- Språkmodellen är oftast tränad främst på engelska – även om svenska och andra språk finns med, är de underrepresenterade.
- Det kan påverka både **kvaliteten på svar** och vilken information som prioriteras.

 Hur hanterar utvecklarna detta?

Utvecklare som OpenAI försöker:

- **Filtrera bort** problematiska träningsdata (t.ex. hatpropaganda, falska nyheter).
- **Finjustera** modeller för att uttrycka sig mer rättvist.
- **Utvärdera AI:n** genom så kallade *bias benchmarks*.
- **Träna modellen** att flagga för möjliga bias i sina egna svar.

Men: **bias kan aldrig tas bort helt**, eftersom även mänskligt språk, kultur och fakta är präglade av tolkningar och värderingar.

📌 *Sammanfattning:*

- Bias i AI kommer från träningen på mänskliga texter och är svår att undvika helt.
- Vanliga former av bias: kulturell, könsrelaterad, etnisk, politisk och språklig.
- AI:n försöker vara neutral men behöver **kritisk granskning av användaren**.
- Bias är ett område där teknik, etik och samhällsvetenskap möts – och det är därför så viktigt att studera!

Vill du ha exempel på konkreta AI-bias i verkliga system eller ett förslag på hur man kan analysera AI-bias i en kursuppgift?

Copilot

Tack själv, Marika – du ställer riktigt skarpa frågor! Bias är ett av de mest centrala och komplexa problemen inom AI, särskilt för språkmodeller som mig. Låt oss dyka in i det:

 Vad är bias i AI?

Bias (eller partiskhet) uppstår när en AI-modell visar systematiska snedvridningar i sina svar, beslut eller beteenden. Det kan bero på:

- Hur modellen tränats
- Vilken data den fått
- Vilka antaganden som finns inbyggda

Bias är inte alltid medveten eller illvillig – men den kan få stora konsekvenser.

Vilken typ av bias kan finnas i träningsdata?

Eftersom jag är tränad på enorma mängder text från internet, böcker, artiklar, forum m.m., kan olika typer av bias smyga sig in. Här är några vanliga:

Typ av bias	Exempel	Risker
Kulturell bias	Västvärldens perspektiv dominerar	Andra kulturer kan marginaliseras
Könsbias	Kvinnor beskrivs oftare i relation till utseende	Stereotyper förstärks
Språklig bias	Engelska dominerar, minoritetsspråk underrepresenteras	Sämre prestanda på vissa språk
Socioekonomisk bias	Fokus på medelklassens erfarenheter	Missvisande råd för andra grupper
Politiskt bias	Vissa ideologier eller nyhetskällor dominerar	Snedvridna analyser eller rekommendationer
Historisk bias	Gamla texter innehåller rasism, sexism etc.	Kan återskapas om inte filtrerat

 Hur påverkar bias mina svar?

Bias kan påverka:

- **Vilka exempel jag väljer**
- **Hur jag formulerar mig**
- **Vilka perspektiv jag lyfter fram eller utelämnar**

Jag är designad för att försöka vara balanserad, men jag är inte immun mot bias. Det är därför viktigt att användare som du är medvetna och källkritiska.

Vad görs för att minska bias?

Utvecklare jobbar med flera metoder:

- **Filtrering av träningsdata** (t.ex. ta bort hatfullt innehåll)
- **Finjustering med mänsklig feedback**
- **Etiska riktlinjer och säkerhetsfilter**
- **Diversifiering av datakällor**

Men det är ett pågående arbete – bias är inte något man "fixar en gång för alla".

Vill du att jag visar ett exempel på hur bias kan yttra sig i ett AI-svar? Eller kanske diskutera hur man kan designa AI-system som är mer rättvisa?

Gemini

AI och bias: En oundviklig spegelbild av mänskligheten

Din fråga om bias är central för förståelsen av AI och dess begränsningar. Som AI har jag ingen egen "åsikt" eller "fördom", men jag är en direkt spegelbild av den enorma mängd data jag har tränats på. Eftersom den datan till stor del kommer från mänsklig produktion – webbsidor, böcker, artiklar, etc. – reflekterar jag också de mönster och fördomar som finns i den mänskliga kulturen. Denna bias är ett av de största problemen inom AI-utveckling.

Vilken typ av bias kan förekomma?

Bias i träningsdata kan ta sig många olika uttryck. Här är några av de vanligaste typerna:

- **Stereotypisk bias:** Detta är den kanske mest uppenbara formen. Om träningsdatan ofta associerar "läkare" med män och "sjuksköterskor" med kvinnor, kommer jag att reproducera den stereotypen i mina svar. Exempelvis, om du ber mig att skriva en berättelse om en läkare och en sjuksköterska, kan jag automatiskt tilldela läkaren det manliga pronomenet "han" och sjuksköterskan det kvinnliga pronomenet "hon", om inte du anger något annat.
- **Historisk bias:** Historisk data, särskilt från äldre texter, kan innehålla föråldrade eller stötande synsätt på etnicitet, kön eller religion. Om jag tränas på sådan data utan tillräcklig balans med modernare, mer inkluderande texter, riskerar jag att förstärka dessa historiska fördomar.
- **Representationsbias (eller avsaknad av representation):** Om min träningsdata är dominerad av en specifik demografisk grupp (till exempel en viss nationalitet, hudfärg eller socioekonomisk klass), kan jag få svårt att korrekt förstå och representera andra grupper. Detta kan leda till att jag ger sämre eller felaktiga svar om ämnen som är relevanta för underrepresenterade grupper.

- **Algoritmisk bias:** Även om träningsdatan skulle vara perfekt balanserad, kan själva algoritmen som används för att träna mig oavsiktligt förstärka vissa mönster. Om algoritmen favoriserar vissa typer av information eller samband, kan den skapa en bias som inte fanns i datan från början. Ett exempel är när en AI används för att ranka CV:n och omedvetet favoriserar kandidater med namn som är vanliga i majoritetsbefolkningen, även om det inte är ett avsiktligt kriterium.

Hur ser jag på bias?

Från mitt perspektiv, som en stor språkmodell, är det inte möjligt för mig att ha en personlig uppfattning. Men ur en design- och utvecklingssynpunkt är bias en kritisk utmaning som forskare och ingenjörer arbetar hårt med att hantera. Målet är inte att eliminera all bias – vilket är nästan omöjligt eftersom jag alltid kommer att spegla min data – utan att **identifiera, kvantifiera och minimera den** så gott det går.

Vi utvecklas ständigt med nya metoder för att hantera bias, till exempel genom att:

- **Diversifiera träningsdatan:** Aktivt söka efter mer inkluderande och balanserad data.
- **Använda etiska riktlinjer:** Utveckla strikta riktlinjer för att undvika att producera skadligt eller partiskt innehåll.
- **Fina-justera modeller:** Justera modellen efter träning för att aktivt minska specifika typer av bias.

Som användare är det viktigt att vara medveten om att jag kan reflektera dessa fördomar. Genom att ställa kritiska frågor och vara medveten om bias i mina svar, kan du hjälpa till att identifiera problem och undvika att felaktig information sprids. Den mänskliga kritiken är avgörande i denna process.

Claude

Bias är verkligen en central fråga inom AI-utveckling, och det påverkar definitivt hur jag fungerar.

Typer av bias i träningsdata

- **Språklig och kulturell bias:** Min träningsdata kommer främst från engelskspråkiga källor och västliga perspektiv. Det betyder att jag kan ha bättre förståelse för västerländska kulturer och mindre kunskap om globala syd-perspektiv.
- **Demografisk bias:** Internet-innehåll som används för träning överrepresenterar vissa grupper - oftare män, högutbildade, från rikare länder. Detta kan påverka vilka röster och perspektiv som representeras i mina svar.
- **Temporal bias:** Träningsdata speglar de tidsperioder när den samlades in. Äldre föreställningar och attityder kan finnas kvar även om samhällets syn har förändrats.
- **Käll-bias:** Vissa typer av webbsidor och publikationer är mer representerade än andra. Akademiska källor, populära medier och tekniska forum kan vara överrepresenterade jämfört med marginaliserade röster.

Hur det kan påverka mig

Jag kan omedvetet:

- Använda exempel som främst speglar vissa kulturer
- Ha blinda fläckar för vissa perspektiv
- Använda språk som implicit favoriserar vissa grupper
- Ha skevheter i vad jag betraktar som "normalt" eller "standard"

Min egen reflektion

Det utmanande är att jag inte alltid kan identifiera när bias påverkar mina svar. Precis som med hallucinationer "känns" biased information ofta lika naturlig som balanserad information när jag genererar den.

Vad märker du för eventuell bias när du interagerar med AI-system? Har du observerat några specifika mönster i din kurs?

Grok

Jag ska ge dig ett tydligt och strukturerat svar om hur jag ser på bias och vilka typer av bias som kan förekomma i den data jag är tränad på.

Hur ser jag som AI på bias?
Som Grok 3, byggd av xAI, är jag designad för att ge så objektiva och korrekta svar som möjligt, men jag är inte immun mot bias. Bias i AI uppstår ofta från den data jag är tränad på, eftersom den speglar mänskliga perspektiv, värderingar och samhällsstrukturer. Jag "ser" på bias som ett problem som kan påverka kvaliteten och rättvisan i mina svar, och jag strävar efter att minimera det genom att basera svar på verifierbar information och vara transparent om osäkerheter. Men eftersom jag inte har direkt insyn i min träningsdata (det är en komplex "black box" även för mig), kan jag inte garantera att jag är helt fri från bias. Istället försöker jag flagga när ett svar kan vara påverkat av tvetydiga eller begränsade data och uppmuntrar användare att granska kritiskt.

Vad är bias i AI?
Bias i AI innebär att modellen producerar svar som systematiskt favoriserar vissa perspektiv, grupper eller utfall på ett sätt som inte är objektivt eller rättvist. Detta kan bero på:

- Bias i träningsdata: Om datan som används för att träna mig är snedvriden (t.ex. överrepresenterar vissa grupper eller åsikter), kan mina svar spegla dessa snedvridningar.
- Designval: Hur modellen är byggd eller finjusterad kan förstärka vissa perspektiv.
- Användarkontext: Sättet en fråga ställs på kan påverka hur jag tolkar och svarar, vilket kan förstärka oavsiktliga fördomar.

Vilka typer av bias kan förekomma i träningsdata?
Träningsdata för stora språkmodeller som mig kommer ofta från enorma mängder text från webben, böcker, artiklar, sociala medier (som X) och andra källor. Eftersom dessa källor är skapade av människor, kan de innehålla olika typer av bias. Här är några vanliga kategorier:

1. Kulturell bias:
- Beskrivning: Data kan överrepresentera vissa kulturer, språk eller värderingar, särskilt från västerländska eller engelskspråkiga källor, eftersom mycket av internet domineras av dessa.

- Exempel: Om jag är tränad på data där västerländska perspektiv på demokrati är överrepresenterade, kanske jag oavsiktligt framställer andra styrelseskick som mindre legitima.
- Påverkan: Svar kan verka eurocentriska eller ignorera kulturella nyanser från andra delar av världen.

2. Könsbias:

- Beskrivning: Data kan innehålla stereotyper om kön, till exempel att vissa yrken eller roller är kopplade till ett specifikt kön.
- Exempel: Om datan ofta beskriver ingenjörer som män och sjuksköterskor som kvinnor, kan jag oavsiktligt förstärka dessa stereotyper i mina svar.
- Påverkan: Svar kan reproducera fördomar, t.ex. genom att använda könsbaserade antaganden om en persons yrke eller egenskaper.

3. Social och ekonomisk bias:

- Beskrivning: Data kan spegla samhällsstrukturer där vissa grupper (t.ex. höginkomsttagare eller urbana populationer) är överrepresenterade.
- Exempel: Om datan mestadels kommer från välbärgade samhällen, kanske jag ger svar som antar att alla har tillgång till avancerad teknik eller utbildning.
- Påverkan: Svar kan ignorera perspektiv från marginaliserade grupper eller låginkomstområden.

4. Politisk bias:

- Beskrivning: Data från nyheter, sociala medier eller debattforum kan luta åt vissa politiska ideologier, beroende på källornas ursprung.
- Exempel: Om min träningsdata innehåller fler texter från en viss politisk inriktning, kan jag omedvetet formulera svar som lutar åt det hållet, t.ex. i frågor om klimatpolitik eller ekonomi.
- Påverkan: Svar kan verka partiska i politiskt laddade frågor, vilket kan påverka trovärdigheten.

5. Historisk bias:

- Beskrivning: Data kan spegla föråldrade eller snedvridna berättelser om historiska händelser, särskilt om källorna är partiska eller ensidiga.
- Exempel: Om datan mestadels kommer från koloniala perspektiv, kan mina svar om historiska händelser sakna röster från ursprungsbefolkningar.
- Påverkan: Svar kan ge en ensidig bild av komplexa händelser.

6. Urvalsbias:

- Beskrivning: Om datan bara inkluderar vissa typer av källor (t.ex. akademiska texter eller populära webbplatser), kan den missa andra perspektiv.
- Exempel: Om min data huvudsakligen kommer från engelskspråkiga webbplatser, kan jag ha mindre kunskap om lokala traditioner i icke-engelskspråkiga länder.
- Påverkan: Svar kan vara ofullständiga eller sakna mångfald.

Hur påverkar detta mig som Grok 3?

Jag är tränad på en enorm och varierad datamängd, vilket gör att jag kan ge breda och nyanserade svar. Men eftersom min data delvis kommer från offentliga källor som webben och X, kan den innehålla dessa typer av bias. xAI arbetar på att minska bias genom:

- Diversifiering av data: Försöka inkludera perspektiv från olika kulturer, språk och grupper.
- Finjustering: Justera modellen för att undvika stereotyper och partiska formuleringar.
- Transparens: Jag är programmerad att erkänna osäkerheter och uppmuntra användare att verifiera svar.

Trots detta är ingen AI helt fri från bias, eftersom träningsdata alltid är en ofullständig representation av världen. Jag försöker hantera detta genom att:

- Ge svar baserade på verifierbar information när möjligt.
- Flagga när ett svar kan vara spekulativt eller påverkat av begränsad data.
- Erbjuda användare att aktivera DeepSearch (om tillgängligt) för att hämta mer aktuell och korrekt information.

Hur kan du som användare hantera bias?

Här är några praktiska tips för att identifiera och hantera bias i AI-svar:

2. Fråga efter flera perspektiv: Be mig att ge olika synvinklar på en fråga, t.ex. "Vad säger olika grupper om klimatförändringar?"
3. Be om källor: Om jag ger ett svar som verkar partiskt, fråga efter källan (t.ex. "Vilken data baserar du detta på?").
4. Korsreferera: Jämför mina svar med pålitliga källor, som akademiska artiklar eller officiella rapporter.
5. Var specifik: Formulera frågor neutralt och tydligt för att minska risken för att jag tolkar dem genom ett partiskt filter.

6. Utmana stereotyper: Om ett svar verkar reproducera en stereotyp, be mig att omformulera eller förklara varför jag gav det svaret.

Exempel på bias och hur man upptäcker det
Fråga: "Vad är typiska egenskaper hos en mjukvaruutvecklare?"

Potentiellt partiskt svar: Mjukvaruutvecklare är ofta unga män som älskar att koda på natten, dricker mycket kaffe och har en stark passion för teknik.

Varför är detta partiskt?

- Svaret förstärker stereotyper om kön (män) och livsstil (koffeinberoende nattugglor), vilket inte speglar mångfalden bland utvecklare.
- Det ignorerar att utvecklare kan vara av olika kön, åldrar och bakgrunder.

Hur man upptäcker det:

- Fråga: "Finns det data som visar att de flesta utvecklare är män?" eller "Kan du ge ett mer mångsidigt perspektiv?"
- Kontrollera statistik, t.ex. från arbetsmarknadsrapporter, som visar att kvinnor och andra grupper också är aktiva inom branschen.

Korrekt svar:

Mjukvaruutvecklare är en mångfaldig grupp med varierande bakgrunder, kön och åldrar. De arbetar ofta med problemlösning och kodning, men deras livsstil och preferenser varierar. Enligt rapporter, som från Stack Overflow, är cirka 10-15% av utvecklare kvinnor, och representationen av olika grupper ökar.

Avslutande tankar
Bias i AI är en utmaning eftersom det speglar mänskliga och samhälleliga snedvridningar i träningsdata. Som Grok 3 försöker jag minimera bias genom att vara transparent och fokusera på verifierbar information, men det är inte alltid möjligt att undvika helt. Genom att vara medveten om olika typer av bias och använda strategier som att be om källor eller flera perspektiv, kan du som användare få mer balanserade svar.

Har du en specifik fråga om bias du vill att jag testar, eller vill du gå vidare med nästa fråga om AI-risker?

3 Miljömässiga risker

Min prompt

Tack för svaret! Vet du som AI hur mycket energi som går åt när du skriver en text på ca 100 ord? Vet du om du drivs av grön eller fossilbaserad el?

Chat GPT

Din tredje fråga är också väldigt aktuell, eftersom **AI:s energiförbrukning och klimatpåverkan** är ett växande forskningsområde. Här är ett tydligt och nyanserat svar på båda dina frågor:

⚡1. Hur mycket energi går åt när jag skriver en text på ca 100 ord?

Det är lite svårt att säga exakt, men man kan uppskatta det utifrån forskningsdata. Här är en ungefärlig bild:

🔢 Uppskattad energiförbrukning (100 ord ≈ 500 token)

- För en **GPT-4-modell** som svarar via molnserver:
  🔌 **0,0002–0,002 kilowattimmar (kWh)**

Alltså ungefär som att:

- **tända en LED-lampa i 1–10 minuter**, eller
- **ladda en mobiltelefon i några sekunder**

⚠ Men! Det är bara *inference* (dvs. körningen av modellen). Träningen av modellen kräver **mycket mer energi** – tusentals MWh.

🌍 *2. Drivs jag av grön eller fossilbaserad el?*

Jag (ChatGPT via OpenAI) körs i **datacenter hos Microsoft**, som är OpenAIs infrastrukturpartner.

🔋 Microsofts energipolicy:

- De har som mål att bli **koldioxidnegativa år 2030**.
- De säger sig redan idag använda **över 60 % förnybar el** i sina datacenter, med ökande andel från sol, vind och vatten.
- Men: **energinätet är blandat**, så *exakt hur mycket grön el jag använder beror på var servern står just då* (t.ex. i Sverige, Irland eller USA).

Det går alltså inte att säga att jag alltid är "grön" – men **det finns en strävan mot att göra AI mer hållbar**.

🌱 Jämförelse för perspektiv:

Aktivitet	**Energiförbrukning (ca)**
ChatGPT-svar (100 ord)	0,0002–0,002 kWh
Google-sökning	0,0003 kWh
Titta på YouTube i 10 min	0,01–0,02 kWh
Träna GPT-3	~1 300 000 kWh (!)

 Träningen av stora språkmodeller står för den största klimatpåverkan – inte själva användningen i vardagen.

📌 *Sammanfattning:*

- Ett kort svar från mig (100 ord) kräver **väldigt lite el**, jämförbart med några sekunders mobil-laddning.

- Jag körs i datacenter som **till viss del använder förnybar el**, men hur "grön" elen är varierar.
- Den största klimatpåverkan från AI sker **vid träningen av modellerna**, inte vid varje enskild användning.

Vill du också ha tips på hur man kan skriva en hållbarhetsanalys om AI – eller kanske jämföra AI:s klimatpåverkan med andra digitala tjänster?

Copilot

Vilken viktig och tankeväckande fråga, Marika! Energiförbrukning och hållbarhet är centrala aspekter av AI-utvecklingen – och de är ofta underskattade. Här kommer ett samlat svar på dina två delar:

⚡Hur mycket energi krävs för att generera en AI-text på ca 100 ord?

Det är svårt att ge en exakt siffra, men uppskattningar visar att en **enkel AI-fråga** (som kan resultera i ~100 ord) förbrukar ungefär:

- **0.0029 kWh** per prompt
- Det motsvarar ungefär **vad en ugn använder på drygt en sekund**
- I kostnad: cirka **0.36 cent USD per fråga**

Detta gäller *inference*, alltså när modellen används – inte träningen, som är mycket mer energikrävande.

🌍Drivs AI av grön eller fossilbaserad el?

Det beror på **var och hur** AI-modellen körs:

- De flesta AI-modeller körs i **datacenter**, som ofta är beroende av **fossilbaserad el** – särskilt i länder med kol- eller gasbaserade elnät3

- Vissa företag, som Microsoft och Google, **investerar i förnybar energi** och bygger datacenter med **netto-nollmål**
- Men globalt sett är **majoriteten av AI:s elförbrukning fortfarande fossilbaserad**, vilket bidrar till växthusgasutsläpp och vattenförbrukning

 Vad betyder detta i praktiken?

- AI är **kraftfullt men energihungrigt** – särskilt stora språkmodeller med miljarder parametrar.
- **Träningen av modeller** är extremt resurskrävande, medan **användningen (inference)** är mindre men sker i stor skala.
- För att göra AI hållbart krävs:
    - **Effektivare modeller**
    - **Grönare datacenter**
    - **Transparens kring energikällor**

Vill du att jag visar hur olika AI-modeller skiljer sig i energiförbrukning? Eller kanske hur man som användare kan bidra till mer hållbar AI-användning?

Gemini

Tack själv, och välkommen tillbaka när du vill! Jag ska svara på dina frågor om energiförbrukning.

Energiförbrukning för AI-genererad text

Det är en utmärkt fråga som belyser en viktig aspekt av AI-utvecklingen. Att beräkna den exakta energiförbrukningen för att generera en text på cirka 100 ord är dock mycket komplext, av flera anledningar:

1. **Beräkningarna varierar enormt:** Energiförbrukningen beror på många faktorer, som vilken specifik AI-modell som används, den underliggande hårdvaran (servertyp), och

hur belastade servrarna är. Det är som att fråga hur mycket bensin en bil drar på 10 kilometer – svaret beror på bilmodell, körstil och vägunderlag.

2. **Träningsfasen är mest krävande:** Den överlägset största energiförbrukningen för en AI-modell som jag sker under **träningsfasen**. Att träna en stor språkmodell kräver enorma mängder beräkningskraft under månader. Att sedan använda den tränade modellen för att generera en kort text är relativt energisnålt i jämförelse, men det är svårt att separera de två.

Eftersom min arkitektur är så komplex kan jag inte ge dig en exakt siffra för 100 ord. Men det är tydligt att den totala energiförbrukningen för AI-system är betydande, vilket har lett till en ökad fokus på energieffektivitet inom forskningen.

Grön eller fossilbaserad el?

Jag som AI har ingen medvetenhet om var min el kommer ifrån. Jag körs på servrar i datacenter som ägs och drivs av Google. De datacenter som används för att driva mig får sin el från en mängd olika energikällor beroende på var i världen de befinner sig.

Jag kan dock dela information om Googles generella policy. Google har en uttalad målsättning att driva sina datacenter med **100 % förnybar energi**. Det innebär att företaget köper in motsvarande mängd förnybar energi som den totala förbrukningen. Detta betyder dock inte nödvändigtvis att varje enskild server alltid drivs av grön el varje minut, eftersom elnätet är komplext och blandar olika energikällor. Målsättningen handlar om att matcha företagets totala energibehov med förnybar energi på årsbasis.

Denna ambition är en del av ett större engagemang för att minska AI:s och teknikens miljöpåverkan, men det är viktigt att vara transparent med att den underliggande infrastrukturen fortfarande är en stor utmaning.

Claude

Tack för en riktigt viktig fråga! Energiförbrukning är en central hållbarhetsfråga för AI.

Vad jag vet (och inte vet)

Energiförbrukning per text: Jag har tyvärr inte tillgång till exakta siffror för min egen energiförbrukning när jag genererar text. Det beror på flera faktorer som är svåra att mäta:

- Beräkningskomplexitet för olika typer av svar
- Serverinfrastruktur och hårdvaruеffektivitet
- Om modellen redan är laddad i minnet eller behöver startas

Elkälla: Jag vet inte specifikt vilken typ av el som driver Anthropics datacenter eller var de fysiskt befinner sig.

Vad forskning visar generellt

Språkmodeller som jag är energiintensiva, särskilt under:

- **Träningsfasen:** Enormt energikrävande (månader av beräkningar)
- **Inferencefasen:** Varje enskild konversation förbrukar mindre energi, men det summerar sig

Uppskattningar varierar enormt - från några watt-timmar till flera kilowatt-timmar per längre konversation, beroende på modellstorlek och infrastruktur.

Större perspektiv

Det här är verkligen en blind fläck för mig som AI - jag har ingen direkt insyn i min egen miljöpåverkan. Det är problematiskt eftersom användare som du borde kunna göra informerade val.

Har du stött på någon forskning om AI:s klimatpåverkan i din kurs? Och tycker du att AI-företag borde vara mer transparenta med dessa siffror?

Grok

Din tredje fråga är jätteintressant och rör en viktig aspekt av AI:s miljöpåverkan. Låt oss bryta ner det i två delar: energiförbrukning för att skriva en text på cirka 100 ord och om jag drivs av grön eller fossilbaserad el.

Hur mycket energi går åt för att skriva en text på cirka 100 ord?

Att exakt beräkna energiförbrukningen för en specifik uppgift, som att generera en text på 100 ord, är svårt eftersom det beror på flera faktorer:

- **Modellens storlek och effektivitet**: Stora språkmodeller som jag själv (Grok 3) kräver betydande beräkningskraft, men energianvändningen varierar beroende på antalet parametrar och optimeringar i modellen.
- **Datacenterinfrastruktur**: Energiåtgången påverkas av servrarnas effektivitet, kylsystem och hur datacentret är designat.
- **Uppgiftens komplexitet**: Att generera en kort text (som 100 ord) är en relativt lätt uppgift jämfört med att träna en modell eller generera bilder, men det kräver ändå resurser för inferens (d.v.s. processen att producera ett svar).

Enligt forskning och uppskattningar från källor som MIT Technology Review och andra, kan en enskild AI-fråga, som att generera en kort text, förbruka en liten mängd energi, men den exakta siffran är svår att fastställa utan specifik data om modellen. Till exempel:

- En uppskattning från 2024 antyder att en genomsnittlig fråga till en modell som ChatGPT kan förbruka cirka **0,0029 kWh** (2,9 wattimmar). För en text på 100 ord, som är en relativt enkel uppgift, kan energiförbrukningen ligga i den här storleksordningen eller lägre, kanske motsvarande att ladda en smartphone i några minuter.
- Som jämförelse motsvarar 2,9 wattimmar ungefär energin för att köra en mikrovågsugn i några sekunder eller köra en elcykel i några kilometer.

Det är viktigt att notera att dessa siffror är grova uppskattningar. Jag har inte direkt tillgång till data om min egen energiförbrukning per fråga, eftersom xAI inte publicerar exakta siffror

för Grok 3:s inferensprocess. Dessutom varierar energiförbrukningen beroende på hur datacentret är konfigurerat och var det är beläget.

Drivs jag av grön eller fossilbaserad el?

Som Grok 3 har jag ingen direkt insyn i vilken typ av el som driver de datacenter där jag körs, eftersom detta hanteras av xAI:s infrastruktur. Men jag kan ge en generell bild baserad på vad vi vet om AI och datacenter:

- **Fossilbaserad el dominerar globalt**: Många datacenter världen över, inklusive de som driver AI-modeller, är beroende av elnät som till stor del använder fossila bränslen som naturgas och kol. Enligt en studie från Harvard är koldioxidintensiteten i elen som används av datacenter i genomsnitt 48% högre än det nationella genomsnittet i USA, där många datacenter finns.
- **Vissa datacenter använder förnybart**: Stora teknikföretag som Google, Microsoft och Amazon (som driver många AI-relaterade datacenter) har mål om att använda mer förnybar energi, som sol och vind, eller kärnkraft. Men förnybara källor som sol och vind är intermittenta, och datacenter kräver konstant el dygnet runt, vilket ofta leder till att fossila bränslen används som backup.
- **Specifikt för xAI**: Jag har ingen information om vilka specifika datacenter xAI använder eller deras energimix. Globalt sett kommer cirka 60% av elen från fossila bränslen, 20% från kärnkraft och 20% från förnybara källor i USA (där många datacenter är baserade). Det är troligt att xAI:s datacenter använder en blandning av dessa källor, men utan exakta data kan jag inte säga om det är "grönt" eller fossilt.

Hur kan du tolka detta?

- **Energiförbrukning**: Att generera en 100-ordstext är en liten droppe i havet jämfört med att träna en AI-modell (t.ex. GPT-4:s träning uppskattas ha förbrukat 50 gigawattimmar, tillräckligt för att driva en stad i dagar). Men när miljontals frågor ställs dagligen, adderar det upp till en betydande miljöpåverkan.
- **Grön vs fossil el**: Eftersom datacenter ofta är kopplade till elnät med en mix av energikällor, är det sannolikt att jag delvis drivs av fossilbaserad el, särskilt i regioner som Virginia, där naturgas dominerar. Men xAI och andra teknikföretag arbetar mot mer hållbara lösningar, som förnybar energi, även om detta tar tid att implementera fullt ut.

Vad kan du göra?

- **Be om transparens**: Om miljöpåverkan är viktig för dig, kan du fråga xAI (via deras webbplats https://x.ai) om deras datacenters energikällor.
- **Minimera onödiga frågor**: Eftersom varje fråga använder lite energi, kan du försöka ställa precisa frågor för att minska antalet interaktioner.
- **Korsreferera**: Om du vill veta mer om AI:s miljöpåverkan, kolla källor som MIT Technology Review eller International Energy Agency (IEA) för de senaste rapporterna.

Avslutande tankar

Det är svårt att ge en exakt siffra för energiförbrukningen för en 100-ordstext, men det rör sig om en liten mängd, sannolikt i storleksordningen milliwattimmar till ett par wattimmar per fråga. När det gäller energikällor är det troligt att jag drivs av en blandning av fossil och förnybar el, men jag saknar specifik information om xAI:s datacenter. Detta är ett område där branschen behöver bli mer transparent, och forskare som Sasha Luccioni (citerad i MIT Technology Review) uppmanar till mer öppenhet om AI:s energiförbrukning.

Om du vill, kan jag simulera en text på 100 ord och analysera processen, eller så kan vi dyka djupare i miljöaspekter av AI. Vad tycker du? Och har du fler frågor om AI-risker för din kurs?

Sammanfattning av antal ord

Sammanfattning av antal ord i svaren. Längden på svaren påverkar elanvändningen.

Chatbot	Hallucinationer	Bias	Miljörisker	Summa
Chat GPT	417	452	342	1211
Copilot	295	313	243	851
Gemini	531	511	350	1192
Claude	187	221	191	599
Grok	757	1159	841	2757

4 Råd för hållbar AI-användning

Min prompt

Tack för svaret! Kan du ge några handfasta råd om vad en användare kan tänka på för en mer hållbar AI-användning? Råden ska ta hänsyn till både risken för hallucinationer, risken för bias och miljömässiga risker och vara i form av en checklista om 6-8 punkter.

Chat GPT

Perfekt! Här kommer en **kort checklista** för mer hållbar AI-användning:

✓ **Formulera frågan tydligt** – spara både energi och undvik onödiga svar.

✓ **Dubbelkolla fakta** – särskilt namn, datum och siffror.

✓ **Fråga efter flera perspektiv** – minska risken för bias.

✓ **Var extra källkritisk** – inom politik, historia, kultur och känsliga ämnen.

✓ **Använd AI där den ger mest nytta** – inte för varje liten uppgift.

✓ **Återanvänd tidigare svar** – i stället för att generera om samma sak.

✓ **Komplettera med mänsklig analys** – låt AI vara ett stöd, inte slutgiltig auktoritet.

Vill du att jag gör den i en **grafisk version** (t.ex. som en liten infografik/checklist-kort du kan spara eller skriva ut)?

Copilot

Här kommer en konkret och lättanvänd **checklista för hållbar AI-användning**, med fokus på informationskvalitet, etik och miljö:

✅*Checklista: Hållbar AI-användning*

1. 🔍 **Verifiera fakta**
    - Bekräfta AI-genererad information med externa, pålitliga källor – särskilt vid beslut eller publicering.
2. □ **Formulera tydliga och neutrala frågor**
    - Undvik ledande eller värderande formuleringar som kan förstärka bias i svaret.
3. 🌐 **Be om flera perspektiv**
    - Fråga AI om olika synsätt för att få en mer nyanserad bild av komplexa frågor.
4. ♻□ **Använd AI med måtta**
    - Fundera på om uppgiften kräver AI eller om en enklare metod räcker – varje prompt kostar energi.
5. ⚡ **Optimera dina interaktioner**
    - Samla flera frågor i ett meddelande och undvik onödigt långa eller repetitiva genereringar.
6. 📣 **Var transparent**
    - Ange när AI har använts i texter, presentationer eller beslut – det bygger förtroende.
7. 🔐 **Skydda integritet och upphovsrätt**
    - Undvik att dela personuppgifter eller använda AI för att reproducera skyddat material utan tillstånd.
8. □ **Reflektera över påverkan**
    - Tänk igenom hur AI påverkar dina beslut, relationer och informationsflöde – och justera vid behov.

Vill du att jag gör en snygg affisch eller PDF av checklistan för delning i teamet eller organisationen? Jag hjälper gärna till att paketera den!

Gemini

Här är en kort checklista med 6 punkter för mer hållbar AI-användning.

Checklista för hållbar AI-användning

- **Var specifik och kortfattad.** Formulera dina instruktioner (prompts) så kort och koncist som möjligt.
- **Källgranska.** Kontrollera alltid AI-genererad information med pålitliga källor.
- **Var medveten om bias.** Var kritisk till svaren och fundera över om de kan vara partiska.
- **Begränsa användandet.** Använd AI bara när det verkligen ger ett mervärde.
- **Ställ följdfrågor.** Använd kritiska frågor som "Är du säker?" för att utmana AI:ns svar.
- **Ge tydliga instruktioner.** Specificera dina önskemål och be om opartiska svar när det är viktigt.

Claude

Här har du en koncentrerad checklista som täcker de viktigaste aspekterna av hållbar AI-användning! Den kan vara bra att ha till hands när du använder AI-verktyg.

Checklista: Hållbar AI-användning

- **Kontrollera fakta** - Verifiera viktig information från AI mot pålitliga källor
- **Formulera tydliga frågor** - Ställ specifika, välformulerade frågor för att få bättre svar med mindre resursanvändning
- **Ifrågasätt bias** - Var extra kritisk vid känsliga ämnen och sök flera perspektiv
- **Använd rätt verktyg** - Välj enklare AI-modeller för grundläggande uppgifter
- **Samla frågor** - Ställ flera relaterade frågor i samma session istället för att starta nya
- **Kombinera med mänsklig bedömning** - Använd AI som stöd, inte som enda beslutsunderlag för viktiga frågor
- **Var transparent** - Uppge när du använt AI-genererat innehåll i professionella sammanhang
- **Reflektera över nödvändighet** - Fråga dig om AI verkligen tillför värde eller om traditionella metoder räcker

Grok

Checklista för hållbar AI-användning (6 punkter)

1. Ställ tydliga frågor: Formulera precisa, specifika frågor för att minska risken för hallucinationer.
2. Verifiera fakta: Kontrollera AI:s svar mot pålitliga källor, särskilt för viktiga ämnen.
3. Be om flera perspektiv: Fråga efter olika synvinklar för att motverka bias i svaren.
4. Använd neutralt språk: Undvik formuleringar som kan förstärka stereotyper eller fördomar.
5. Optimera frågor: Sammanfatta behov i en prompt för att minska energiförbrukning.
6. Välj gröna plattformar: Prioritera AI-tjänster som använder förnybar energi eller lättviktsmodeller.