**Til:** Utvalget om kunstig intelligens i høyere utdanning

**Fra:** Thobias Melfjord Knudsen, **vinner av NM i AI 2026** og B.Sc-student i informatikk ved NTNU

**Dato:** 18. mai 2026

**Tema:** Forslag om fagspesifikke emner i kritisk og effektiv KI-bruk

> **TL;DR:** Etabler fagspesifikke emner i kritisk og effektiv KI-bruk der oppgavene er **kalibrert vanskelig nok til at direkte LLM-besvarelse gir få poeng**, slik at studenten lærer å tenke kritisk med KI som hjelpemiddel. **Gjør ukritisk KI-bruk selvutslettende.** Effekten **måles empirisk via pre/post-test** med samme institusjonsmodell. Hvert studieprogram avgjør selv om emnet skal være obligatorisk. Forslaget operasjonaliserer Påstand III og Anbefaling 4 i utvalgets foreløpige notat. Det utfordrer ikke retningen, men fyller den ut.

> **Om grunnlaget:** Forslaget bygger på praktisk erfaring med formatet det foreslår. NM i AI 2026 er bygd etter samme prinsipp: deltakerne får komplekse, virkelighetsnære problemstillinger og fri tilgang til alle KI-verktøy de selv velger. Det er ikke en algoritmekonkurranse. Det er evnen til kritisk og effektiv KI-bruk under tidspress som skiller deltakerne. Det er også slik mange yrker fungerer i dag. Flere av designgrepene under (særlig 5.1.5 om oppgavekalibrering og 5.6 om tilgang under eksamen) er operasjonaliseringer av det jeg observerte virker i det formatet.

## 1. Sammendrag

Utvalget peker i notatet *Foreløpige vurderinger* (16. desember 2025) på en sentral spenning: KI gir både risiko for «metakognitiv latskap» (Påstand IV) og potensial for å styrke læring (Påstand III). De foreløpige anbefalingene er hovedsakelig forsvarsorienterte. Altså flere kontrollerte eksamensformer for å beskytte vurderingsintegriteten. Det er nødvendig, men ikke tilstrekkelig.

Jeg foreslår et komplementært tiltak: utvalget bør anbefale at hvert fagrelevant studieprogram tilbyr minst ett emne i **kritisk og effektiv KI-bruk**, med felles pedagogisk struktur men fagspesifikt innhold. Hvert studieprogram avgjør selv om emnet skal være obligatorisk for sine studenter eller valgfritt. Emnet skal gi oppgaver som er for komplekse og nyanserte til å løses ved ukritisk KI-bruk («gjør X for meg»), og trene studenten både i kritisk evaluering av KI og i effektiv samhandling med den. Forslaget operasjonaliserer notatets Påstand III og Anbefaling 4, ikke ved å forby KI, men ved å gjøre ukritisk bruk selvutslettende.

To designgrep er sentrale og skiller forslaget fra prinsipperklæringer: (a) **oppgavekalibrering** der direkte LLM-besvarelse gir få poeng, slik at karakterfordelingen drives av studentens egen kritiske syntese, ikke av modellytelse (en strategisk omgåelse av KI-immun-debatten); (b) **pre/post-test** med samme institusjonsmodell, som gjør modellen empirisk testbar og lar designet justeres over tid. Sammen er disse to grepene det som gjør modellen kalibrert mot KIs egen styrke og målbar i effekt.

Forslaget bygger på samme premiss som **UH-didaktikk-nettverkets innspill** (17.02.2026, kap. 2): «Hovedutfordringen knyttet til KI i høyere utdanning er ikke primært teknologisk, men pedagogisk og epistemisk.» Det svarer også på **NSOs etterlysning** (innspill 12.01.2026) av at «utvalget ikke foreslår gode tiltak knyttet til studentens egen bruk av KI».

## 2. Det jeg støtter i notatet

- **2.1 Anbefaling 4:** Universiteter bør ikke jage KI-immune oppgaver.
- **2.2 Påstand III:** God KI-bruk kan styrke læring.
- **2.3 Påstand II:** KI vil sannsynligvis lede til nye arbeidsmåter og økt automatisering, men vi mangler kunnskapsgrunnlag for å foregripe utviklingen. Læringsinnholdet må derfor tilpasses kontinuerlig.
- **2.4 Notatets kap. 2:** Fagspesifikk KI-kompetanse må inn i bredden av studietilbud.

Jeg støtter dette uten reservasjoner. Forslaget under er ment som operasjonalisering av disse, ikke som alternativ.

## 3. Premisset: KI som hjelpemiddel, ikke erstatning

Mange er redde for KI. Den frykten er forståelig, men ikke produktiv. Vi må lære å bruke KI godt heller enn å unngå den, fordi det er gjennom faktisk bruk vi tydeligst ser hvor verktøyet ikke er nok alene.

Jeg har brukt språkmodeller aktivt i tre år. Forbedringen i KIs evne til å løse komplekse, nyanserte problemstillinger har vært beskjeden det siste året. Mye av den oppfattede framgangen kommer fra optimalisering mot benchmarks. Det er flere grunner til at narrativet om «konstant forbedring» bør møtes med kritisk distanse:

- **3.1 Insentiv-skjevhet hos leverandøren.** KI-selskapene tjener på inntrykket av kontinuerlig forbedring; markedsføring, prising og kapitaltilgang henger på det. Effekten er todelt: (a) brukere overvurderer faktisk fremgang, og oppfattet ytelse på eldre modeller faller før nye lanseres; (b) narrativet om «eksponensiell forbedring» holdes oppe selv når reell ytelse på

komplekse oppgaver flater ut. Konsekvensen for studenten: institusjonen kan ikke stole på ekstern API som stabilt måleinstrument, og studenten må læres opp til å vurdere modellens reelle ytelse på egne komplekse problemer i stedet for å akseptere leverandørens benchmark-narrativ, og å bytte leverandør når en modell faktisk blir dårligere.

- **3.2 Benchmark-optimalisering ≠ reell forbedring.** Benchmarks er noe selskapene aktivt optimaliserer mot. Forbedring på benchmark er ikke det samme som forbedring på faktiske problemstillinger som ligger utenfor treningsdata.
- **3.3 Dokumentert ytelsesdrift uten varsel.** Stanford-studien til Chen, Zou & Zaharia (2023, *«How Is ChatGPT's Behavior Changing over Time?»*) fant at GPT-4 sin nøyaktighet på primtall-identifikasjon falt fra 97,6 % til 2,4 % mellom mars og juni 2023. Tilsvarende uvarslede modellendringer er bekreftet av flere kommersielle leverandører (OpenAI, Google, Anthropic) i 2025–2026. Konsekvensen er den samme uansett leverandør: institusjonen kan ikke stole på at en ekstern API leverer samme produkt i mai som i januar.
- **3.4 Konklusjon: operasjonell definisjon av «samme modell».** «Samme modell» gjennom semesteret må operasjonelt bety **samme vekter + samme system-prompt + samme sampling-parametre + samme inference-konfigurasjon + ingen kontinuerlig finetuning på studentdata**. Det er praktisk bare oppnåelig med institusjonsmodell hostet lokalt med versjonslås, ikke med ekstern API. Pre/post-tester med slik låst modell-konfigurasjon er det eneste pålitelige måleinstrumentet for å skille mellom modellforbedring og brukerlæring.

Det fundamentale: en stor språkmodell er statistikk. Den gjengir mønstre fra treningsdata. Den tenker ikke kritisk på samme måte som et menneske som veier verdier, vurderer kontekst og bringer egen erfaring til bordet. På problemer som ligger utenfor treningsdata, eller som krever syntese av motstridende hensyn, treffer den dårligere.

Tekna formulerer det godt i sitt innspill (06.10.2025): «Kritisk bruk av KI er kun mulig hvis man har noe å tenke kritisk ut fra ... Bruk av kalkulator hjelper ikke en som ikke kan regne.» Det er nettopp dette emnet som foreslås under, er ment å trene: fagkunnskap først, KI som forsterker, ikke som erstatning.

## 4. Det notatet mangler

Notatet beskriver godt *hva* KI-kompetanse innebærer, men gir få konkrete eksempler på *hvordan* undervisning bør designes for å fremme kritisk KI-bruk i praksis. Uten konkrete modeller risikerer anbefalingene å bli generelle prinsipper som hvert fagmiljø tolker forskjellig, eller ignorerer.

Gapet er ikke teoretisk. UHRs innspill (30.10.2025) viser til at omtrent 80 % av studentene bruker KI i studiene, mens nesten 70 % oppgir at de i ingen eller liten grad har fått opplæring i bruk av KI-verktøy. Studentene bruker altså verktøyet uten å lære å bruke det godt.

UH-didaktikk-nettverket peker på samme gap, men forblir på prinsippnivå. NSO peker eksplisitt på at utvalget mangler tiltak rettet mot studentens egen KI-bruk. Mitt innspill prøver å fylle dette gapet med en konkret modell.

## 5. Konkret modell: Fagspesifikke emner med felles malstruktur

Utvalget bør anbefale at fagrelevante studieprogrammer **etablerer pilot av** et emne i kritisk og effektiv KI-bruk, med felles pedagogisk struktur men fagspesifikt innhold, og skalerer ved bekreftet effekt fra pre/post-test. Hvert studieprogram avgjør selv om emnet skal være obligatorisk for sine studenter eller valgfritt. Anbefalingen formuleres slik at den respekterer institusjonelt selvstyre og akademisk frihet (jf. mandatet): fagmiljøene selv beslutter om og hvordan emnet skal inngå i eget studieprogram, og hvilket fagspesifikt innhold som er relevant. Anbefalt studiepoengvekt: **7,5 stp** som utgangspunkt. Det holder emnet under terskelen for to-sensor-kravet i uhl. § 11-7 (15 stp eller mer) og reduserer sensurkostnad i pilotfasen uten å gå på akkord med faglig dybde. Eksempler på emnetitler:

- **Arkitektur-design med praktisk KI-bruk**
- **Juss med praktisk KI-bruk**
- **Etisk hacking med praktisk KI-bruk**
- **Forretningsutvikling med praktisk KI-bruk**
- **Drøfte samfunnsproblemer med praktisk KI-bruk**
- **Utvikling med praktisk KI-bruk**

Felles for emnene: oppgavene er for komplekse og nyanserte til å løses ved kort prompting. Studenten må bryte ned problemet, utforske med KI som læringspartner, evaluere kritisk, og syntetisere et helhetsforslag.

### 5.1 Pedagogisk struktur

Emnet kombinerer fem elementer:

- **5.1.1 Forelesninger om metoder for effektiv KI-bruk.** Sentrale temaer: (1) **promptdesign** med evidensbaserte teknikker (chain-of-thought, tree-of-thoughts, few-shot, self-consistency, ReAct); (2) **iterativ samhandling** over flere runder med korreksjon og presisering; (3) **verifikasjonsteknikker** mot primærkilder, kjøring av kode og eksperimentell testing; (4)

**dialog-divergens** for å avsløre når modellen er usikker; (5) **modellforståelse på brukernivå**: hvorfor modeller hallusinerer, hva de er gode og dårlige på, og hvordan oppdage at en modell er blitt dårligere så man kan bytte leverandør; (6) **komparativ modellevaluering** på flere modeller (åpne/lukkede, eldre/nyere), som gir både arbeidslivskompetanse og empirisk grunnlag for å skille modellforbedring fra brukerlæring; (7) **strategier mot metakognitiv latskap**: forstå, utfordre, formulere egne meninger og bruke KI til å utfordre dem; (8) **etikk og opphavsrett** (attribusjon, personvern, juridiske rammer). Pensumet må holdes oppdatert; feltet endrer seg raskt.

- **5.1.2 Ukentlige øvinger** med komplekse fagrelevante problemstillinger. Vurderes av studassistenter mot rubrikk. Læringseffekten kommer fra mengdetrening og rask, konsentrert tilbakemelding, i tråd med notatets vurderinger i kap. 3. Øvingene veksler mellom **individuelle øvinger** og **gruppeøvinger med felles innlevering og peer-review**, for å styrke det sosiale læringsmiljøet og motvirke individualiserende effekt av samtaleroboter (jf. Påstand IV og UH-didaktikk-nettverkets innspill om kollektiv kunnskapsutvikling).
  - **5.1.2.1 Øvingsformer.** *Individuelle øvinger* gjennomføres som en tidsavgrenset økt (f.eks. 4 timer) som studenten selv velger tidspunkt for innen ukas frist. *Gruppeøvinger* gjennomføres som en synkron tidsavgrenset blokk (f.eks. 4 timer) der hele gruppen jobber samtidig; studentene har hver sin LLM-dialog men koordinerer menneskelig (i samme rom, over video eller skriftlig) om hvordan oppgaven skal løses og syntetiseres. LLM-tilgangen er konstant gjennom hele emnet; det er tidsrammen rundt øvingen, ikke verktøytilgangen, som driver kalibreringen mot effektiv bruk. Tre fordeler med gruppemodellen: (i) hver student utvikler egen prompt-disiplin og KI-bruk-ferdighet gjennom egen dialog, ikke som passiv tilskuer, noe som er nødvendig forberedelse til den individuelle eksamen og pre/post-testen; (ii) øvingsplattformen (Inspera, Wiseflow, Canvas) trenger bare håndtere autentisering og tidsavgrensning, ikke dyp LLM-integrasjon; (iii) koordinasjonskostnaden i synkrone blokker er en del av treningen for reelle arbeidssituasjoner.
  - **5.1.2.2 Refleksjonsnotat per student.** For å motvirke kollusjon og sikre at alle gruppemedlemmer faktisk deltar i fagarbeidet, leverer hver student et kort individuelt refleksjonsnotat (~1 side) etter at tidsrammen for selve øvingen er ute: hvilke deler av LLM-dialogen som var nyttige, hvor LLM-en bommet, og hva hen lærte. Notatet vurderes på sine egne premisser mot rubrikken.
  - **5.1.2.3 Dialog-lagring og personvern.** Studentens egen LLM-dialog kan lagres lokalt for studenten som læringsverktøy, men brukes ikke som grunnlag for sensurvedtak. Refleksjonsnotatet i 5.1.2.2 er studentens egen menneskelige syntese av dialogen og er det som vurderes. Det er studentens fagtekst som sensureres, ikke maskinens output. Rådialog som vurderingsgrunnlag ville utløst personvernforordningen art. 22 (automatisert behandling med rettsvirkning) og krevd DPIA, uten pedagogisk merverdi over refleksjonsnotatet.
  - **5.1.2.4 Rubrikkbasert sensur med KI-assistert utkast.** Sensur skjer mot detaljert rubrikk; rubrikken er det bindende vurderingsinstrumentet. For å øke konsistens og redusere ventetid kan studassistenten bruke institusjonsmodellen til å generere et *utkast* til poenggivning og tilbakemelding mot rubrikken. Studassistenten verifiserer utkastet mot rubrikken, korrigerer der det trengs, og signerer av med egen faglig vurdering. Det er sensuranbefalingen som er KI-assistert, ikke sensurvedtaket. Rubrikken anvendes raskere, ikke erstattes. NTNUs gjeldende retningslinje (rev. 30.09.2025) åpner eksplisitt for KI som støtte ved sensur og kvalitetssikring; den juridiske rammen utdypes i 5.7.1.
  - **5.1.2.5 Karaktervekting og sensur.** Gruppeøvinger og individuelle øvinger gir samlet sett en delkomponent av karakteren der **fagmiljøet selv velger vekting**, mens **pre/post-test og eksamen alltid er individuelle**. Endelig sensur av eksamen gjøres av faglærer mot rubrikk; KI kan inngå som utkast og kalibreringshjelp, men sensors faglige avgjørelse skal være selvstendig og dokumenterbar (jf. 5.7.1).
- **5.1.3 Pre/post-test** med samme modell for å måle reell læringseffekt gjennom emnet (jf. 5.5).
- **5.1.4 Eksamen** som store syntese-oppgaver over fire timer, under tilsyn, med tilgang til samme KI-modell brukt gjennom semesteret. Oppgavevolum og tidsrammen tvinger studenten til å prioritere, parallellisere og bruke KI effektivt, som er selve kjernen i emnets «effektiv KI-bruk». Rubrikkbasert vurdering: rubrikken dekker flertallet av faglig forsvarlige løsningsretninger og kan også vise hva som gir trekk der en feil i ekte sammenheng ville vært kritisk. Rubrikken utvikles og kvalitetssikres av fagfeltseksperter. En sentral forutsetning er at rubrikkforfatterne aktivt jakter på løsningsretninger de selv ikke har tenkt på, slik at en student som finner en faglig forsvarlig sti utenfor rubrikken ikke straffes urettmessig.
- **5.1.5 Designgrep for oppgavene.** To designgrep er sentrale:
  - **5.1.5.1 Oppgaver med ufullstendig eller internt motstridende informasjon.** I virkeligheten kommer problemstillinger sjelden med korrekte premisser. Kunder beskriver feil krav, vitner forteller ulike historier, måledata er feilkalibrert. Studenten skal lære å være kritisk til selve oppgaveteksten, ikke bare til KIs svar. Også den som har gitt oppgaven er en kilde studenten må vurdere. Rubrikken skal eksplisitt belønne at studenten identifiserer og drøfter premissfeilen.
  - **5.1.5.2 Oppgavekalibrering mot direkte LLM-besvarelse.** Oppgavene kalibreres slik at en direkte LLM-besvarelse (oppgaven gitt rått til KI-en uten kritisk bearbeidelse) gir få eller ingen poeng. Kalibreringen valideres ved å kjøre direkte LLM-besvarelse mot rubrikken *før* eksamen settes. Det tvinger studenten til å engasjere seg kritisk og bidra med

egen syntese for å få uttelling, og sikrer at karakterfordelingen drives av kvaliteten i studentens egen tenkning, ikke av modellytelse.

Eksempel fra juss: et etterforskningscase der flere kilder gir motstridende fremstillinger, oppgaveteksten inneholder upresise eller misvisende formuleringer, og påtalemyndighet (oppgavegiver i casen) er én av flere kilder studenten må veie kritisk.

Tilsvarende fagspesifikt: en samfunnsvariant utforsker hvilke perspektiver et politikkforslag treffer og misser; en arkitektur-design-variant gir krav som internt motsier hverandre slik at studenten må identifisere konflikten før hen kan løse den skikkelig. En utviklingsvariant kan gi i oppgave å implementere et API der spesifikasjonen er ufullstendig, internt motstridende eller har skjulte ytelses- og sikkerhetshensyn studenten må avdekke og adressere.

### 5.2 Norsk fagspråk

Mandatet inkluderer «tilegnelse av norsk fagspråk» som et hensyn. Språkrådet (innspill 18.11.2025) og Senter for norsk fagspråk i bruk / Norbruk-UiO (innspill 12.03.2026) advarer om at språkmodeller uten finetuning på norsk fagstoff gir upresist fagspråk. Det er en korrekt beskrivelse av dagens tilstand, men det leder ikke til at norske fagmodeller bør være standardvalget.

- **5.2.1 Modellvalg: fagretningsspesifikt.** Valget av primærmodell bør gjøres per fagretning, ikke som generelt prinsipp. Ledende internasjonale open-weight-modeller (f.eks. GLM- og Llama-familiene) håndterer norsk akseptabelt via system prompt og er for de fleste fagspesifikke oppgaver overlegne dedikerte nordiske initiativer på faglig kapasitet. **For fagretninger der KI-kapasitet er den knappe ressursen, gir det dårlige avveininger å velge en svakere modell utelukkende for bedre norsk-evne**, fordi det reduserer det studenten faktisk skal lære. Nordiske initiativ (NorLLM, EuroLLM, Sikt-koordinerte arbeider) bør følges aktivt og prioriteres som standardmodell der de er konkurransedyktige på faglig oppgaveløsning.
- **5.2.2 System prompt.** Systemprompten skal kreve at modellen svarer på bokmål eller nynorsk etter studentens valg, og bruker norsk fagterminologi der den finnes. På de sterkeste open-weight-modellene er dette enkelt å implementere og dekker en stor del av norsk-behovet.
- **5.2.3 Praktisk fagvalg.** For fagretninger der undervisning og litteratur allerede i hovedsak er på engelsk (store deler av teknologi, naturvitenskap, økonomi og anvendte fag), er det lite å miste på å holde emnet på engelsk eller på sterkeste modell med norsk system prompt. For fag der norsk fagspråk er kjerne i kompetansen (juss, lærerutdanning, klinisk helse), kompenseres et eventuelt språklig gap i den faglig sterke modellen med fagspesifikk språklig kvalitetssikring i øvinger og vurdering, ikke ved å nedgradere modellens faglige styrke.
- **5.2.4 Vurdering.** Rubrikker kan premiere bruk av korrekt norsk fagterminologi i sluttsyntesen der det er pedagogisk meningsfullt. Det gjør fagspråk til en aktiv læringskomponent, ikke bare en ramme.

Konklusjonen er pragmatisk: norsk fagspråk er et reelt hensyn, men ikke et hensyn som forsvarer at studenten må jobbe på en faglig svakere modell enn det som ellers er beste tilgjengelige verktøy.

### 5.3 Sertifiseringsfunksjonen

En nær innvending mot modellen er at den svekker universitetenes sertifiseringsfunksjon (jf. notatets kap. 1.1): et vitnemål skal si noe pålitelig om hva kandidaten kan, ikke hva modellen kan. Innvendingen forutsetter at sertifisering må skje uten substansielle bidrag fra eksterne kilder eller verktøy. Det er ikke hvordan høyere utdanning faktisk fungerer i dag. Flere veletablerte vurderingsformer tester nettopp studentens evne til å drive en faglig prosess med eksterne bidrag inn:

- **5.3.1 Bachelor- og masteroppgaver** skrives med veileder, fagfellelitteratur, peer review og fri biblioteksbruk. Det som vurderes er studentens evne til å drive prosessen (formulere spørsmål, velge metode, integrere kilder, syntetisere), ikke evnen til å produsere hver setning alene.
- **5.3.2 Klinisk vurdering** i medisin, psykologi og helsefag gjennomføres med pasient som «input», med tilgang til retningslinjer, oppslagsverk og kolleger. Det som testes er studentens diagnostiske og kliniske prosess.
- **5.3.3 Ingeniør- og bachelor-/masterprosjekter** bruker ofte ekstern fagrådgivning, industri-veiledere, simuleringsverktøy og kodebibliotek. Kvalifikasjonen handler om studentens prosjektledelse og integrasjon, ikke om at hver komponent er bygd fra bunnen.

Det disse har felles er at sertifiseringen tester studenten som **aktiv driver av en prosess**, ikke som isolert produsent. Den studenten som ikke kan drive prosessen (som ikke vet hva som skal spørres om, ikke kan vurdere svaret, ikke kan integrere på tvers, ikke kan syntetisere) blir avslørt fordi prosessen krever løpende beslutninger som verktøyene ikke kan ta for hen.

Det som sertifiseres med det foreslåtte emnet er på samme måte: **kandidaten kan drive en kompleks faglig prosess der KI er ett av flere tilgjengelige verktøy.** Det er en reell, dokumenterbar kompetanse i seg selv, og en kompetanse som arbeidslivet etterspør. Kalibreringen av oppgavene (direkte LLM-besvarelse gir få poeng, jf. 5.1.5.2) er den tekniske garantien for at det er studentens drivkraft som måles, ikke modellens output.

Den vesentlige variasjonen mellom LLM-er er hvor godt modellen løser komplekse, fagspesifikke oppgaver. Selve brukermønsteret (promptdesign, verifikasjon, kritisk evaluering, iterativ dialog) er stabilt på tvers av modeller. Studenten som

har lært disse prinsippene på institusjonsmodellen overfører ferdigheten til en hvilken som helst tilsvarende modell i arbeidslivet. Selv for studenter som senere jobber uten LLM-tilgang er det kritiske bidraget (den disiplinerte kritiske tenkningen rundt informasjon fra en uperfekt kilde) direkte overførbar.

Empirisk evaluering hviler primært på pre/post-test innen hver kohort (jf. 5.5). Det måler om studenten faktisk blir bedre på effektiv KI-bruk gjennom semesteret. Langtidsoppfølging av kandidater (kandidatundersøkelser, arbeidsgivervurdering) er ønskelig, men ikke nødvendig for å beslutte om pilot skal skalere.

### 5.4 Læringsutbyttebeskrivelser (forslag)

Etter fullført emne skal studenten kunne:

- **5.4.1 Anvende språkmodeller** som læringspartner i komplekse, fagrelevante problemstillinger.
- **5.4.2 Vurdere språkmodellers styrker og begrensninger** innen fagets domene.
- **5.4.3 Verifisere KI-genererte påstander** mot primærkilder, eksperimentell testing og kjøring av kode.
- **5.4.4 Formulere problemstillinger presist** nok til å få faglig nyttig respons.
- **5.4.5 Identifisere feil og motstridende informasjon i selve problemstillingen**, ikke bare i KIs svar.
- **5.4.6 Detektere feil**, både i KI-output og i egne resonnementer. Dette er anvendt kritisk tenkning i KI-tiden.
- **5.4.7 Holde egen tenkning aktiv** gjennom hele arbeidsprosessen, fordi besvarelsens kvalitet drives av studentens egen kritiske bidrag; ukritisk KI-bruk gir dårligere resultat.

Kjernen i læringsutbyttet er at studenten får direkte hands-on erfaring med bruk av KI i komplekse problemløsningssituasjoner, slik at hen lærer rammene for hva språkmodeller faktisk er gode og dårlige på.

### 5.5 Pre/post-test for målbar læringseffekt

Ved semesterstart løser alle studentene individuelt en standardisert kompleks problemstilling med KI. Testen gjentas rett før eksamen. **Begge er individuelle, men inngår ikke i karakteren**: studenten ser sin egen score og kan måle egen framgang. Sammenligningen viser om studenten faktisk har blitt bedre på effektiv KI-bruk og om emnet har målbar effekt.

**Kritisk forutsetning: nøyaktig samme KI-modell brukes i pre-test, gjennom hele semesteret, i post-test og på eksamen.** Uten dette kan modellens egen utvikling forveksles med studentens læring. Pre/post-data fra flere kohorter gir empirisk grunnlag for å beslutte om emnet skal skaleres, justeres eller skrotes.

**Modell-versjonering mellom kohorter:** Ved hvert semesterstart vurderer det fagansvarlige miljøet om en sterkere open-weight-modell er tilgjengelig. Hvis ja, den legges til som standardmodell for kommende kohort. Pre/post-data er da ikke direkte sammenlignbare *på tvers av* kohorter med ulik modell, men er fortsatt valide *innen* hver kohort. Tidligere modellversjoner beholdes parallelt så lenge de er nyttige. En eldre modell kan vise seg å være bedre på enkelte fagspesifikke oppgaver. Komparativ bruk av nyere mot eldre modell under øvinger gir studenten direkte erfaring med modellevaluering; eksamen kjøres alltid på én definert primærmodell.

Hardware-implikasjonen er moderat: 8×H100 (640 GB VRAM) håndterer én primær 70B-modell og én eldre 70B-modell parallelt med kvantisering (FP8/INT4) og dynamisk batch-styring, eller én 70B pluss mindre modeller hvis fagmiljøet trenger det.

### 5.6 Tilgang under eksamen: åpent designspørsmål

Et reelt designspørsmål: hvilken tilgang skal KI-en ha utenfor seg selv under eksamen? To hovedalternativer:

- **5.6.1 Begrenset internett-tilgang via whitelist** (offisielle dokumenter, fagbibliotek, etablerte oppslagsverk). Mer realistisk i forhold til arbeidslivet, men teknisk og sikkerhetsmessig komplisert. Krever hardnet eksamensnettverk på eksamen-PCene for å hindre kommunikasjon mellom studenter eller eksterne hjelpere. Siden LLM-en uansett hostes lokalt, er dette en standardøvelse i eksamenssikkerhet.
- **5.6.2 Lokal, fagspesifikk RAG-løsning** der KI får tilgang til relevante filer (forelesningsnotater, regelverk, standarder, fagets eget oppdaterte materiale) gjennom verktøykall som åpner og søker i filene. Modellen kan eksplisitt åpne og lese hele filer eller utdrag, søke med nøkkelord eller fritt søk, og dermed gi gjennomsiktige, sporbare resultater. For mange fag er pretraining alene tilstrekkelig og en dokumentsamling gir lite ekstra; verdien ligger først og fremst i fagspesifikt materiale som ikke er godt dekket av modellens pretrening, eksempelvis Lovdata for juss, NORSOK-standarder for petroleumsfag, kliniske retningslinjer for medisin, eller nylig oppdatert fagstoff etter modellens treningskutt.

Utvalget bør anbefale at institusjonene utforsker begge alternativer der det er relevant, og at det utvikles felles standarder for sikker eksamensinfrastruktur i samarbeid mellom institusjoner. Bruk av opphavsrettslig beskyttet faglitteratur i RAG-løsningen forutsetter avklaring av Kopinor-avtalen (jf. juridisk avsnitt under).

### 5.7 Juridisk forenlighet

**EUs AI Act:** Studentens egen bruk av KI under tilsynseksamen er ikke høyrisiko-bruk. Selve sensurvedtaket gjøres av fagpersoner mot ekspert-utviklet rubrikk og forblir et menneskelig ansvar. Tre forhold må likevel adresseres før pilot:

- **5.7.1 Underviserens og studassistentens KI-bruk i vurderingsarbeid.** Annex III pkt. 3(b) klassifiserer som høyrisiko «AI systems intended to be used for evaluating learning outcomes». Art. 6(3)(d) unntar systemer som bare utfører «preparatory tasks to an assessment», og unntaket gjelder eksplisitt **ikke** summativ vurdering. UiT KI-teamets innspill (10.04.2026) peker på at rettstilstanden ikke er klart avklart for en modell der KI foreslår sensur og et menneske kvalitetssikrer. Det er ikke uten videre gitt at slik praksis faller utenfor høyrisiko-kategorien. NTNUs retningslinje (rev. 30.09.2025) åpner samtidig for at KI brukes som støtte i sensurarbeid (utkast, kvalitetssikring, kalibrering) så lenge sensor selv står for den faglige avgjørelsen, og åpner for tidsavgrenset utprøving med automatisert KI-sensur. Anbefalt praksis for emnet er derfor:
  - **5.7.1.1** Studassistenter kan bruke KI fritt som hjelpemiddel ved retting av ukentlige øvinger (formativ kontekst, klart innenfor unntaket i Art. 6(3)(d)).
  - **5.7.1.2** Ved summativ sensur av eksamen kan KI brukes som utkast og kalibreringshjelp, men sensors faglige avgjørelse må være selvstendig og dokumenterbar, og det må kunne vises at avgjørelsen ikke er rent automatisk videreført fra KI-utkastet.
  - **5.7.1.3** Studenten informeres om at KI inngår i vurderingsstøtten, og institusjonen gjennomfører DPIA. Hvis tilsynsmyndighet senere lander på en strengere tolkning i tråd med UiTs varsel, må pilotinstitusjonene kunne falle tilbake på rent menneskelig summativ sensur uten å miste pedagogisk substans i emnet.
- **5.7.2 Institusjonsmodellen som eksameninfrastruktur.** En bredere tolkning av Annex III kan trekke selve institusjonsmodellen (når den eksponeres som verktøy under summativ eksamen) inn under høyrisikoklassen. Det bør avklares med juridisk fagmiljø før første pilot, og veiledning fra Datatilsynet/Digitaliseringsdirektoratet bør innhentes.
- **5.7.3 Opphavsrett til faglitteratur i lokal modell.** UH-didaktikk-nettverkets innspill (17.02.2026, kap. 4) peker på at gjeldende Kopinor-avtale for UH-sektor ikke åpner for at forskningsartikler og faglitteratur lastes opp og bearbeides i interne KI-løsninger. Det betyr at fagspesifikke RAG-løsninger under eksamen ikke kan implementeres lovlig i dag for de fleste fagbøker og artikler. Bare CC-lisensiert materiale kan trygt brukes.

**Handlingspunkt for utvalget:** NOU-en bes om eksplisitt å etterspørre veiledning fra Datatilsynet, Digitaliseringsdirektoratet og Kopinor om punktene 5.7.1–5.7.3, slik at pilotinstitusjonene starter på avklart juridisk grunn. Dette er den enkleste juridiske avklaringen utvalget kan be om. Alternativene er å pilotere på usikker grunn eller å vente til hver enkelt institusjon avklarer selv.

**Norsk regelverk:** Universitets- og høyskoleloven og statlig forskrift har ingen begrensninger som hindrer denne modellen. Hjelpemidler defineres på emnenivå. NTNUs gjeldende retningslinje (rev. 30.09.2025) angir at det ved NTNU er tillatt å bruke KI som hjelpemiddel ved utarbeidelse av studentoppgaver, men at studenten selv står for besvarelsen og må redegjøre for bruken. Emneansvarlig kan eksplisitt angi mer eller mindre tillatelse i hvert enkelt emne. Det betyr at emnespesifikke unntak der KI-bruk inngår i læringsmålet og vurderingsformen kan implementeres innenfor eksisterende rammeverk. Tilsvarende retningslinjer ved andre institusjoner (UiO, UiB, OsloMet, Nord, Innlandet) bør verifiseres før pilot starter, og bør om nødvendig oppdateres til samme nivå av fleksibilitet som NTNU har. Statlig forskriftsendring trengs ikke.

## 6. Forskningsgrunnlag

Forslaget bygger på etablert pedagogisk forskning, mye av den allerede sitert i utvalgets eget notat:

- **6.1 Metakognitiv latskap og kognitiv avlastning:** Fan et al. (2025), Handa, Bent et al. (2025), Gerlich (2025), alle i notatets referanseliste. Disse studiene viser at ukritisk KI-bruk gir bedre prestasjoner på enkeltoppgaver, men ikke nødvendigvis bedre læring. Mitt forslag adresserer dette direkte gjennom oppgavedesign og rubrikk.
- **6.2 Formative tilbakemeldinger og produktiv strid:** Engeness & Gamlem (2025), Hardman (2025), Kestin et al. (2025), også i notatet. Disse støtter design med ukentlige øvinger og rask tilbakemelding.
- **6.3 «Desirable difficulties»** (Bjork & Bjork). Det klassiske pedagogiske prinsippet om at det er den mentale anstrengelsen, ikke effektiviteten, som gir varig læring. KI senker som regel innsatsnivået; oppgavekalibreringen (jf. 5.1.5.2) *gjenoppretter* den ønskede vanskelighetsgraden ved å kreve studentens egen syntese for poeng. Studieutvalget på matematisk-naturvitenskapelig fakultet ved UiO viser i sitt innspill (19.11.2025) til samme idé under navnet «den produktive kampen», og argumenterer for at studentene må lære å løse problemer både med og uten KI for å sikre dybdelæring.
- **6.4 Mollick & Mollick (2024)** *«Instructors as Innovators: a Future-focused Approach to New AI Learning Opportunities, With Prompts»* (Wharton School Research Paper, SSRN: 4802463). Gir konkrete pedagogiske strategier for KI i undervisning, og er et naturlig referansepunkt for det foreslåtte emnet.
- **6.5 Chen, Zou & Zaharia (2023)** *«How Is ChatGPT's Behavior Changing over Time?»* dokumenterer empirisk drift i kommersielle modeller og underbygger kravet om institusjonsmodell med versjonslås i pre/post-design.

Listen er ikke uttømmende, men illustrerer at det finnes solid forskningsforankring for designet.

## 7. Infrastruktur og innføringsrekkefølge

For at modellen skal være praktisk gjennomførbar, og for at pre/post-testen skal være meningsfull, må institusjonene løse modell-stabilitet og likhet.

### 7.1 Innføringsstegene

- **7.1.1 Utvikle pensum, rubrikker, øvingsoppgaver og pre/post-test** forankret i den nyeste forskningen på effektiv KI-bruk (jf. kap. 6). Pensumet må være dynamisk og oppdateres mellom semestre fordi feltet endrer seg raskt. Strukturert studentfeedback hvert semester bør være en del av oppdateringsprosessen. Sentral utvikling deles på tvers av institusjoner via HKdir, lokal tilpasning gjøres av fagmiljøet selv.
- **7.1.2 Rekruttere undervisere og studassistenter** med praktisk erfaring i kritisk KI-bruk.
- **7.1.3 Etablere stabil modell-infrastruktur lokalt.** Institusjoner (f.eks. NTNU) hoster en åpen modell, f.eks. Llama, GLM eller tilsvarende konkurransedyktig open-weight-modell, på egen compute, med tilgangskontroll og datasikkerhet. **Hver institusjon velger sin egen primærmodell**, noe som skaper konstruktiv konkurranse om hvem som setter opp den best fungerende KI-infrastrukturen. Pre/post-data forblir valide innen hver institusjon, selv om absolutte scores ikke kan sammenlignes direkte på tvers. Når institusjonsmodellen er konkurransedyktig, gir studentens egen tilgang til private modeller liten ekstra fordel. Mye av denne infrastrukturen finnes allerede: NTNUs IDUN-kluster har 56 NVIDIA H100-GPUer (samt 104 A100, 40 V100), og Sigma2-konsortiets Olivia-system (satt i drift juni 2025) har 448 NVIDIA GH200 Grace-Hopper-superchips delt mellom UiB, UiO, UiT og NTNU. Pilotfasen kan derfor kjøres uten ny kapitalinvestering i hardware. Anskaffelse og drift kan koordineres via Sikt eller BOTT-samarbeidet, jf. HK-dirs anbefaling (19.11.2025) om felles sektor-anskaffelser av KI-løsninger. Det vil samtidig svare på Studenttinget NTNUs bekymring (08.12.2025) om at KI ikke må bli et privilegium for studenter som har råd til egne abonnementer. Institusjonen må også ha **redundant kapasitet og dokumentert prosedyre for modellbytte**, slik at hardware-feil ikke får semesteret til å kollapse.
- **7.1.4 Pilotfase: start med ett emne som proof-of-concept.** Institusjon og fagmiljø velger selv. En liten oppstart gjør det mulig å oppdage designsvakheter før de skaleres, og holder ressursbruken proporsjonal med usikkerheten i tidlig fase. Studentene får fri tilgang til institusjonsmodellen via institusjons-API. Utvidelse til flere emner og institusjoner skjer i takt med dokumentert pedagogisk effekt fra pre/post-data og kandidatoppfølging, ikke som forutsatt nasjonal utrulling.
- **7.1.5 På sikt:** integrere LLM-tilgang i institusjonens øvings- og eksamenssystem (f.eks. Inginious for programmeringsøvinger, Inspera eller Wiseflow for eksamen; Inspera dekker NTNU, UiO, UiB, OsloMet og flere, og Wiseflow brukes ved Innlandet og Nord). Slik tett integrasjon gjør det enklere å verifisere at studenten har brukt institusjonsmodellen og ikke eksterne verktøy under tilsynseksamen, og gir aggregert anonymisert grunnlag for forskning på pedagogisk effekt. Individuelle dialoger inngår ikke i sensurvedtak (jf. 5.1.2.3). Dette er ikke realistisk i første pilotfase, men bør være målbilde.

### 7.2 Fallback hvis modellen ikke er tilgjengelig

Selv med redundant kapasitet kan det oppstå tilfeller der institusjonsmodellen ikke er tilgjengelig under en konkret øving eller eksamen. Prinsippet bør være at studenten **ikke skal tape studiepoeng eller godkjenninger på grunn av infrastrukturfeil**:

- **7.2.1** Hvis modellen er nede under en øving, godkjennes øvingen automatisk på grunnlag av oppmøte eller alternativ levering.
- **7.2.2** Hvis modellen blir utilgjengelig under eksamen, avbrytes eksamen og det settes opp utsatt eksamen så snart infrastrukturen er gjenopprettet. Studenten skal ikke tape studiepoeng eller forsinkelse på grunn av driftsfeil. Institusjonen må ha eksplisitt dokumentert prosedyre for utsatt eksamen, slik at avbrudd ikke kompromitterer sertifiseringsverdien.

### 7.3 Grovt kostnadsanslag

**Kostnaden ved å la være:** Norge taper mer på å utdanne studenter som får mindre og mindre nytte av utdanningen sin i en KI-preget arbeidsdag. Personlige prosjekter og tidligere jobberfaring veier allerede tyngre enn studiepoeng i mange ansettelsesprosesser. Hvis universitetene ikke integrerer KI som verktøy i studiene, blir det stadig mindre relevant å studere for arbeidslivet. En målrettet investering i et fagspesifikt emne i kritisk og effektiv KI-bruk er en investering i utdanningssektorens fortsatte relevans, ikke en marginalkostnad over en eksisterende kjerne.

Kostnadene deles i to klart adskilte kategorier: **sektorinvesteringer** (engangskostnader som deles på tvers av institusjoner og emner) og **per-emne-kostnader** (kostnader hvert enkelt emne med 100–150 studenter genererer). De skal ikke amortiseres sammen.

**Sektorinvesteringer (engang, deles på tvers av institusjoner):**

| Post | Beløp | Beskrivelse |
|---|---|---|

| | | |
|---|---|---|
| Pensumutvikling, rubrikker, veiledningsmateriell | ~25 MNOK over 2 år | Kan dekke 8–12 fagretninger; koordineres via HKdir |
| Infrastruktur-engineering per institusjon | 5–10 MNOK | vLLM-deployment, autentisering, logging, Inspera/Wiseflow-integrasjon |
| Juridisk avklaring, retningslinje-revisjon, Datatilsynet-veiledning | 1–2 MNOK | Fellesnivå for sektoren |

**Per-emne-kostnader (100–150 studenter):**

| Fase | Kostnad per emne per år | Hovedposter |
|---|---|---|
| **Pilotfase (år 1–2)** | 3–6 MNOK første år, 2–4 MNOK år 2 | Studassistent (1–4 årsverk), lokal pensumtilpasning, ekstra oppstartsbyrde |
| **Etablert fase (år 3+)** | 1,5–3 MNOK/år | Studassistent dominerer, drift på eksisterende infrastruktur, lokal vedlikehold |
| **Modne fase (år 5+)** | 1,5–2,5 MNOK/år | Effektivisert drift, studassistent fortsatt hovedpost |

Til sammenligning har et regulært emne med 100–150 studenter ved NTNU typisk 1–1,5 MNOK/år i drift, dominert av 1–2 årsverk studassistent. Et KI-emne av denne typen er altså i etablert fase **1,5–2× dyrere ved typiske verdier (opptil 3× ved høyeste estimat)**, og i pilotfasen 2–4× dyrere. Forskjellen drives av at synteseoppgaver mot rubrikk krever mer studassistent-tid enn flervalgs- eller standardøvinger, og av oppstartsbyrden ved nytt fagområde.

**Detaljerte poster per emne:**

- **7.3.1 Studassistent-ressurser:** ~1–4 årsverk per emne avhengig av øvingsfrekvens. Dominerende driftspost. KI som hjelpemiddel i selve rettearbeidet (jf. 5.1.2.4) reduserer time-per-besvarelse, men studassistent må fortsatt verifisere mot rubrikk og stå faglig for vurderingen.
- **7.3.2 Compute:** marginal kostnad i pilotfasen. Eksisterende NTNU IDUN-infrastruktur (56 H100) og Sigma2 Olivia (448 GH200) holder. Realistisk yteevne på 8×H100 med vLLM og PagedAttention er 100–200 samtidige brukere på en 70B-modell og 200–300 samtidige brukere på en 30B-modell, med ~20–40 tokens/sek per stream ved typisk eksamensbelastning, på linje med kommersielle API-er. Hvis sektoren skalerer ut over eksisterende kapasitet, koster en ny 8×H100-node ~2,4 MNOK i GPU-bundle (ca. 230 000 USD per mai 2026) pluss ~1–2 MNOK i chassis/kjøling/nettverk.
- **7.3.3 Lokal pensumtilpasning og fagansvarlig** dekkes innenfor faglærers normale undervisningsplikt, men kommer med oppstartsbyrde i pilotfasen som er reflektert i de høyere år 1-tallene.

### 7.4 Kompetanse hos undervisere: et høne-og-egg-problem som må løses ved å starte

En vanlig innvending mot et nytt emne av denne typen er at det finnes for få ansatte med ekspertkompetanse på kritisk og effektiv KI-bruk. Det er riktig som beskrivelse, men feil som konklusjon. Vi får ikke flere eksperter ved å vente: feltet modnes når institusjonene faktisk underviser i det, samler erfaring, og videreutvikler pensumet iterativt.

Forutsetningen for at en oppstart skal fungere uten et fullt modent fagmiljø er at **pensumet og rubrikkene bærer kvaliteten**, ikke at hver enkelt underviser er ekspert. Med tett rubrikkbeskrivelse, gode eksempelbesvarelser og systematisk erfaringsdeling mellom institusjoner kan studassistenter og undervisere uten dyp KI-erfaring likevel gi pålitelig vurdering tidlig.

Det viktigste kravet til underviser er ikke ekspertise på KI som verktøy, men solid faglig kompetanse i det feltet KI brukes på i emnet. En jusprofessor som leder *Juss med praktisk KI-bruk* trenger ikke å være ledende på prompt-design. Vedkommende trenger å være god på juss, så hen kan vurdere om studentens KI-assisterte resonnement holder faglig. KI er i denne sammenhengen bare en annen samtalepartner studenten bringer inn; det er studentens faglige bidrag, kritikk og syntese som vurderes. Det burde ikke være vanskelig for en faglig ekspert å vurdere.

En naturlig rekrutteringskilde er undervisere som allerede er aktivt engasjerte i KI og som ønsker å undervise i feltet. De finnes ved alle institusjoner, men har sjelden hatt et formelt mandat. Et formelt emne av denne typen gir det mandatet, samtidig som sentral pensumutvikling (jf. 7.1.1) reduserer den enkeltes oppstartsbyrde.

## 8. Pensumimplikasjon: T-formet kompetanse

Selv om dette i utgangspunktet er ett emne, peker det på en bredere implikasjon for pensum i fagene det forholder seg til:

- **8.1 Dybde:** grunnleggende prinsipper, fundamenter og teori som ikke endrer seg når KI-verktøyene utvikles videre.
- **8.2 Bredde:** spesifikke implementasjoner, syntaks og fakta som KI uansett har tilgang til, håndteres med KI som verktøy, ikke gjennom memorering.

Denne omfordelingen krever at fagene aktivt identifiserer hva som er prinsipielt fundament og hva som er teknisk overflate. Det er en faglig vurdering, og må gjøres av fagmiljøene selv.

## 9. Avgrensning

Modellen treffer fag der anvendelse og syntese er læringsmålet, der studentens evne til å drive en faglig prosess med flere informasjonskilder er kjerne i kompetansen. Det dekker ingeniørfag, IT, økonomi og administrasjon, juss (med ekstra disiplin for kildeverifikasjon), samfunnsvitenskap, store deler av medisin, og tverrfaglige problemstillinger som samfunnsutfordringer og policy-design.

Modellen krever betydelig tilpasning for fag der læringskjernen ligger i:

- **9.1** Den personlige kampen med primærteksten (humaniora, filosofi, litteratur).
- **9.2** Mønstergjenkjenning bygd på mengdetrening (klinisk medisinsk vurdering).
- **9.3** Innsiktsmomentet i seg selv (matematiske bevis).

For disse fagene trengs andre tilnærminger, som ligger utenfor dette innspillets ambisjon.

## 10. Om generaliserbarhet

En berettiget innvending er om modellen fungerer for gjennomsnittsstudenten eller bare for høy-motiverte studenter. Min vurdering: dagens karaktersystem belønner i stor grad evnen til å gjenfortelle pensum og underbelønner selvstendig kritisk tenkning. Før internettet var det mye viktigere at man husket ting, derfor belønnet man det. I dag trenger vi ikke å memorere like mye. Det vi trenger er kritisk tenkning som kan anvende tilgjengelig informasjon fra hele verden. Studiene blir mindre relevante slik de er nå, og studenter har derfor ikke fått trening i akkurat den ferdigheten som KI gjør mer tilgjengelig, og som arbeidsmarkedet utlyser. Modellen kan bygges med gradvis økende kompleksitet og formative tilbakemeldinger underveis, slik at overgangen for studenter som ikke har trent denne ferdigheten tidligere, blir håndterbar.

Pre/post-testen gir empirisk grunnlag for å besvare innvendingen over tid: data fra de første kohortene vil vise om læringseffekten varierer med studentprofil, og om designet bør justeres for å treffe bredere.

## 11. Avslutning

Utvalgets foreløpige anbefalinger er gode, men forsvarsorienterte. For at norske universiteter skal lykkes med formålet om samfunns- og arbeidslivsrelevans (notatets kap. 1.1) i en KI-preget tid, bør NOU-en også inneholde konkrete operasjonaliseringer av Påstand III og Anbefaling 4, ikke bare prinsipper.

Jeg foreslår at NOU-en eksplisitt anbefaler institusjonene å etablere fagspesifikke emner i kritisk og effektiv KI-bruk, der hvert studieprogram selv avgjør om emnet skal være obligatorisk for sine studenter, med:

- **11.1** Felles pedagogisk struktur: forelesninger om metode, ukentlige øvinger med studassistent-vurdering (KI som hjelpemiddel mot rubrikk), pre/post-test med samme modell, rubrikkbasert eksamen under tilsyn.
- **11.2** Institusjonelt hostet, stabil KI-modell som infrastruktur.
- **11.3** Empirisk evaluering gjennom pre/post-data, slik at designet kan justeres over tid.

Det som skiller forslaget fra prinsipperklæringer er to designgrep som gjør modellen kalibrert mot KIs egen styrke og målbar i effekt: oppgaver kalibrert slik at direkte LLM-besvarelse gir få poeng, og pre/post-test som lar effekten verifiseres empirisk.

Forslaget bygger på samme premiss som UH-didaktikk-nettverkets innspill og svarer på det gapet NSO peker på. Det utfordrer ikke utvalgets retning, det operasjonaliserer den. **Gjør ukritisk KI-bruk selvutslettende.** Grunnprinsippet er enkelt: KI er kommet for å bli, og svaret er å gjøre oppgavene faglig vanskeligere, ikke å skjerme studiene fra et verktøy som er en del av yrkeslivet studentene utdannes til.

Jeg stiller meg gjerne til disposisjon for utvalget for utdypning eller spørsmål.

Med vennlig hilsen, Thobias Melfjord Knudsen Vinner av NM i AI 2026 B.Sc-student i informatikk, NTNU thobknu@gmail.com