

# Fra hallusinasjoner til kontroll: Når kommer AI-agentene til revisjon og regnskap?

På bare tre år har generativ kunstig intelligens utviklet seg fra rent tekstbaserte verktøy til multimodale systemer som kan resonnerer og utføre oppgaver dels autonomt. Samtidig har utfordringene med hallusinasjoner, manglende presisjon, og risiko for kompetanseforvitring blitt tydelige – særlig i tillitsbaserte tjenester som revisjon, regnskap og jus. Selv om AI-verktøyenes presisjon stadig øker, er det viktig med robuste kontrollmekanismer, kompetanseutvikling og en bevisst strategi for å lykkes med å ta AI i bruk.



Ph.D.  
**Lars Erlend Leganger**  
Chief Technical Officer,  
24SevenOffice

## Fra gjennombrudd til hallusinerende hverdagsverktøy

Da OpenAI lanserte ChatGPT i november 2022, ble generelt anvendbar kunstig intelligens (AI, fra engelsk «artificial intelligence») for første gang tilgjengelig for folk flest. De underliggende store språkmodellene (LLM, fra engelsk «large language models») kunne forstå nesten enhver tekstlig forespørsel, og svarene var ofte overraskende gode. Men LLM-teknologi baserer seg på sannsynligheter, og det hender at modellene genererer svar som kanskje er sannsynlige sett i kontekst av modellens treningsdata, men som rent faktisk er feil – såkalte hallusinasjoner. De fleste har vel nå hørt om den herostratisk berømte advokat Schwartz, som misforstod teknologiens kapabiliteter og begrensninger, og endte opp med å forarge en dommer da han refererte til hallusinert ikke-eksisterende rettsprak-

sis i et prosesskriv<sup>1</sup>. I en studie ved Stanford FRA januar 2024 stilte Dahl og kolleger generelle juridiske spørsmål til et utvalg kommersielle AI-verktøy fra 2023 for generell bruk, og fant hallusinasjoner i over halvparten av svarene<sup>2</sup>. I tillegg til de generelle AI-verktøyene finnes domenespesifikke løsninger som lover mer presisjon, men heller ikke disse er feilfrie. I en oppfølgingsstudie fra mai 2024 fant Stanford-forskerne at

jus-spissede AI-verktøy fra samme periode, som i enkelte tilfeller ble promotert som «garantert hallusinasjonsfrie», fortsatt gav hallusinerte svar på mellom 17 og 33 % av test-spørsmålene<sup>3</sup>.

Arbeidsflyter i tillitsbaserte tjenester – som juridisk rådgivning, revisjon og regnskap – har lav toleranse for feil, kanskje spesielt feil forårsaket av brudd med gjengs metodikk, noe bruk av AI fort kan oppfattes som. Bredt anerkjent beste praksis metodebeskrivelser endrer seg langsommere enn teknologifronten. Kjennetegn på dem som lykkes med bruk av kunstig intelligens de siste årene, har vært nysgjerrighet og mot til å pilotere AI-teknologi bredt i verdikjedene, kombinert med målrettet opplæring av ansatte rundt teknologiens styrker og svakheter, anvendelige risikomitigeringsprosesser og vektlegging av at AI-bruk skal være sporbart og etterprøvbart<sup>4</sup>.

- 1 Det hører til historien at det som gjorde dommeren virkelig sint, var at Schwartz insisterte på at de hallusinerte dommene eksisterte selv etter at deres eksistens hadde blitt trukket i tvil: «... there is nothing inherently improper about using a reliable artificial intelligence tool for assistance. But existing rules impose a gatekeeping role on attorneys to ensure the accuracy of their filings. Rule 11, Fed. R. Civ. P. Peter LoDuca, Steven A. Schwartz and the law firm of Levidow, Levidow & Oberman P.C. (the «Levidow Firm») (collectively, «Respondents») abandoned their responsibilities when they submitted non-existent judicial opinions with fake quotes and citations created by the artificial intelligence tool ChatGPT, then continued to stand by the fake opinions after judicial orders called their existence into question ... But if the matter had ended with Respondents coming clean about their actions shortly after they received the defendant's March 15 brief questioning the existence of the cases, or after they reviewed the Court's Orders of April 11 and 12 requiring production of the cases, the record now would look quite different. Instead, the individual Respondents doubled down and did not begin to dribble out the truth until May 25, after the Court issued an Order to Show Cause why one of the individual Respondents ought not be sanctioned». <https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/>
- 2 Dahl et al. 2024, «Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models». <https://arxiv.org/abs/2401.01301>

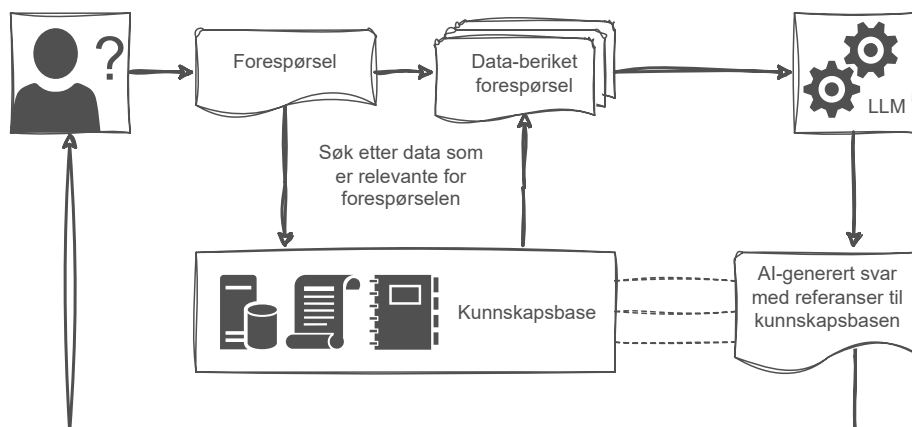
- 3 Magesh et al. 2024, «Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools». <https://arxiv.org/abs/2405.20362>
- 4 McKinsey & Company 2024, «The state of AI in early 2024: Gen AI adoption spikes and starts to generate value». <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024>

## Profesjonsspissing, dataforankring og resonnering gir bedre presisjon

De siste årene har imidlertid modellene blitt kraftigere, fått tilgang på mer data under trening og bruk, og fått evnen til å forstå flere typer input. I dag kan bredt tilgjengelige generelle AI-verktøy analysere og generere tekst, bilder, tall og lyd i én og samme arbeidsflyt. For eksempel kan et hylleware-verktøy lese en samling skannede bilag, hente ut og sammenstille relevante tall i et regneark, og svare muntlig på oppfølgings-spørsmål i én og samme prosess. Slik multimodalitet gjør verktøyene mer fleksible og reduserer behovet for manuell «klipp-og-lim» i AI-støttede arbeidsflyter.

Generalistverktøy som ChatGPT, Gemini og Claude har fått konkurranse av selskap som spesialiserer seg på verktøy skreddersydd for spesifikke profesjoner og arbeidsflyter, som Harvey (for jurister) og Medbric (for leger). Profesjonsspesifikke verktøy kan gjøre det enklere og tryggere å anvende kunstig intelligens i kritiske og fagtungte kjerneprosesser, spesielt der det finnes profesjonsspesifikke behov knyttet til datakilder, integrasjoner, sertifisering og godkjenninger.

Mange selskap har tatt i bruk retrieval-augmented generation (RAG) for å redusere hallusinasjoner i sine AI-verktøy. Før LLM-modellen mottar brukerens forespørsel, suppleres den med resultatene fra søk etter relevant kontekst i egne databaser (f.eks. lovtekster, standarder, regnskapsregler osv.), og modellen instrueres i å bygge på disse kontekstuelle dataene når den genererer svar, se figur 1. Slik kan AI-svarene forankres i faktiske kilder som brukeren kan etterprøve. Men RAG avhenger av god datakvalitet i kontekstdatabasene – gitt data som er ufullstendige eller har uventet utforming, kan modellene fort falle tilbake til hallusinererte svar. RAG-implementeringer styrker verktøyenes presisjon, men det må ikke bli en hvilepute for menneskelig kvalitetssikring.



Figur 1: RAG forankrer språkmodellens svar i faktiske kilder, slik at brukeren kan kontrollere grunnlaget for svaret.

I 2024 ble også de første resonnerende modellene bredt tilgjengelige. Dette er modeller som lager en plan for hvordan oppgavene de får bør løses (ofte bestående av mange trinn), før de til slutt setter sammen arbeidet til et endelig svar, se figur 2. I deres egne laboratorietester har OpenAIs nylig lanserte GPT-5 betraktelig lavere feilrater enn deres tidligere modeller, spesielt når den jobber i resonneringsmodus<sup>5</sup>.

ChatGPT 5 Thinking

Hvilken påvirkning har formuesskatten på norsk økonomi?

Tenkte i 1m 31s

- Jeg må svare på norsk om formuesskattens innvirkning på den norske økonomien. Dette er et politisk og økonomisk tema. Jeg bør oppdatere med relevant informasjon og referere til pålitelige kilder som SSB, Norges Bank, og nyere studier. Skal også undersøke emigrasjon og kapitalflukt siden skatteendringene i 2022.
- Jeg vurderer relevante kilder som SSB-rapporten 2023/36, NHH og Frisch Senterets forskning, samt papirene fra Halvorsen & Høiseith om formuesskatt og entreprenørskap. Jeg bør søke etter tall om utflytting til Sveits, investeringseffekter og skatteutvalgets vurderinger. Jeg planlegger søk etter flere akademiske kilder.

Searching for analysis on formuesskatt i Norge

www.ssb.no
www.regjeringen.no
www.huseierforbundet.no
10 mer

- Vi må være nøye med kildene. E24 er pålitelig for økonominyheter, Bergens Tidende gir generell informasjon, og Civita er et kjent tankesmie. Jeg bør inkludere data fra SSB, forskning fra Frisch Senteret og NOU-rapporter. Jeg vurderer også å bruke en akademisk artikkel om formuesskatt og entreprenørskap.

Verifisere kilder og finne relevant forskning

www.regjeringen.no

- Forskningen ser ut til å vise at formuesskatten har en positiv effekt på sysselsetting, men liten effekt på investeringer, og at likviditets

Analyserer effekter av formuesskatt på arbeid

www.ssb.no

Figur 2: Resonnerende modeller planlegger løsningen steg for steg før de setter sammen et endelig svar, noe som kan redusere feilrate i komplekse oppgaver.

5 OpenAI 2025, «Introducing GPT-5». <https://openai.com/nb-NO/index/introducing-gpt-5/>

## Økt adopsjon og høye forventninger

Innen revisjons- og rådgivningstjenester har økt AI-fleksibilitet og -presisjon ført til økt adopsjon og høyere fremtidsforventninger. I 2025-utgaven av Thomson Reuters' Future of Professionals-undersøkelse oppgir 53 % av respondentene at de allerede ser avkastning på sine AI-investeringer, og 8 av 10 respondenter forventer at AI vil medføre store endringer i arbeidshverdagen deres i løpet av de neste fem årene<sup>6</sup>. Samtidig har de klassiske adopsjonsbarrierene ikke forsvunnet: Manglende nøyaktighet anses som den viktigste barrieren for videre utrulling.

Det er også en utbredt frykt blant fagfolkene for at utstrakt AI-bruk skal føre til at faglig kompetanse forvitrer. Her er strategi og ledelse viktig. Selskap som har en tydelig AI-strategi, og der ledelsen er aktivt involvert i AI-arbeidet, er også de som oftest rapporterer avkastning på AI-investeringene.

## Er neste steg autonome AI-agenter?

Multimodalitet og resonnering er viktige skritt på veien til autonome AI-agenter: Systemer som planlegger og utfører oppgaver på egen hånd via integrasjoner, eller direkte styring av grensesnittet i brukerens datamaskin. Det er en utbredt oppfatning at AI-agenter er det neste steget i AI-kappløpet, og de store aktørene posisjonerer seg deretter, med (i skrivende stund fremdeles noe umodne) produkter som Anthropic's Claude Computer Use Tool<sup>7</sup>, og OpenAI's ChatGPT Agent<sup>8</sup>.

Wolters Kluwer anslår stort potensiale for AI-agenter innen revisjon og

regnskap<sup>9</sup>. De ser en nær fremtid der AI-agenter for eksempel kan lytte til det som sies i et kundemøte, automatisk identifisere hva kunden etterspør, planlegge hvordan oppgaven kan løses, hente relevante data fra f.eks. regnskapssystemet, og klargjøre (utkast til) leveransen uten menneskelig detaljstyring. På lengre sikt spår de at AI-agenter vil kunne jobbe enda mer uavhengig, og for eksempel få ansvar for å overvåke regelendringer, holde i compliance-prosesser og produsere individtilpasset kundekommunikasjon i stor skala. Som vanlig er tanken at dette skal frigjøre (menneskelige) revisorer og regnskapsførere til mer rådgivende oppgaver.

## *Multimodalitet, resonnering og formell verifikasjon endrer hvordan kunstig intelligens kan brukes i profesjonelle tjenester.*

Store vyer og muligheter til tross, i skrivende stund er AI-agentene relativt trege og kostbare, og det oppstår betydelige sikkerhetsutfordringer når AI-agenter får tilgang på menneskelige brukers verktøy og data. Den som vil pilotere autonome AI-agenter i kritiske deler av verdikjeden, bør ha stålkontroll på tilgangsstyring og monitorering, ellers kan advokat Schwartz fort få selskap i «ufrivillig berømt for å stole for mye på AI»-kategorien.

## Høyere krav til kvalitet hos AI enn hos mennesker

Et interessant funn i Thomson Reuters' Future of Professionals 2025-rapport er at 91 % av respondentene mener at AI-systemer må være mer nøyaktige enn mennesker, og 41 % mente at resultatene må være 100 % korrekte før man kan bruke dem uten menneskelig kontroll. Da er det langt frem til

autonome AI-systemer i tillitsbaserte tjenester. Enn så lenge må jurister kontrollere alt Harvey's verktøy produserer, og leger som bruker Medbrics' journal-AI går gjennom notatene før de lagres i pasientjournalen. Slik manuell kontroll skalerer dårlig når dagens AI-verktøy og morgendagens AI-agenter skal gjøre mer komplekse oppgaver. For å hindre at menneskelig arbeidskraft blir en automatiseringsflaskehals, trengs metoder som automatisk kan validere og sikre at AI-genererte leveranser tilfredsstiller en gitt verdikjedes krav til presisjon og nøyaktighet.

## Deterministisk logikk kan tøyle probabalistisk AI

En lovende tilnærming er formell spesifisering og verifikasjon. Dette er teknikker fra matematikkens verden der man erstatter potensielt tvetydige prosessbeskrivelser, policyer og krav med formell logikk. I stedet for å gjøre stikkprøver og kontroller, inkluderer man alle mulige inputs og bruker matematikk og logikk for å bekrefte eller avkrefte at et systems adferd er i tråd med de formelt spesifiserte kravene, se figur 3.

Formelle metoder gir sterkere garanti for sikkerhet og pålitelighet enn tradisjonell programvaretesting, og brukes mye innen høyrisikoområder som atomkraft, luftfart, medisin og kritisk infrastruktur. Med løsninger som Amazons Automated Reasoning Checks kan man, med utgangspunkt i prosessbeskrivelser, standarder og policyer generere maskinlesbare formelle regler som muliggjør automatisk kontroll av om AI-systemenes konklusjoner er holdbare, gitt den formelle modellen av den aktuelle prosessens «internrett».

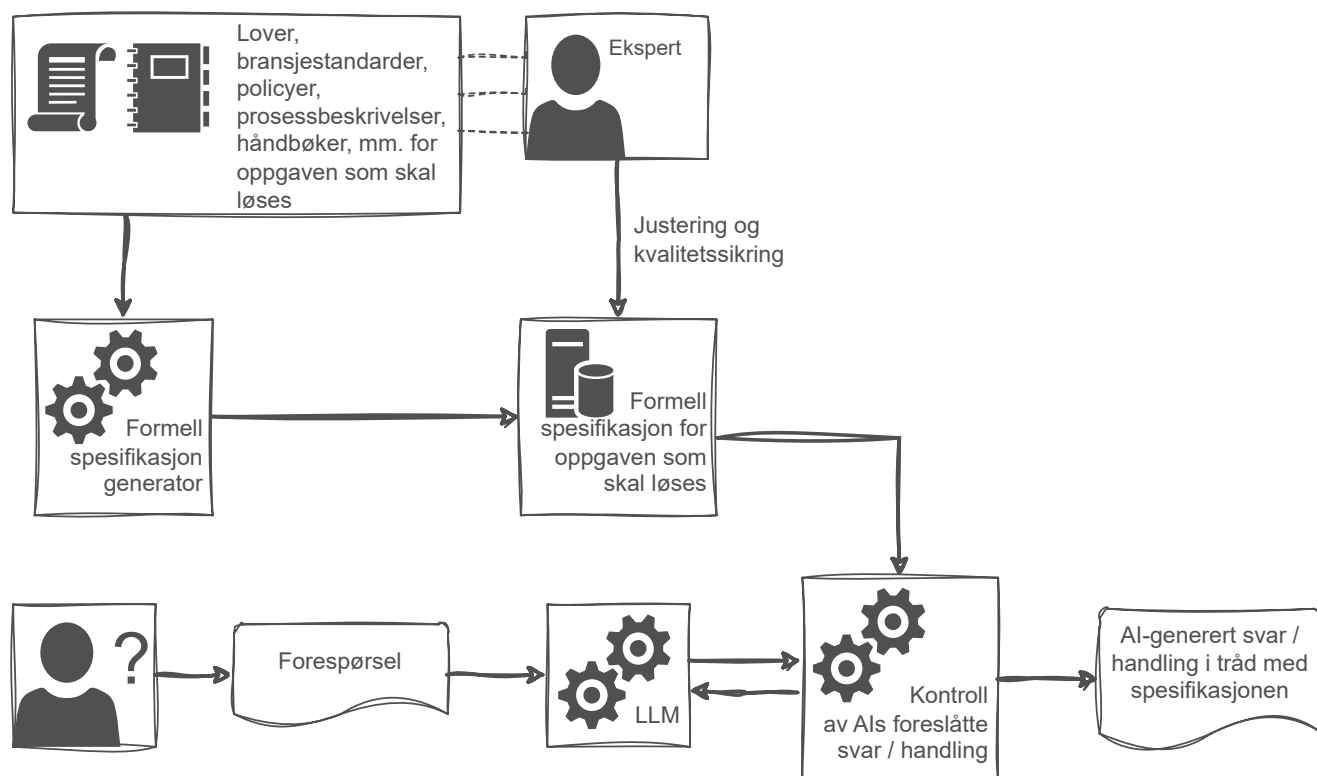
Formell verifikasjon av AI-systemer er relativt ny teknologi (Amazon lanserte sin løsning i august 2025), og det er foreløpig begrenset med erfaringer fra praktisk bruk, men for bedrifter med høye krav til presisjon som ønsker å ta AI i bruk dypt inne i bedriftskritiske prosesser, kan det være verdt å utforske

6 Thomson Reuters 2025, «Future of Professionals Report 2025». <https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/reports/future-of-professionals-report-2025.pdf>

7 Anthropic 2025, «Computer Use Tool». <https://docs.anthropic.com/en/docs/agents-and-tools/tool-use/computer-use-tool>

8 OpenAI 2025, «Introducing ChatGPT Agent». <https://openai.com/nb-NO/index/introducing-chatgpt-agent/>

9 Wolters Kluwer 2025, «The power of AI: What accounting and tax professionals need to know». <https://www.wolterskluwer.com/en/expert-insights/the-power-of-ai>



Figur 3: Policyer, standarder og krav kan oversettes til logiske regler som muliggjør formell deterministisk kontroll av mulighetsrommet for AI-systemers handlinger.

bruk av formell verifikasjon på de mest risikofylte delene av arbeidsflyten, i første omgang kombinert med tradisjonell menneskelig kvalitetssikring.

### Balansegangen fortsetter

På bare tre år har vi gått fra enkle tekstgeneratorer til multimodale modeller som kan resonnerer og (i begrenset grad) handle autonomt. Samtidig har LLM-modellers iboende svakheter – hallusinasjoner, ugjennomsiktighet og risiko for misbruk – blitt tydelige. For å lykkes med AI handler det om mer enn å anskaffe den nyeste teknologien.

Tillitsbaserte tjenester som revisjon, regnskap og juridisk rådgivning må kombinere og balansere mot å innovere med robuste prosessuelle og tekniske kontrollmekanismer, og kontinuerlig kompetanseutvikling. Ansatte som forstår teknologiens styrker og svakheter, kan få mer frihet under ansvar, og kan bidra med sin domenekunnskap til å finne nye anvendelsesmuligheter.

Tydelige og risikodifferensierte policyer kan gjøre det enklere å automatisere egnede oppgaver, og samtidig hindre at AI slippes løs på oppgaver som fremdeles krever grader av menneskelig

kontroll. RAG kan brukes for å forankre AI-genererte svar i verifiserbare kilder.

Og i fremtiden kan kanskje teknikker fra formelle metoder brukes til å omdanne policyer, standarder og regelverk til maskinlesbare formelle regler som kan danne ytre rammer for AI-verktøy og -agents handlingsrom. Leveranser fra fullstendig autonome AI-agenter i høyrisiko-bransjer ligger fjernere frem i tid, men i det myldrende AI-feltet, der store og små gjennombrudd har kommet nærmest månedlig de siste årene, er ikke nødvendigvis «fjern fremtid» så langt unna.