

KURS 3 · ØKT 3

Hvorfor høres den sikker ut?

8. trinn

40 av 60 min

Øvelsestekster på papir, lærerdemo (valgfri)

Mål for økta: elevene kan forklare hvorfor en KI-chatbot kan gi ulike svar på samme spørsmål og høres sikker ut selv når den tar feil, og at det som er likt i flere svar, ikke dermed er riktig.

Kort fortalt: elevene leser tre svar på samme spørsmål om Galdhøpiggen, skrevet for å ligne svar fra en KI-chatbot. De finner hva som er likt og ulikt, hvilke ord som gjør svarene sikre, og sjekker tre påstander mot et kildekort fra Store norske leksikon. Elevene skal gå ut av timen med at det som står i alle svarene, ikke dermed er sant, og at en sikker tone ikke betyr at modellen vet.

Før timen

rundt 15 min

- Skriv ut tre svar, kildekort og svarark, ett sett per gruppe på tre: [tre-svar.pdf](#).
- Legg kildekortet (side 2) til side. Det deles ut først i trinn 3.
- Elevene trenger KI-loggen sin ([ki-logg.pdf](#) hvis noen trenger nytt ark).
- Les «Det du trenger å vite» under, og fasiten i fasiten for kurset (egen PDF på øktsiden).
- Valgfritt, lærerdemo: sjekk at skolens KI-tjeneste er godkjent av skoleeier, har databehandleravtale, ikke lagrer eller trener på ledetekster, og passer elevenes alder. For Oslo-skoler er Osloskolens chatbot et eksempel. Prøv spørsmålet på forhånd, med en ny samtale for hvert svar. Ikke skriv inn personopplysninger.
- Åpne presentasjonen: [okt-3.html](#). Trykk N for å se lærernotatene under lysbildet. PDF-en er reserve hvis nettleseren ikke virker.

Det du trenger å vite

Om tekstene: de tre svarene er øvelsestekster med innebygde feil. Vi har skrevet dem for å ligne tre svar fra en KI-chatbot på spørsmålet «Hvor høy er Galdhøpiggen, og hvem kom først til toppen?» Det står på arket. Si det til elevene.

Hva som stemmer: Galdhøpiggen er 2469 meter, Norges og Nord-Europas høyeste fjell, i Jotunheimen i Lom. Steinar Sulheim, Ingebrigt Flotten og Lars Arnesen gikk til toppen i 1850. Det er første kjente gang i nyere tid. En pil fra yngre jernalder er funnet på toppen, så noen kan ha vært der før. Stien fra Juvasshytta går over Styggebrean, med fører.

Feilene i svarene:

- **Storjuvbrean** står i alle tre svarene, men stien går over Styggebrean. Storjuvbrean finnes, men ligger et annet sted. En feil med et ekte navn er vanskeligere å oppdage.
- **Ola Storjuv** i svar 2 finnes ikke. Navnet er satt sammen av biter som passer, som i heftet fra økt 2.
- **2496 meter** i svar 3 er tallene i feil rekkefølge.
- **Sulheim gikk alene** i svar 3 er feil. De var tre.
- «**Alltid regnet som Norges høyeste**» er feil. Glittertinden ble lenge regnet som høyest når snøskavlen ble tatt med.
- «**De første**» står sikkert i alle svarene, men kilden sier «i nyare tid».

Likt i alle tre: høyeste fjell, Jotunheimen, 1850, Juvasshytta og Storjuvbrean. Fire av fem stemmer. Storjuvbrean er feil i alle tre.

Sikre ord: «de første», «den kjente fjellføreren», «de fleste», «alltid», «den vanligste ruten». Ingen av svarene sier at noe er usikkert.

Hvorfor KI høres sikker ut: modellen har lært av tekst skrevet i en sikker tone. Forskere som lager språkmodeller, skriver også at treningen har belønnet å gjette framfor å si «jeg vet ikke».

Om KI-bruk: økta fungerer fullt ut uten skjerm. Lærerdemoen er et tillegg. Elevene bruker ikke KI selv før det er avklart hvilken tjeneste elever på 8. trinn trygt kan bruke.

I kurs 1 til 3 prøver elevene først og får forklaringen etterpå. Derfor kommer det praktiske før faktadelen.

Slik går timen

Del	Lysbilde	Tid
Start: hva tror du?	1 og 2	3 til 5 min
Praktisk: likt og ulikt, sikre ord, sjekk	3 til 5	14 til 18 min
Lærerdemo (valgfri)	6	5 til 8 min
Fakta: hvorfor høres den sikker ut?	7 til 10	6 til 10 min
Diskusjon	11	6 til 10 min
KI-logg	12	3 til 5 min

Lysbilde 13 er kreditering, og trenger ikke vises i timen. Uten lærerdemo hopper du over lysbilde 6.

Start: hva tror du?

3 til 5 min

KJERNE

Lysbilde 1 og 2. Ble KI-loggen fra økt 2 flyttet, gjør elevene den nå.

Ellers deler du ut en lapp til hver elev. **Si:** «Hvis du stiller en KI-chatbot samme spørsmål tre ganger, får du samme svar? Skriv svaret og hvorfor.» Lappen legges i konvolutten.

1. Praktisk

14 til 18 min

KJERNE

Grupper på tre.

Lysbilde 3. Trinn 1: likt og ulikt (5 til 6 min). Del ut de tre svarene og svararket.

Si: «Dette er tre svar på samme spørsmål. De er ikke ekte KI-svar, men laget slik KI-svar ofte er, og de har feil i seg. Les dem høyt i gruppen. Skriv hva som står i alle tre, og hva som bare står i ett eller to.»

Lysbilde 4. Trinn 2: sikre ord (2 til 3 min).

Si: «Finn ord og uttrykk som får svarene til å høres sikre ut. Skriv dem på svararket.»

Lysbilde 5. Trinn 3: sjekk tre påstander (6 til 8 min). Del ut kildekortet.

Si: «Dette er fra Store norske leksikon, og det er på nynorsk. Velg tre påstander fra svarene. Minst én skal stå i alle tre. Finn påstanden på kildekortet, og skriv hva kilden sier. Stemmer det?»

Se etter, og spør:

- Grupper som tror at det som er likt i alle tre, må være riktig. Spør: «Hva sier kilden om breen?» (Styggebreen, ikke Storjuvbreen.)
- Grupper som velger det svaret som høres best ut. Svar 2 høres kunnskapsrikt ut. Spør: «Finner dere Ola Storjuv på kildekortet?» (Nei.)
- «De første». Spør: «Hva sier kilden om hvem som var først?» (Første gang i nyere tid. En pil fra jernalderen er funnet på toppen.)

Lærerdemo

5 til 8 min

UTDYPING

Lysbilde 6. Bare med skolens godkjente tjeneste, på storskjerm. Skriv samme spørsmål tre ganger, hver gang i en ny samtale: «Hvem kom først til toppen av Galdhøpiggen? Svar med tre setninger.» Ingen personopplysninger.

Si: «Se etter tre ting. Er svarene ordrett like? Hva står i alle tre, og stemmer det med kildekortet? Viser tjenesten kilder, og hva står i dem?»

Svarene vil variere fra gang til gang. Det er poenget. Ikke lov elevene at verktøyet tar feil. Svarer det riktig hver gang, spør du: «Hvordan vet vi det, uten å sjekke?»

2. Fakta og info

6 til 10 min **KJERNE**

Lysbilde 7. Si: «En KI-chatbot trekker ord med tilfeldighet, som terningen i økt 1 og 2. Derfor blir svarene ulike når du spør flere ganger.»

Lysbilde 8. Si: «Det som kommer igjen i flere svar, er det modellen oftest gjetter. Det kan være riktig. Det kan også være en feil den gjør hver gang, som Storjuvbreen.»

Lysbilde 9. Si: «Modellen har lært av tekst skrevet i en sikker tone. Forskere som lager språkmodeller, skriver også at treningen har belønnet å gjette. En gjetning gir av og til poeng. Jeg vet ikke gir aldri poeng.»

Lysbilde 10. Si: «Noen tjenester søker på nettet og viser kilder. Det hjelper, men da må du sjekke at kilden faktisk sier det svaret påstår. I kurs 5 lærer dere en fast rutine for det.»

3. Diskusjon

6 til 10 min

Lysbilde 11. Still ett spørsmål om gangen. La elevene tenke alene i et halvt minutt, snakke med den som sitter ved siden av i ett minutt, og si det til hele klassen etterpå.

- **Kjernespørsmål:** «Kan vi stole på det som er likt i alle tre svarene?» (Nei. Fire av fem ting stemte, men Storjuvbreen var feil i alle tre.)
- **Utdyping:** «Hvilket svar hørtes best ut før dere sjekket? Var det det beste?»
- **Utdyping:** «Hvorfor tror dere modellen fant på et navn som Ola Storjuv?» (Som heftet fra økt 2: biter som passer sammen, uten at noen har sjekket.)
- **Utdyping:** «Hva ville dere sjekket først hvis svaret handlet om noe viktig, for eksempel medisin eller en lekse?»

4. Skrivning og refleksjon

3 til 5 min **KJERNE**

Denne skrivingen avslutter kurset og bør ikke flyttes.

Lysbilde 12. Elevene ser på lappene fra alle tre øktene. Så skriver de i KI-loggen:

- Dagens spørsmål: «Hvordan lager en maskin tekst uten å vite noe?»
- «Før trodde jeg...»
- «Nå tenker jeg...»
- «Jeg lurte på...»

Samle inn KI-loggene, eller legg dem i konvoluttene.

Misforståelser å følge med på

- At et svar er riktig fordi det kommer igjen. Modellen kan gjøre samme feil hver gang.
- At en sikker tone betyr at modellen vet. Tonen kommer fra teksten den har lært av og fra treningen.
- At modellen slår opp i et leksikon. Uten nettsøk lager den teksten ut fra det den har lært. Med nettsøk kan den fortsatt gjengi kilden feil.
- At KI alltid tar feil. Mye av det som står i de tre svarene, er riktig. Poenget er at vi ikke vet hva som er riktig, før vi har sjekket.

Tilpasning

- **Kort tid:** dropp trinn 2, og sjekk bare påstanden om breen i trinn 3.
- **Mer tid:** gruppene skriver et fjerde svar på tre setninger, der hver påstand er støttet av kildekortet.
- **Uten lærerdemo:** økta er laget for å fungere uten. Bruk tiden på diskusjonen.

Kilder og kreditering

- Store norske leksikon, «Galdhøpiggen» (Lars Mæhlum og Per Roger Lauritzen, sist endret 14. september 2026): høyde, beliggenhet, stien over Styggebrean, Glittertinden, pila fra yngre jernalder og turen i 1850. Kildekortet siterer korte utdrag.
- Store norske leksikon, «språkmodell» (Sondre Wold, sist oppdatert 16. desember 2025): generert tekst virker ofte troverdig, men kan inneholde faktafeil.
- Kalai, Nachum, Vempala og Zhang, «Why Language Models Hallucinate» (OpenAI, 4. september 2025, arxiv.org/abs/2509.04664): trening og testing belønner gjetting framfor å innrømme usikkerhet. Kilden er skrevet av et selskap som lager språkmodeller.
- Tilfeldighet i ordvalget: Google Cloud, «Adjust parameter values» (Vertex AI-dokumentasjonen).
- NDLA, «Kan KI være en kilde?» (Marthe Johanne Moe, oppdatert 16. juni 2026): om KI-svar med kilder. Lenket, ikke kopiert.
- De tre svarene er øvelsestekster skrevet for KI på papir.