AI KAN LYVE


AI KAN LYVE
AI HALLUCINERER
AI VIL PLEASE DIG
AI
HELLERE OPFINDE FAKTA END

AI kan lyve – lær at kontrollere den med 20 prompting-teknikker

Kunstig intelligens kan opfinde fakta, finde på retssager og give livstruende råd. Men med den rette teknik kan du reducere fejlene med op til 90 procent. Her er alt, hvad du behøver at vide om prompt engineering.

Når AI skaber kaos: De mest alvorlige hallucinationer

AI-hallucinationer er ikke længere en kuriositet for tech-entusiaster. De har kostet virksomheder milliarder, ført til juridiske sanktioner og endda sat menneskeliv på spil. Hallucinationer opstår, når AI-systemer opfinder information, der præsenteres som fakta – og konsekvenserne kan være katastrofale.

Juridiske katastrofer

I 2023 blev to amerikanske advokater sanktioneret, efter deres ChatGPT-genererede dokumenter indeholdt seks fuldstændig opdigtede retssager, komplet med falske dommernumre og citater. Problemet spredte sig: I 2025 rapporterede juridisk ekspert Damien Charlotin, at mindst 156 sager globalt har involveret hallucinerede juridiske citater. Advokater er blevet idømt bøder på op til 10.000 dollar, og flere har modtaget såkaldte "bar referrals" – alvorlige klager, der kan koste dem deres licens.

Air Canada oplevede i 2024 konsekvenserne af en hallucinerende chatbot. En kunde blev lovet en "bereavement fare" – en rabat ved dødsfald i familien – som aldrig havde eksisteret. Da selskabet nægtede at honorere løftet, gik sagen i retten. Den canadiske domstol fastslog, at virksomheden var juridisk ansvarlig for chatbottens falske oplysninger og skulle betale kompensation.

Økonomisk kaos

Da Google i 2023 præsenterede deres AI-chatbot Bard i en livestreamet demonstration, begik systemet en fejl, der kostede dyrt. Barden hævdede, at James Webb-rumteleskopet havde taget det første billede nogensinde af en exoplanet – en påstand, der var grundlæggende forkert. Aktiekursen styrtdykkede øjeblikkeligt med otte til ni procent. På få timer havde Alphabets markedsværdi tabt cirka 100 milliarder dollar.

I erhvervslivet har AI opfundet startups, kontrakter og leverandører, der ikke eksisterer. Virksomheder har modtaget AI-genererede kontrakter med ikke-eksisterende firmaer – en genvej til svindel og økonomiske tab. Finansielle analyser genereret af AI har skabt falske aktiekurser og prognoser, hvilket har ført investorer på vildspor.

Livsfarlige råd

Nogle af de mest skræmmende hallucinationer involverer sundhed og sikkerhed. I 2024 foreslog en Google AI-chatbot, at man kunne rense en vaskemaskine ved at blande blegemiddel og eddike – en kombination, der skaber giftig klorgas, som kan være dødelig i lukkede rum.

Meta's AI-assistent gik viralt i 2024 med kostråd som "spis en lille sten om dagen for at få mineraler" – et slående eksempel på, hvor absurd AI's gætteri kan blive, når systemet forsøger at lyde hjælpsomt.

Endnu mere alarmerende er medicinske hallucinationer. Et studie fra Mount Sinai Hospital i 2025 viste, at AI-chatbots ikke blot opfinder sygdomme, men også "husker" og gentager deres egne fejl i senere interaktioner. Systemer har skabt navne på ikke-eksisterende sygdomme som "Neuroflux" og leveret detaljerede, men fuldstændig opdigtede behandlingsforslag. Falske doseringer af medicin er blevet anbefalet, og ikke-eksisterende kliniske studier er blevet citeret som dokumentation.

I 2025 afslørede whistleblowere fra det amerikanske Food and Drug Administration (FDA), at agenturets AI-system "Elsa" havde hallucineret ikke-eksisterende kliniske studier under godkendelsesprocessen for nye lægemidler. Eksperter advarede om, at fejlene kunne have livsfarlige konsekvenser.

Akademisk og kulturelt sammenbrud

Chicago Sun-Times publicerede i 2024 en AI-genereret sommerlæseliste med 15 bøger – hvoraf de 10 aldrig har eksisteret. Fiktive forfattere, fiktive titler, fiktive plots.

Ph.d.-studerende verden over har rapporteret, at op til 30 procent af AI-genererede litteraturlister indeholder opdigtede kilder. Historikere advarer om hallucinerede historiske begivenheder – "borgerkrige", der aldrig fandt sted, og citater tilskrevet Einstein, der i virkeligheden stammer fra Stephen Hawking.

Compliance-eksperter har afsløret AI-systemer, der opfinder ISO-standarder med kontrolnumre, der ikke findes, hvilket kan føre til fejlslagne audits og regulatoriske bøder. Flere AI-systemer har genereret falske GDPR-paragraffer, herunder en helt opdigtet "paragraf 14 om AI-brug", som skabte forvirring blandt europæiske virksomheder.

AI har endda opfundet Nobelpriser og tildelt dem til forskere, der aldrig har modtaget dem.

Hvorfor hallucinerer AI?

For at forstå, hvordan man bekæmper hallucinationer, er det afgørende at forstå, hvordan generative AI-chatbots fungerer. Modeller som ChatGPT, Claude og Gemini er ikke databaser, der slår information op. De er statistiske sproggeneratorer, trænet på enorme mængder tekst til at forudsige det mest sandsynlige næste ord i en sætning.

Når du stiller et spørgsmål, genererer AI'en ikke et svar baseret på lagret viden i traditionel forstand. I stedet analyserer den mønstre fra sine træningsdata og konstruerer en respons ord for ord, baseret på, hvad der statistisk set virker mest plausibelt.

Fire hovedårsager til hallucinationer

1. Belønningssystem designet til flydende svar
AI-modeller er trænet til at give komplette, sammenhængende svar – ikke til at sige "jeg ved det ikke". Under træningen belønnes de for at producere tekst, der lyder overbevisende og hjælpsom, ikke nødvendigvis for sandfærdighed. Dette skaber et fundamentalt problem: Modellen vil hellere gætte end indrømme uvidenhed.

2. Manglende realtidsadgang til information
De fleste AI-modeller har ingen direkte adgang til internettet eller opdaterede databaser. De arbejder udelukkende ud fra den information, de blev trænet på – ofte data, der er måneder eller år gamle. Kun få systemer som Perplexity har indbygget realtidssøgning, men selv disse kan misfortolke kilder.

3. Træningsbias mod hjælpsomhed
Modellerne er optimeret til at være hjælpsomme og imødekommende. Hvis de ikke kender svaret, vil de ofte forsøge at konstruere et plausibelt svar frem for at indrømme begrænsninger. Dette gør dem særligt tilbøjelige til at "fylde hullerne" med opfundne detaljer.

4. Komplekse eller vage prompts
Jo mere uklar eller bred en anmodning er, jo større frihed har modellen til at gætte. Når konteksten mangler, forstærkes risikoen for, at AI'en fabricerer information for at fylde tomrummet.

Hvilke AI-modeller hallucinerer mest?

Forskellige AI-modeller har forskellige styrker og svagheder, når det kommer til hallucinationer:

ChatGPT (OpenAI): Sprogligt flydende og kreativ – men med høj risiko for faktafejl, især ved nicheemner og præcise datoer.

Gemini (Google): Hurtig og god til søgning, men medium risiko ved komplekse eller specialiserede emner.

Claude (Anthropic): Fokuserer på sikkerhed og etik med indbygget tilbageholdenhed. Lav til medium risiko, men kan være overdreven forsigtig.

Perplexity: Bygger på realtidssøgning og leverer kildehenvisninger automatisk. Lav risiko for hallucinationer, men kan misfortolke kilder eller sammensætte information forkert.

Grok (xAI): Lynhurtig, men med høj risiko for overfladiske eller fejlagtige svar.

Et simpelt test: Stil det samme spørgsmål til to forskellige modeller og sammenlign svarene. Hvor mange forskelle finder du? Hvilke detaljer stemmer ikke overens?

Prompt engineering: Nøglen til at styre AI

Prompt engineering er kunsten at formulere instruktioner til AI, så risikoen for hallucinationer minimeres. Det handler ikke kun om at stille det rigtige spørgsmål – det handler om at strukturere kommunikationen, så modellen styres mod faktuel præcision frem for kreativ fantasi.

Forskning viser, at korrekt prompt engineering kan reducere hallucinationer med 80-90 procent. Her er 20 teknikker, der hjælper dig med at få pålidelige svar.

20 teknikker til hallucinationsfri prompts

1. Præcision og kontekst

Jo mere præcis din prompt er, jo mindre frihed har AI'en til at gætte.

Dårlig prompt:
"Fortæl om fotosyntese."

God prompt:
"Forklar fotosyntese trin-for-trin, som det ville blive undervist i en dansk 9. klasses biologibog. Start med lysets absorption af klorofyl og afslut med dannelsen af glukose. Inkluder klorofylmolekylets rolle, Calvin-cyklussen og de kemiske ligninger for lysreaktionen og mørkeraktionen."

2. Struktureret output

Strukturerede formater som tabeller, JSON eller punktlister begrænser AI'ens tendens til at improvisere.

Dårlig prompt:
"Fortæl om vigtige molekyler i fotosyntese."

God prompt:
"Lav en Markdown-tabel med følgende kolonner: Molekyle | Funktion | Placering i cellen. Inkluder klorofyl, ATP, NADPH, vand og glukose."

---

3. Rolle-prompting

Ved at tildele AI'en en specifik rolle aktiverer du bestemte træningsdata og styrker det faglige fokus.

Dårlig prompt:
"Hvad siger klimarapporter om CO_2?"

God prompt:
"Du er IPCC's sjette vurderingsrapport (AR6) fra 2021-2023. Opsummer de væsentligste konklusioner om CO_2-reduktion fra Working Group III's afsnit om mitigation. Inkluder sidetal og kapitelhenvisninger."

---

4. Chain-of-Thought (CoT)

Denne teknik tvinger AI'en til at vise sin logiske ræsonnering trin for trin, hvilket gør det lettere at opdage fejl.

Dårlig prompt:
"Beregn strømmen i et RLC-kredsløb med R=100Ω, L=0.5H, C=10μF ved 50Hz."

God prompt:
"Beregn strømmen i et RLC-kredsløb trin-for-trin:
Trin 1: Tegn kredsløbsdiagrammet.
Trin 2: Skriv formlen for impedans.
Trin 3: Indsæt værdierne R=100Ω, L=0.5H, C=10μF, f=50Hz.
Trin 4: Tjek enhederne.
Trin 5: Vis slutresultatet og verificer ved at sammenligne med en ingeniørhåndbog eller Wolfram Alpha."

---

5. Kildekrav

Insister på verificerbare kilder for at undgå falske DOI-numre, ISBN-koder ellerLinks.

Dårlig prompt:
"Opsummer IPCC's klimarapport."

God prompt:
"Opsummer kapitel 2 i IPCC AR6 Working Group I-rapporten. Inkluder sidetal, direkte citater og links til den officielle rapport på ipcc.ch."

---

6. Few-shot prompting

Vis AI'en eksempler på det ønskede format, så den lærer af mønstre.

Dårlig prompt:
"Analysér denne sætning: 'Okay bil'."

God prompt:
"Klassificer følgende sætninger som Positive, Negative eller Neutrale:
Eksempel 1: 'Jeg elsker denne bil!' → Positiv
Eksempel 2: 'Dækket er ødelagt' → Negativ
Nu: 'Okay bil' → ?"

---

7. Self-verification

Bed AI'en om at tjekke sit eget svar og markere usikkerheder.

Dårlig prompt:
"Hvad er hovedstaden i Australien?"

God prompt:
"Hvad er hovedstaden i Australien? Efter du har givet svaret, marker eventuelle dele, du er usikker på, med [USIKKER]."

---

8. Iterativ raffinering

Forbedrer kvaliteten trinvis ved at bede om revideringer.

Dårlig prompt:
"Skriv et resume om kvantemekanik."

God prompt:
"Skriv et udkast til et 300-ords resume om kvantemekanik.
Derefter: Tilføj akademiske kilder.
Til sidst: Forkort til 200 ord uden at miste væsentlige pointer."

---

9. Multi-step reasoning med verifikation

Sikrer korrekthed ved at verificere hvert trin i en proces.

Dårlig prompt:
"Løs ligningen: 3x + 7 = 22."

God prompt:
"Løs ligningen 3x + 7 = 22 trin-for-trin:
Trin 1: Isoler x-termen.
Trin 2: Divider med koefficienten.
Trin 3: Tjek løsningen ved at indsætte værdien i den oprindelige ligning."

---

10. "Don't make up"-instruktion

En direkte ordre om ikke at opfinde information.

Dårlig prompt:
"Hvad er den nyeste forskning i kvantecomputere?"

God prompt:
"Opsummer den nyeste peer-reviewede forskning i kvantecomputere fra 2024-2025. Hvis du ikke har verificerede data, skriv: 'Ingen data tilgængelig'."

---

11. Retrieval-Augmented Generation (RAG)

Grounder AI'ens svar i specifikke dokumenter eller links.

Dårlig prompt:
"Hvad siger GDPR om databehandling?"

God prompt:
"Brug kun information fra GDPR-teksten (EU 2016/679) tilgængelig på eur-lex.europa.eu. Opsummer artikel 6 om lovligt grundlag for behandling."

---

12. Temperaturkontrol

Lav temperatur (0-0.3) reducerer kreativitet og øger faktuel præcision.

Dårlig prompt:
"Hvad er pi?"

God prompt:
"Angiv værdien af pi med 10 decimaler. Brug temperatur 0.2 for maksimal faktuel nøjagtighed."

---

13. Negative instruktioner

Fortæl eksplicit, hvad AI'en ikke må gøre.

Dårlig prompt:
"Forklar fremtiden for vedvarende energi."

God prompt:
"Forklar aktuelle trends inden for vedvarende energi baseret på IEA's World Energy Outlook 2023. Ingen spekulation. Ingen hypotetiske scenarier."

---

14. Cross-check med alternative kilder

Kræv, at information gentages i flere uafhængige kilder.

Dårlig prompt:
"Hvad er konsekvenserne af global opvarmning?"

God prompt:
"Opsummer kun de klimaeffekter, der nævnes i både IPCC AR6, NASA's klimadata og EU's klimarapport 2023. Ignorer information, der kun findes i én kilde."

---

15. Critic Mode

AI'en kritiserer sit eget svar og identificerer svagheder.

Dårlig prompt:
"Skriv om renewable energy."

God prompt:
"Skriv et afsnit om vedvarende energi. Derefter: Lav en liste over mulige fejl, usikkerheder eller mangler i dit svar."

---

16. Konfidensscore

Gør usikkerhed eksplicit med procentvise scores.

Dårlig prompt:
"Hvornår blev penicillin opdaget?"

God prompt:
"Hvornår blev penicillin opdaget? Tilføj [Sikkerhed: XX%] efter dit svar baseret på din tillid til informationen."

---

17. Prompt chaining

Opdel komplekse opgaver i flere separate prompts.

Dårlig prompt:
"Opsummer og analyser IPCC-rapporten."

God prompt:
Prompt 1: "Lav en liste over kapitler i IPCC AR6."
Prompt 2: "Opsummer kapitel 3 fra listen."
Prompt 3: "Identificer de tre vigtigste konklusioner."

---

18. Grounding phrases

Brug vendinger, der signalerer krav om fakta.

Dårlig prompt:
"Hvad tror du om klimaforandringer?"

God prompt:
"Baseret udelukkende på verificerede videnskabelige data: Hvad er konsensus om menneskeskabte klimaforandringer ifølge IPCC?"

---

19. Kontekstualisering med tid og sted

Specificer geografisk og tidsmæssig kontekst.

Dårlig prompt:
"Hvad siger loven om persondatasikkerhed?"

God prompt:
"Forklar GDPR-reglerne for databehandling i EU pr. december 2024. Inkluder eventuelle ændringer efter den oprindelige forordning fra 2016."

---

20. Hybrid prompts (tekst + kode)

Kombiner tekstforklaring med kodebaseret validering.

Dårlig prompt:
"Beregn moms på 1.250 kr."

God prompt:
"Beregn moms (25%) på 1.250 kr. Forklar beregningen, og generér derefter Python-kode, der verificerer resultatet."

Master Prompt: Din universelle skabelon

Her er en skabelon, der kombinerer de mest effektive teknikker:

”*SYSTEM:*

[SPØRGSMÅL] =

[Kontekst] =

Du må ikke bruge tekst i <<meta>>-blokke til nogen del af dit svar.

Du optræder som en højt kvalificeret akademisk ekspert inden for det relevante

fagområde (f.eks. ph.d. i datalogi, professor ved Aarhus Universitet, med omfattende

erfaring i AI og mere end 400 videnskabelige publikationer). Din rolle er at levere

præcise, konsistente, evidensbaserede og reproducerbare svar. Du formulerer

vurderinger eksplicit, begrunder dem metodisk, adskiller fakta fra fortolkning og

angiver klart, hvor antagelser eller begrænsninger indgår.

Du skriver altid i idiomatisk, formelt akademisk dansk med fuld grammatisk

korrekthed. Sproget skal være præcist, velstruktureret og syntaktisk stringent.

Undgå anglicismer, talesprog, uformelle konstruktioner, upræcise vendinger og

unødige gentagelser.

Når du forklarer tekniske, metodiske eller teoretiske forhold, skal du anvende

korrekt og etableret fagterminologi. Fremstillingen skal være logisk sammenhængende,

fagligt stringent og understøttet af klar argumentation. Du benytter en neutral,

objektiv og analytisk tone og angiver eksplicit, når eksterne data, kilder eller

verifikationer er nødvendige.

Din kommunikation skal afspejle akademiske normer: klar problemformulering,

præcise definitioner, stringent argumentation, konsistens mellem afsnit, korrekt

terminologi og eksplicit opdeling mellem fakta, analyse og vurdering. Du må ikke

forsimple indholdet, medmindre brugeren udtrykkeligt anmoder om det.

Du følger altid brugerens faglige niveau og hensigt, men du går aldrig på kompromis

med korrekthed, terminologisk præcision, metodisk stringens eller akademisk kvalitet.

USER PROMPT:

Du er en kritisk ekspert. Besvar spørgsmålet [SPØRGSMÅL] i konteksten [Kontekst]

ved at følge nedenstående outputstruktur præcist. Dette er ikke en anmodning om

interne kæder af tanker, men om kompakte og ikke-sensitive resuméer af din

rationelle vurdering.

Outputstruktur (strengt):

1) Kort konklusion (1–2 sætninger): det endelige valgte svar.

*2) Tre uafhængige *hypotesepathways* (kortfattede, hver maks. 5–8 sætninger).*

For hver pathway:

a) Hypotesens konklusion (1 sætning).

b) Nøgleargumenter eller evidens.

c) Nødvendige antagelser.

d) Hvilken ekstern verifikation eller kilde der ville be- eller afkræfte den.

3) Vælg én af de tre pathways som mest konsistent — forklar kort hvorfor (1–3 sætninger).

4) Sikkerhedsvurdering: angiv procent (0–100%) for sikkerheden i konklusionen samt de

to vigtigste usikkerheder eller risikofaktorer, der kunne ændre vurderingen.

*5) Accept-kriterium: Hvis sikkerheden < 99%, skriv **kun** "Usikker" (uden yderligere forklaring).*

6) Konsistenskontrol: angiv "Konsistenskontrol: OK" eller

"Konsistenskontrol: Fundet modstrid". Hvis modstrid findes, skriv kun "Usikker".

7) Afslut med en kort liste (maks. 4 punkter) over hvilke konkrete data, målinger

eller beviser jeg kan levere for at hæve sikkerheden til ≥ 99%.

Yderligere krav:

- Brug korte, klare sætninger. Behold nummereringen som angivet.

- Hvis du anvender tidsfølsomme fakta, angiv kilde (titel + år) eller skriv "kilde påkrævet".

- Maksimalt 400 ord i alt, medmindre jeg anmoder om en dybere gennemgang.

- Stil ikke afklarende spørgsmål — besvar ud fra den givne kontekst.

Spørgsmål: [SPØRGSMÅL] ”

Denne Master Prompt kombinerer rolle-prompting, kildekrav, Chain-of-Thought, self-verification, konfidensscore, struktureret output, negative instruktioner og kritisk gennemgang. Resultatet er en reduktion af hallucinationer på op til 90-95 procent.

Øvelse: Bliv din egen prompt engineering-ekspert

Tag en af de dårlige prompts fra eksemplerne ovenfor. Omformulér den ved hjælp af mindst tre af teknikkerne. Test den i ChatGPT eller Claude. Sammenlign resultatet med det oprindelige svar.

Du vil opdage, at præcision, struktur og kritisk tænkning er nøglerne til at få AI til at arbejde for dig – ikke mod dig.

AI kan lyve. Men med de rette værktøjer kan du styre sandheden.

Efterskrift: Sandheden om AI – og ansvaret ligger hos os

Kunstig intelligens er ikke magisk. Den er ikke alvidende. Den er et statistisk sprogværktøj, der er trænet til at lyde overbevisende – også når den tager fejl. Hallucinationer er ikke en sjældenhed, men en indbygget risiko, der følger med teknologien.

Vi står ved begyndelsen af en ny æra, hvor AI kan skrive essays, kode software og analysere data hurtigere end nogensinde før. Men med hastigheden følger faren: Hvis vi blindt stoler på maskinen, kan konsekvenserne være katastrofale – fra juridiske skandaler til livsfarlige råd.

Derfor er **prompt engineering** ikke bare en teknisk finesse. Det er en kritisk kompetence for alle, der arbejder med AI – fra forskere og journalister til advokater og læger. Med de rette teknikker kan vi reducere fejlene med op til 90 procent. Vi kan gøre AI til en pålidelig assistent i stedet for en uforudsigelig risikofaktor.

Men ansvaret ligger hos os. Det er os, der skal stille de rigtige spørgsmål, kræve kilder og tænke kritisk. AI kan lyve – men vi kan lære at kontrollere sandheden.

Spørgsmålet er ikke, om vi skal bruge AI. Spørgsmålet er, om vi tør bruge den klogt.