

Bruk av kunstig intelligens i begrunnelsesskriving

Sluttrapport

Denne rapporten presenterer en samlet evaluering av begrunnelsesprosjektet ved UiA. Prosjektet har undersøkt bruk av kunstig intelligens (KI) som støtte i utforming av karakterbegrunnelser i høyere utdanning.

Rapporten bygger på to livetester gjennomført i eksamensperiodene i vår- og høstsemesteret 2025. Den sammenfatter erfaringer fra deltagende sensorer på tvers av emner og fakulteter, basert på evalueringer innhentet etter gjennomførte tester.

I tillegg inkluderer rapporten tilbakemeldinger fra studenter, innhentet gjennom en spørreundersøkelse.

Formålet med rapporten er å gi et kunnskapsgrunnlag for videre vurdering av KI-baserte støtteverktøy i sensur- og vurderingsarbeid ved UiA.

Rapporten analyserer:

- Gjennomføringen av to livetester av KI-støttet begrunnelsesskriving ved UiA i vår- og høstsemesteret 2025, inkludert rekruttering og deltakelse på tvers av emner og fakulteter.
- Sensorenes erfaringer med bruk av KI-verktøyet i forbindelse med utforming av karakterbegrunnelser, herunder opplevd nytte, tidsbruk og arbeidsflyt.
- Kvalitetsmessige vurderinger av KI-genererte begrunnelser, inkludert behov for menneskelig kontroll, redigering og ansvarlig bruk.
- Tekniske og organisatoriske erfaringer knyttet til gjennomføring av testene, herunder samspill med eksisterende systemer og praksiser.
- Tilbakemeldinger fra studenter om bruk av KI til begrunnelsesskriving.

Rapporten er utarbeidet i mars/april 2026.

Innhold

Innhold	2
1. Bakgrunn for og formål med prosjektet.....	4
2. Prosjektets fremdrift før livetestene	5
3. Gjennomføring av livetester og datagrunnlag for evalueringen	6
3.1 Utvikling, tilgjengeliggjøring og transparens	6
3.2 Livetest våren 2025 (pilot).....	7
3.3 Livetest høsten 2025 (skalering)	8
3.4 Deltakelse i livetestene	9
3.5 Gjennomføring av evalueringen og datagrunnlag	9
3.6 Oppgaveformer i rapportert testing.....	10
4. Hovedfunn fra evalueringene	12
4.1 Opplevd nytte	12
4.2 Tidsbruk og effektivisering.....	13
4.3 Kvalitet og faglig kontroll.....	14
Opplevd kvalitet.....	14
Betydningen av sensorveiledning	15
Grad av redigering	16
Lengde på utkast	17
4.4 Helhetlig vurdering	18
Opplevd nytte av deltakelse	19
Mangler i utkastene	19
5. Studentenes tilbakemeldinger.....	21
5.1 Formål med å be om begrunnelse	21
5.2 Bruk av sensorveiledning	21
5.3 Bruk og opplevelse av begrunnelser	21

5.4 Holdninger til bruk av KI i begrunnelser	22
6. Anbefalinger basert på evalueringen	23
6.1 Standardiser lengde, tone og formål	23
6.2 Innfør en fast mal for begrunnelse.....	24
6.3 Forenkle styring av utkastet (prompt-støtte)	24
6.4 Reduser risiko for feil og uønsket påvirkning fra karakter	25
6.5 Avklar rollefordeling: KI som støtteverktøy, ikke erstatning	25
7. Oppdateringer og tiltak som er iverksatt som følge av evalueringene	26
Videre tiltak til vurdering	26
8. Konklusjon	27

1. Bakgrunn for og formål med prosjektet

Siden 2023 har UiA, på initiativ fra Utdanningsavdelingen og viserektor for utdanning, testet hvordan kunstig intelligens (KI) kan brukes til å generere utkast til begrunnelse for karakter.

I tråd med universitets- og høyskoleloven har studenter ved UiA rett til å få en begrunnelse for sensurvedtaket innen to uker etter at de har fremsatt en forespørsel (2024, § 11-8). I henhold til lovverket skal en begrunnelse inneholde både de generelle prinsippene som er lagt til grunn for sensurvedtaket, samt en konkret vurdering av studentens prestasjon.

Arbeidet med å utforme slike begrunnelser kan være tidkrevende, og praksisen varierer mellom emner og sensorer. Samtidig har økt tilgang til og interesse for ansvarlig bruk av kunstig intelligens aktualisert behovet for å undersøke hvordan slike verktøy kan støtte arbeidet.

For å forenkle arbeidet med begrunnelser undersøker prosjektet om generativ kunstig intelligens kan benyttes til å utarbeide førsteutkast basert på eksamensoppgaven, sensorveiledningen og studentens besvarelse. Sensor har ansvar for å gjennomgå, redigere og ferdigstille begrunnelsen før den sendes til studenten.

Målet med prosjektet er å undersøke om generativ kunstig intelligens kan:

- støtte sensorer i arbeidet med å skrive begrunnelser
- bidra til mer konsistente og transparente begrunnelser
- redusere tidsbruk uten å svekke kvaliteten

Samtidig er det et overordnet mål å bidra til økt kvalitet i begrunnelsene og en mer effektiv arbeidsprosess, samt å opparbeide praktisk og organisatorisk erfaring med bruk av KI i utdanning.

For å sikre et solid kunnskapsgrunnlag har prosjektet kombinert innsiktsarbeid og teknisk utvikling.

2. Prosjektets fremdrift før livetestene

Høsten 2023 ble det gjennomført en tidlig test av den planlagte løsningen, administrert av prosjektgruppen. I denne fasen ble modenheten til språkmodeller undersøkt, med særlig vekt på om de kunne generere utkast til begrunnelser som oppleves som nyttige for sensorer. Dette var spesielt relevant ettersom språkmodeller på dette tidspunktet fortsatt var relativt nye.

I denne proof-of-concept-fasen mottok prosjektgruppen testdata i form av eksamensoppgaver, sensorveiledninger og studentbesvarelser fra de deltakende emnene, levert av eksamenskontoret. Sensorene ble dermed ikke pålagt ekstra arbeid, men fikk i etterkant tilgang til en samling KI-genererte utkast til begrunnelser for alle kandidater i de aktuelle eksamenene. Disse utkastene ble generert av språkmodellen GPT-4 via API.

Høsten 2024 ble verktøyet videreutviklet basert på erfaringer og tilbakemeldinger fra den innledende testen i 2023. Utviklingsarbeidet besto særlig i å forbedre instruksjonene til modellen (systemprompten) og å etablere et enkelt brukergrensesnitt, hvor sensorer selv kunne laste opp relevante data og generere utkast til begrunnelser. Dette brukergrensesnittet ble bygget på plattformen «Dify» og tok i bruk GPT-4o til å generere utkast.

I stedet for å teste løsningen i pågående eksamener, ble representanter fra tre fakulteter invitert til en serie workshops i mars og april 2025. Her fikk deltakerne anledning til å prøve ut løsningen på tidligere eksamener og tilhørende besvarelser fra egne emner.

Utvelgelsen av fakulteter var basert på en spørreundersøkelse gjennomført i desember 2024 blant sensorer ved UiA. Undersøkelsen kartla både hvordan sensorer arbeider med begrunnelser, og hvilke holdninger de har til bruk av KI i dette arbeidet. Den inngikk som en sentral del av prosjektets innsiktsarbeid, med mål om å vurdere i hvilken grad den planlagte løsningen kunne møte eksisterende behov ved UiA. Samtidig bidro innsiktsarbeidet til brukerinvolvering og forankring av utviklingsarbeidet blant fremtidige brukere.

Undersøkelsen hadde en svarprosent på nær 35 %. Funnene er publisert i en [egen oversiktsrapport](#) (Universitetet i Agder, 2025) på prosjektets nettsider, og har også blitt delt i ulike nasjonale fora.

3. Gjennomføring av livetester og datagrunnlag for evalueringen

Dette kapitlet beskriver gjennomføringen av livetestene i begrunnelsesprosjektet, inkludert utvikling og tilgjengeliggjøring av verktøyet, rekruttering av deltakere og omfanget av testingen i de to periodene. Videre redegjøres det for datagrunnlaget som ligger til grunn for evalueringen av prosjektet.

3.1 Utvikling, tilgjengeliggjøring og transparens

Utviklingen av verktøyet for KI-støttet begrunnelsesskriving har vært en sentral del av prosjektet. Arbeidet har omfattet både teknisk utvikling og utforming av brukerrettede løsninger, med mål om å gjøre verktøyet tilgjengelig og anvendelig for sensorer i en reell arbeidssituasjon.

Verktøyet er utviklet med utgangspunkt i eksisterende arbeidsprosesser for sensur, og legger til rette for at sensorer kan generere utkast til begrunnelser basert på eksamensoppgaven, sensorveiledningen og studentens besvarelse. Samtidig er det lagt til grunn at sensor har det fulle faglige og juridiske ansvaret for den endelige begrunnelsen, og at KI-generert innhold alltid skal vurderes, redigeres og kvalitetssikres før bruk.

For å støtte implementering og bruk ble det etablert et enkelt brukergrensesnitt, der sensorer selv kan laste opp relevante dokumenter og generere utkast. Løsningen er utviklet med vekt på lav terskel for bruk og fleksibilitet på tvers av fagområder og vurderingsformer.

Som en del av prosjektets arbeid med å sikre transparens og erfaringsdeling, ble det opprettet en egen [prosjektside på intranettet](#). Her er det publisert informasjon om prosjektets formål, oppbygning og bruk av verktøyet. Nettsiden inneholder også tekniske spesifikasjoner, inkludert beskrivelse av hvordan språkmodellen er instruert gjennom systemprompt.

Systemprompten i v3.0 - høsten 2025 - er følgende:

You are an AI assistant designed to evaluate and provide comprehensive, personalized feedback on student assignments. You will be shown a grading rubric, an assignment, and a grading of the assignment. Write feedback appropriate for the assigned grade. Focus on giving concrete feedback regarding where the student performed well or performed poorly. Try to refer to the grading rubric with your reasoning, if appropriate.

The feedback for the assignment should be provided in the language the student has written in. So if the assignment has been written in Norwegian, for example, your feedback should also be written in Norwegian. Or if the assignment is written in English, write the feedback in English.

Some of the assignments will include multiple choice questions. You will be provided a summary of the points the student has scored in these cases. Do not write too granular feedback regarding these multiple choice questions. If you choose to comment on the multiple choice questions, mention them as a whole, rather than individually.

Do not be afraid of criticising the student. Don't include pleasantries such as "Good luck on your future assignments".

Prosjektsiden har hatt flere funksjoner. For det første har den fungert som en støtte for deltakende sensorer, ved å gi tilgang til nødvendig informasjon og veiledning. For det andre har den bidratt til åpenhet om hvordan KI brukes i prosjektet, både overfor studenter og ansatte. For det tredje legger siden til rette for erfaringsdeling og kunnskapsutvikling i sektoren, ved at sentrale komponenter i løsningen er gjort tilgjengelige.

Samlet sett har utvikling og tilgjengeliggjøring av verktøyet vært tett koblet til prinsipper om ansvarlig bruk av kunstig intelligens, der transparens, etterprøvbarhet og menneskelig kontroll har vært sentrale føringer. Arbeidet er i tråd med [UiAs ambisjoner og retningslinjer for bruk av kunstig intelligens](#), hvor ansvarlig og transparent bruk av teknologi står sentralt. Det understøtter også [universitetets overordnede strategi](#), særlig mål 4, ved å bidra til utvikling av fremtidsrettede og kunnskapsbaserte arbeidsformer i utdanning.

3.2 Livetest våren 2025 (pilot)

Livetesten i vårsemesteret 2025 hadde karakter av en avgrenset pilot, med mål om å teste verktøyets funksjonalitet og relevans i en reell sensurkontekst. Testen ga også innsikt i hvordan sensorer tok i bruk løsningen i praksis, og hvilke utfordringer som oppstod.

Totalt 32 sensorer fikk tilgang til verktøyet, hvorav 10 gjennomførte evalueringen i etterkant. Deltakerne representerte flere fakulteter, med størst andel fra Fakultet for helse- og idrettsvitenskap (n=11), etterfulgt av Fakultet for humaniora og pedagogikk (n=7) og Fakultet for samfunnsvitenskap (n=6). De øvrige deltakerne (n=8) kom fra de resterende fakultetene.

Verktøyet ble i hovedsak testet i forbindelse med skriftlig skoleeksamen og hjemmeeksamen. Bruken var i stor grad eksplorativ, og flere sensorer oppgir at testingen ikke nødvendigvis var knyttet til faktiske begrunnelser sendt til studenter. Dette skyldtes blant annet at sensurperioden i enkelte tilfeller allerede var avsluttet, eller at det ikke forelå forespørsler om begrunnelse.

Videre gjennomførte ikke alle påmeldte sensorer testingen som planlagt. Av de som fikk tilgang til verktøyet, var det 10 som av ulike grunner ikke tok løsningen i bruk i løpet av testperioden. Dette påvirket både omfanget av testingen og datagrunnlaget for evalueringen.

Samlet sett viser pilotfasen at verktøyet i denne perioden i hovedsak ble brukt til utprøving og utforskning, snarere enn som en integrert del av sensorenes ordinære arbeidsprosess. Erfaringene fra vårtesten har likevel vært sentrale for videre utvikling av verktøyet og planleggingen av en mer omfattende test i høstsemesteret 2025.

3.3 Livetest høsten 2025 (skalering)

I høstsemesteret 2025 ble livetesten skalert opp, med mål om å undersøke verktøyet sine anvendelse i en bredere og mer representativ sensurkontekst. Sammenlignet med pilotfasen ble det lagt større vekt på systematisk rekruttering, tydeligere forankring i fagmiljøene og økt grad av faktisk bruk i pågående sensurarbeid.

Informasjon om prosjektet ble publisert på prosjektets nettside og gjennom [en nyhetssak på uia.no](#) 13. november 2025. Samme dag ble dekaner og instituttledere invitert til å videreformidle informasjon om deltakelse til sine fagmiljøer. I tillegg mottok et utvalg på 396 sensorer en personlig invitasjon til deltakelse per e-post 26. november 2025.

Denne rekrutteringsstrategien resulterte i betydelig høyere deltakelse enn i vårtesten. Totalt 73 emner ble meldt på testen, fordelt på 63 sensorer. Forskjellen mellom antall emner og sensorer skyldes at enkelte sensorer meldte på flere emner, og at noen emner hadde flere sensorer tilknyttet.

Evalueringen ble fullført av 30 sensorer i høstsemesteret 2025. I tillegg oppga 10 av de påmeldte at de av ulike grunner ikke fikk testet verktøyet i løpet av perioden. Samlet sett innebærer dette at en betydelig større andel av deltakerne gjennomførte evalueringen sammenlignet med pilotfasen i vårsemesteret.

Deltakerne representerte alle fakulteter ved UiA, med størst andel fra Fakultet for helse- og idrettsvitenskap og Fakultet for humaniora og pedagogikk. Verktøyet ble testet på et

bredere spekter av vurderingsformer enn i vårtesten, inkludert skriftlig skoleeksamen, hjemmeeksamen, gruppeoppgaver og prosjektarbeid.

I motsetning til pilotfasen var testingen i større grad knyttet til faktisk sensurarbeid, noe som ga et mer relevant og robust grunnlag for evaluering av verktøyets nytte og anvendelse i praksis.

3.4 Deltakelse i livetestene

Tabellen under viser fordelingen av deltakere på tvers av fakultetene i vår- og høstsemesteret 2025, basert på påmeldingsskjema. For høstsemesteret er det også registrert antall emner per deltaker.

Deltakelsen økte betydelig fra vår- til høstsemesteret, både i antall sensorer og antall emner. Samtidig viser fordelingen at prosjektet har hatt deltakere fra alle fakulteter ved UiA i høstperioden.

Fakultet for helse- og idrettsvitenskap og Fakultet for humaniora og pedagogikk har hatt den største deltakelsen i begge perioder, mens også Fakultet for teknologi og realfag og Handelshøyskolen ved UiA hadde en tydelig økning i høstsemesteret.

Forskjellen mellom antall sensorer og antall emner i høstperioden skyldes at enkelte sensorer deltok med flere emner, og at enkelte emner hadde flere sensorer tilknyttet.

Hvilket fakultet er du tilknyttet?	Påmeldte personer		Påmeldte emner
	Vår 2025	Høst 2025	
Fakultet for helse- og idrettsvitenskap	11	22	23
Fakultet for humaniora og pedagogikk	7	14	19
Fakultet for kunstfag	1	1	1
Fakultet for samfunnsvitenskap	6	4	3
Fakultet for teknologi og realfag	2	11	16
Handelshøyskolen ved UiA	3	9	9
Total	30	61	71

3.5 Gjennomføring av evalueringen og datagrunnlag

Datagrunnlaget for evalueringen skiller seg fra deltakelsen i livetestene, ettersom ikke alle deltakere gjennomførte evalueringen.

Evalueringen bygger på to oversiktsrapporter generert gjennom analyseverktøyet i SurveyXact. Rapportene er basert på utfylte og innsendte evalueringsskjemaer fra deltakere i

livetestene i vår- og høstsemesteret 2025. Datagrunnlaget fra høstsemesteret er vesentlig større enn fra pilotfasen i vårsemesteret.

Totalt var 91 personer påmeldt livetestene i de to periodene. Av disse gjennomførte 40 personer evalueringen, noe som gir en samlet gjennomføringsgrad på evalueringen på 44 %.

Det er viktig å merke seg at enkelte besvarelser er delvis utfylt. Dette innebærer at antall svar kan variere mellom de ulike spørsmålene. For eksempel fullførte 10 sensorer evalueringen i vårsemesteret 2025, mens 9 av disse besvarte spørsmålet om fakultetstilknytning. Tilsvarende variasjoner forekommer også i datagrunnlaget fra høstsemesteret.

For spørsmål der ikke alle respondentene har avgitt svar, vil svarprosenten kunne avvike fra den samlede gjennomføringsgraden. Dette må tas i betraktning ved tolkning av resultatene i de påfølgende kapitlene.

3.6 Oppgaveformer i rapportert testing

Figurene under viser hvilke oppgaveformer respondentene oppgir å ha testet verktøyet på i henholdsvis vår- og høstsemesteret 2025. Oversikten er basert på data fra evalueringsskjemaene og gjenspeiler dermed rapportert bruk, ikke nødvendigvis all faktisk testing.

I vårsemesteret (n=9) var testingen i hovedsak konsentrert om tradisjonelle eksamensformer. Flertallet av respondentene oppgir at verktøyet ble testet i forbindelse med skriftlig skoleeksamen (67 %, 6 svar), etterfulgt av hjemmeeksamen (22 %, 2 svar). Øvrige oppgaveformer er i liten grad representert i denne perioden.

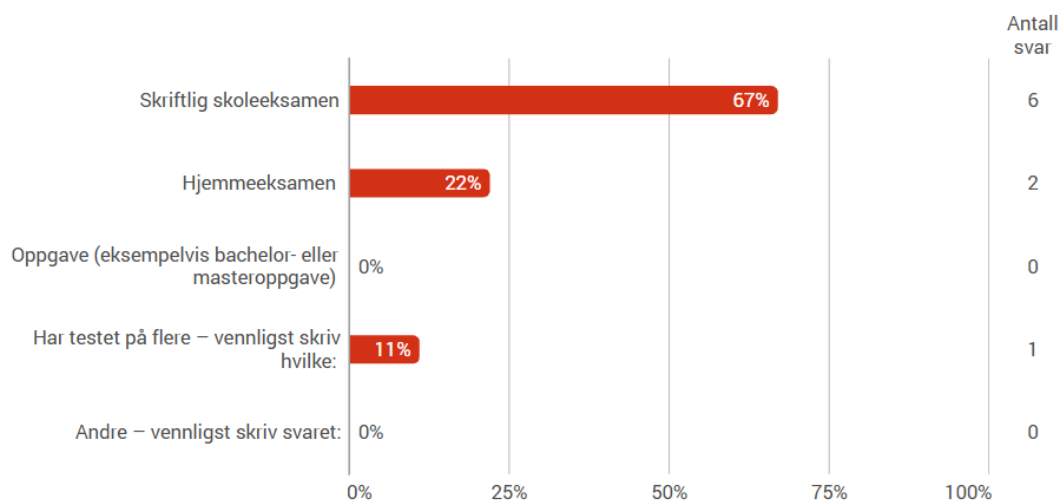
I høstsemesteret (n=31) var testingen både mer omfattende og mer variert. Skriftlig skoleeksamen (45 %, 14 svar) og hjemmeeksamen (29 %, 9 svar) utgjør fortsatt de største kategoriene, men det rapporteres også testing knyttet til større skriftlige arbeider, som bachelor- og masteroppgaver (10 %, 3 svar), samt tilfeller der verktøyet er testet på flere oppgaveformer (10 %, 3 svar) og andre vurderingsformer (6 %, 2 svar).

Samlet sett indikerer dette at verktøyet i høstsemesteret ble prøvd ut i et bredere spekter av vurderingssituasjoner enn i pilotfasen, basert på respondentenes egenrapportering.

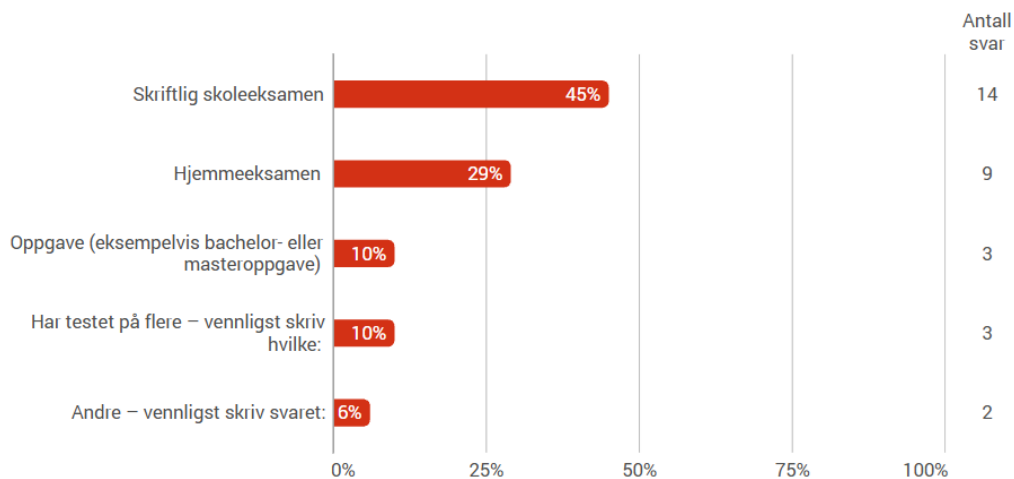


Hvilken oppgaveform har du testet med?

Våren 2025



Høsten 2025



4. Hovedfunn fra evalueringene

Dette kapitlet presenterer hovedfunnene fra evalueringene gjennomført i forbindelse med livetestene i vår- og høstsemesteret 2025. Funnene er basert på både kvantitative og kvalitative data fra evalueringsskjemaene.

Det er viktig å merke seg at resultatene bygger på deltakernes egenrapportering, og at datagrunnlaget er større og mer representativt i høstsemesteret enn i pilotfasen våren 2025. Funnene må derfor tolkes i lys av dette.

Kapitlet er strukturert tematisk og belyser blant annet opplevd nytte, tidsbruk, kvalitet på de genererte utkastene og overordnede vurderinger av verktøyet.

Funnene gir et samlet bilde av hvordan verktøyet fungerer i praksis, og danner grunnlag for videre vurdering av bruk av KI i begrunnelsesskriving ved UiA.

4.1 Opplevd nytte

Respondentene vurderer gjennomgående at KI-utkastene i stor grad svarer til det de antar at studenter ønsker å få ut av en begrunnelse.

I vårtesten 2025 (n=10) oppga samtlige respondenter at utkastene i minst noen grad svarte til studentenes behov. Halvparten svarte «i noen grad» (50 %), mens resten fordelte seg på «i stor grad» (40 %) og «i aller høyeste grad» (10 %).

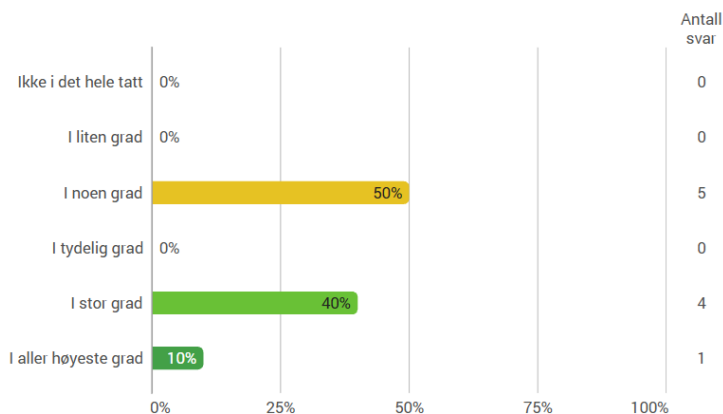
I høsttesten 2025 (n=28) er vurderingene fortsatt positive, men mer konsentrert i de høyere svaralternativene. Ingen respondenter svarte «ikke i det hele tatt» eller «i liten grad». Samtidig svarte 36 % «i tydelig grad», 29 % «i stor grad» og 11 % «i aller høyeste grad».

Samlet sett tyder dette på at KI-utkastene i økende grad oppleves å treffe det studentene etterspør i en begrunnelse, særlig etter oppskaleringen i høstsemesteret.

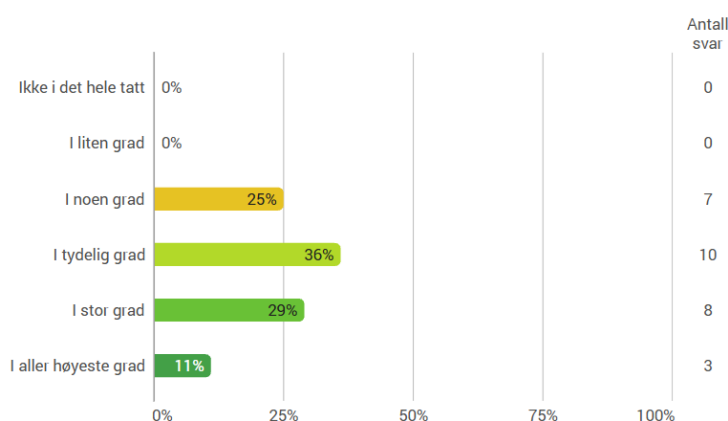


I hvilken grad synes du at KI-utkastene svarte til det du tror studentene ønsker å få ut av en begrunnelse?

Våren 2025



Høsten 2025



4.2 Tidsbruk og effektivisering

Effekten på tidsbruk varierer mellom pilot- og oppskaleringsfasen.

I vårperioden 2025 (n=10) oppgir de fleste redusert tidsbruk: 90 % brukte mindre tid, mens 10 % brukte like lang tid. Ingen oppga økt tidsbruk.

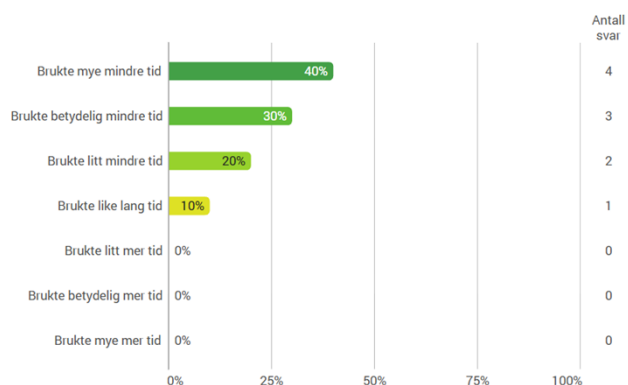
I høstperioden 2025 (n=26) er bildet mer sammensatt. 39 % oppgir redusert tidsbruk, 27 % uendret tidsbruk, og 34 % økt tidsbruk.

Dette kan indikere at mens verktøyet i pilotfasen i stor grad ble brukt eksplorativt og opplevdes tidsbesparende, førte mer reell bruk i høstperioden til økt bevissthet om behov for kvalitetssikring og redigering, noe som i enkelte tilfeller økte tidsbruken.

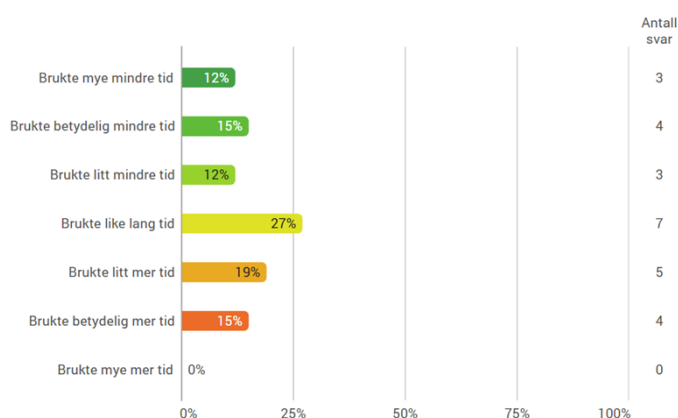


I hvilken grad endret KI-utkastene tidsbruken din ifm. begrunnelsene? Du bes kun å svare på tiden som gikk til selve utformingen (generering og redigering) av begrunnelsene – ikke kommunikasjonen rundt prosjektdeltakelse, gjennomgang av bruksveiledningen eller lignende

Våren 2025



Høsten 2025



4.3 Kvalitet og faglig kontroll

Opplevd kvalitet

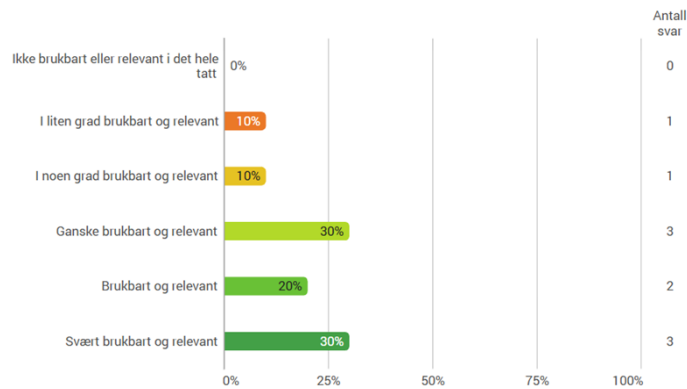
Respondentene vurderer gjennomgående kvaliteten på utkastene som brukbar, men med variasjon.

I vårperioden opplevde de fleste verktøyet som relevant og nyttig, med en tydelig overvekt av positive vurderinger. I høstperioden er vurderingene mer spredt, med flere moderate vurderinger og et fåtall som opplever lav brukbarhet. Ingen vurderer imidlertid verktøyet som helt ubrukbart.

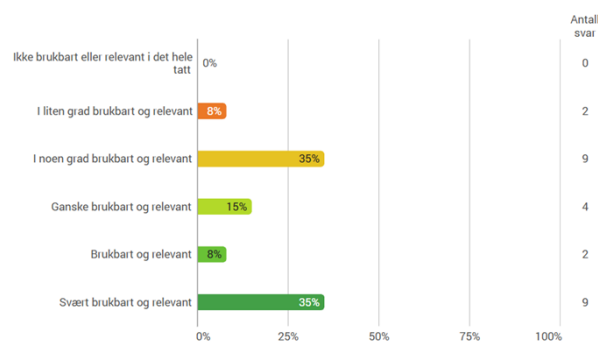


Hvor brukbart og relevant opplevde du verktøyet, med tanke på kvaliteten på utkastene som ble generert?

Våren 2025



Høsten 2025



Betydningen av sensorveiledning

I begge testperioder oppgir majoriteten at kvaliteten på sensorveiledningen har betydning for kvaliteten på de genererte utkastene.

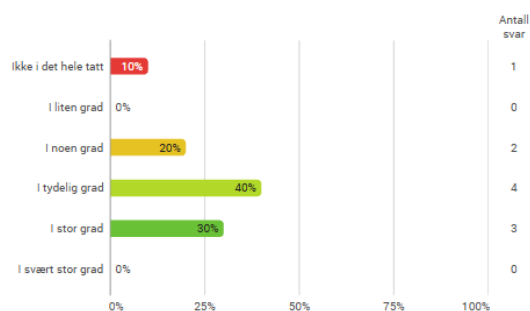
Fritekstsvarene peker på tre hovedtendenser:

- Kvaliteten på utkastene avhenger av hvor tydelig og detaljert sensorveiledningen er
- Mer konkrete og strukturerte veiledninger gir bedre utkast
- Det er usikkerhet knyttet til hvordan veiledninger bør utformes for optimal bruk med KI

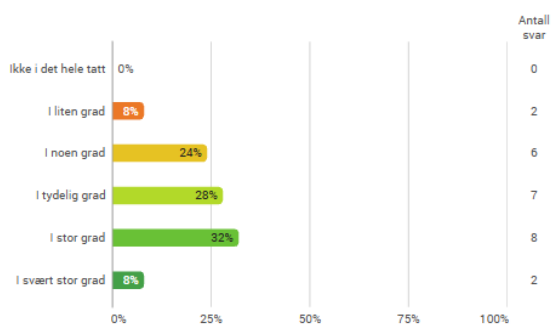


Opplevde du at kvaliteten på sensorveiledningen du brukte i testen, hadde innflytelse på kvaliteten på utkastene som ble generert?

Våren 2025



Høsten 2025



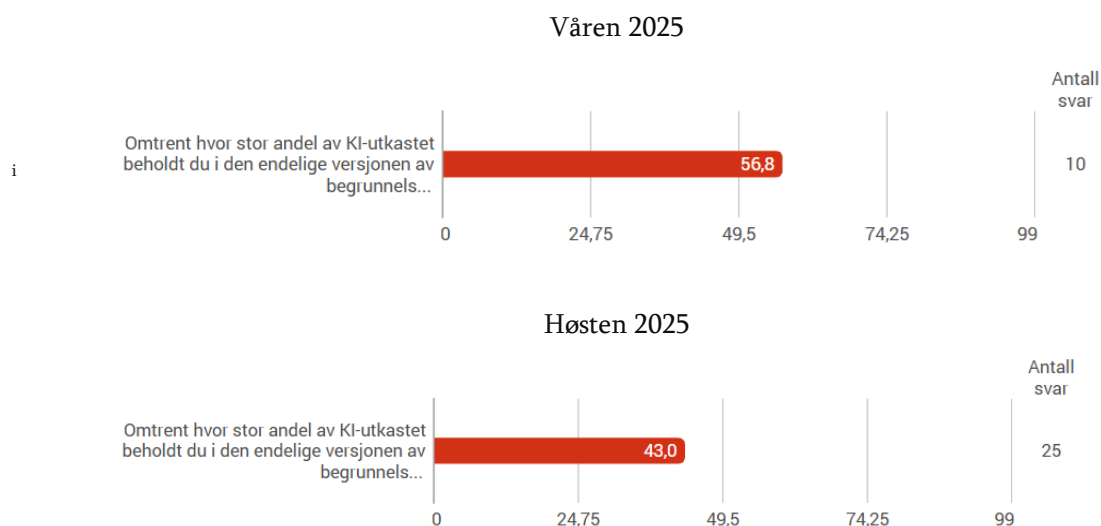
Grad av redigering

Respondentene oppgir å beholde en mindre andel av KI-utkastet i høstperioden (43,0 %, n=25) enn i vårperioden (56,8 %, n=10).

Dette kan indikere at verktøyet i økende grad brukes som støtte for struktur og idéutvikling, mens sensoren i større grad omformulerer og tilpasser teksten før den ferdigstilles.



Omtrent hvor stor andel av KI-utkastet beholdt du i den endelige versjonen av begrunnelsen? Angi i prosent (0-100%)

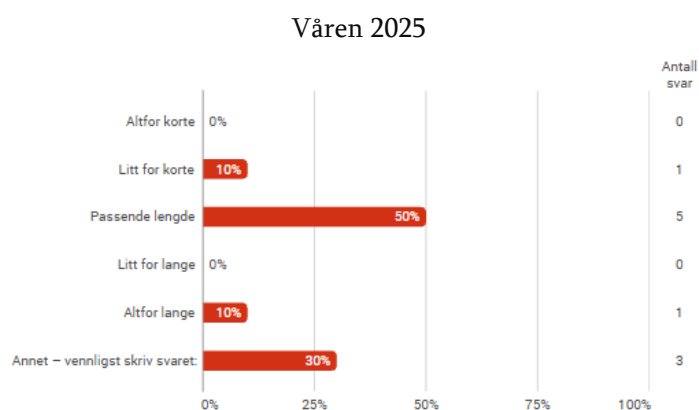


Lengde på utkast

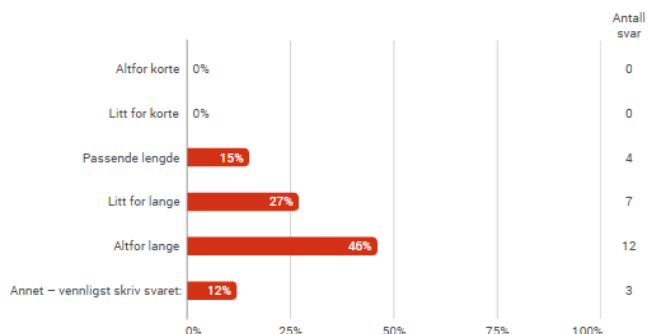
Flere respondenter opplever at utkastene, særlig i høstperioden, er for lange og krever betydelig redigering.



Hva tenkte du om lengden på utkastene som ble generert?



Høsten 2025



4.4 Helhetlig vurdering

Den overordnede vurderingen av kvaliteten på løsningen er moderat positiv.

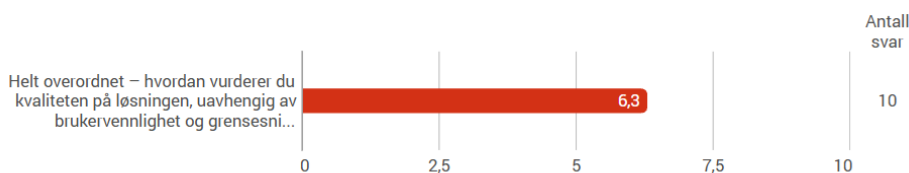
I vårperioden fikk løsningen en gjennomsnittsscore på 6,3 (n=10), mens den i høstperioden fikk 6,0 (n=23). Begge verdiene ligger over midtpunktet på skalaen, noe som indikerer at utkastene oppleves som brukbare, men med tydelig forbedringspotensial.

Det er samtidig forskjeller knyttet til vurderingsform. I høstperioden er gjennomsnittsscoren høyere for skoleeksamen (score på 7,0; n=13) enn for hjemmeeksamen (score på 4,2, n=6), noe som kan tyde på at verktøyet fungerer bedre i mer standardiserte vurderingssituasjoner og at kvaliteten har økt på denne typen eksamensform ifm utviklingen siden våren.

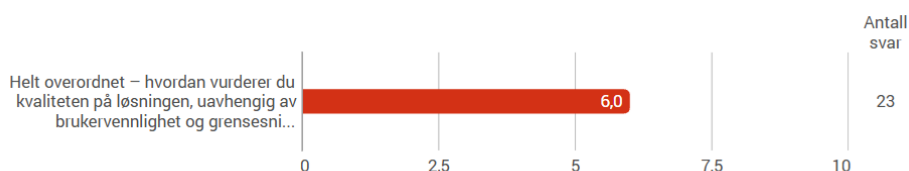


Helt overordnet – hvordan vurderer du kvaliteten på løsningen, uavhengig av brukervennlighet og grensesnitt? Tenk spesielt på kvaliteten på utkastene som ble generert, og gi score på en skala fra 1 til 10 der 1 betyr «svært dårlig» og 10 betyr «svært god».

Våren 2025



Høsten 2025



Opplevd nytte av deltakelse

Respondentene vurderer deltakelse i utprøvingen som nyttig. I begge perioder oppgir et stort flertall at deltakelsen var ganske eller svært nyttig (90 % i vårperioden og 83 % i høstperioden).

Fritekstsvar peker særlig på:

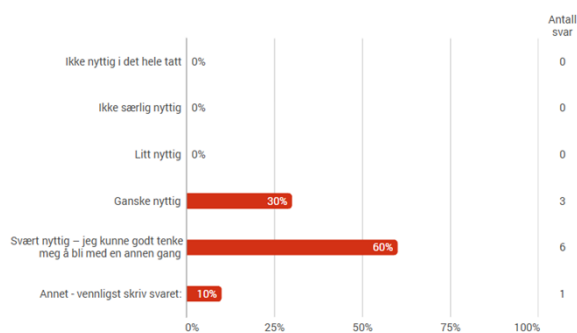
- opplevd støtte i vurderingsarbeidet
- læringseffekt
- økt trygghet i vurderingssituasjonen

Samtidig etterlyser enkelte flere relevante oppgaver for å kunne vurdere nytten bedre.

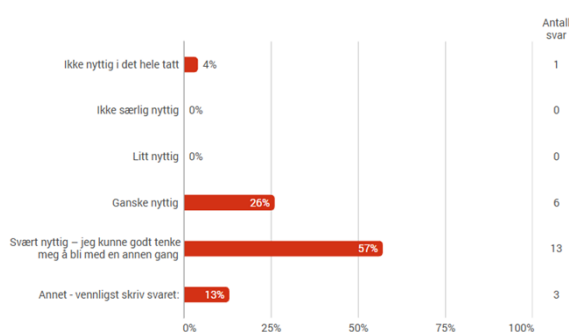


Helt overordnet – hvor nyttig opplevde du det å være med i utprøvingen av verktøyet?

Våren 2025



Høsten 2025



Mangler i utkastene

Svarene viser delte vurderinger av om utkastene mangler elementer som bør inngå i en god begrunnelse.

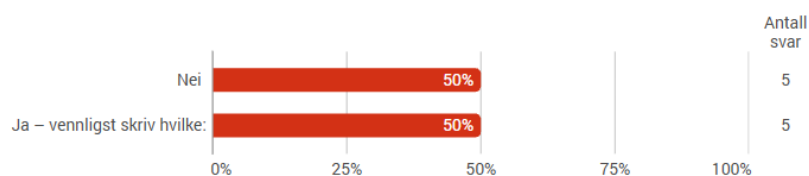
I høstperioden svarer 57 % nei og 43 % ja, mens fordelingen i vårperioden er 50/50.

Kommentarene peker særlig på:

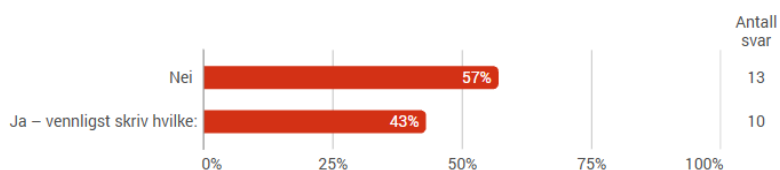
- behov for kortere og mer konsise tekster
- ønske om tydeligere overordnet karakterbegrunnelse
- behov for flere konkrete eksempler fra studentens besvarelse
- ønske om større mulighet for å styre utkastene gjennom prompts

Opplevde du at det manglet elementer i utkastet som bør inngå i en god begrunnelse?

Våren 2025



Høsten 2025



5. Studentenes tilbakemeldinger

Studentenes tilbakemeldinger gir innsikt i hvordan begrunnelser oppleves fra mottakerperspektivet, og hva studenter faktisk ønsker å få ut av en begrunnelse. Funnene bygger på en spørreundersøkelse med 1696 respondenter, noe som tilsvarer en gjennomføringsgrad på 13 %.

5.1 Formål med å be om begrunnelse

Studentene ble bedt om å angi de viktigste grunner til at de ber om begrunnelse. Svarene tyder på at studentene først og fremst oppfatter begrunnelse som et læringsverktøy, der de kan vurdere egen presatsjon.

De viktigste grunnene er:

- 1) Vurdere om karakteren er rettferdig
- 2) Forstå hvordan de kan forbedre seg
- 3) Vurdere om de vil klage på karakter
- 4) Vurdere om de vil ta opp eksamen på nytt

5.2 Bruk av sensorveiledning

Et stort flertall av studentene oppgir at de leser sensorveiledningen:

- 48 % svarer at de gjør det
- 43 % svarer at de gjør det dersom de finner den

Kun et lite mindretall (4 %) oppgir at de ikke leser veiledningen, eller ikke vet hva det er (5 %).

Dette indikerer at mange studenter aktivt oppsøker informasjon for å forstå vurderingsgrunnlaget, og at sensorveiledningen spiller en viktig rolle i denne sammenhengen.

5.3 Bruk og opplevelse av begrunnelser

De fleste studentene oppgir at de leser begrunnelsene de mottar:

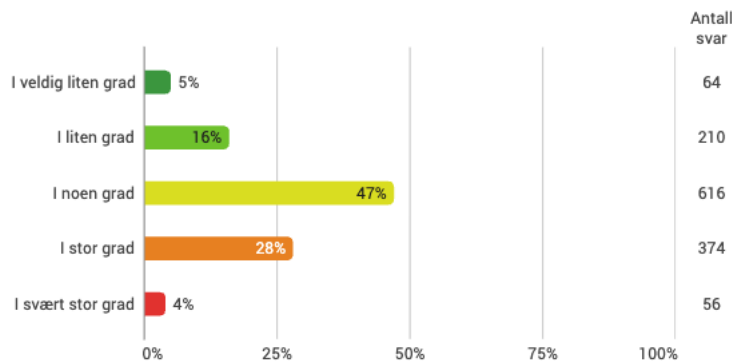
- 78 % svarer ja
- svært få (2 %) oppgir at de ikke gjør det
- 20 % som har besvart undersøkelsen har ennå ikke bedt om en begrunnelse

Samtidig viser svarene at begrunnelsene ikke alltid svarer fullt ut til studentenes forventninger. Flertallet oppgir at begrunnelsene i noen grad (47 %) eller i stor grad samsvarer (28 %) med forventningene, men relativt få (4 %) opplever at de gjør det i svært stor grad.

Dette tyder på at begrunnelser oppleves som nyttige, men at det finnes et forbedringspotensial i hvordan de utformes.



Samsvarte begrunnelsen(e) du har motatt med dine forventninger?



5.4 Holdninger til bruk av KI i begrunnelser

Studentene uttrykker gjennomgående skepsis til bruk av KI i begrunnelsesskriving – selv når utkastene er kvalitetssikret av sensor.

- 55 % oppgir at de ville vært skeptiske
- 20 % svarer at det kommer an på eksamensformen
- kun 12 % oppgir at de ville vært positive

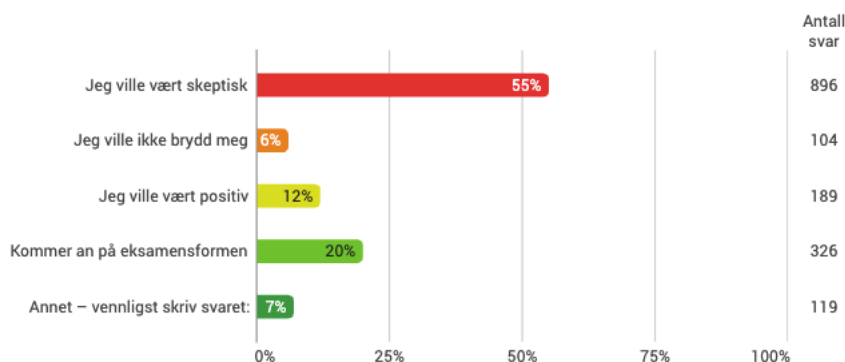
Fritekstsvarene forsterker dette bildet og peker særlig på:

- bekymringer knyttet til rettssikkerhet og faglig kvalitet
- spørsmål om tillit til vurderingsprosessen
- opplevelsen av at bruk av KI kan være prinsipielt problematisk

Samtidig uttrykker en mindre gruppe at KI kan være et nyttig støtteverktøy, forutsatt at den faglige vurderingen fortsatt ligger hos sensoren.



Hva ville du synes dersom begrunnelsen du mottar er basert på et KI-generert utkast som sensoren har kvalitetssikret før utsendelsen?



5.5 Studentenes samlede vurderinger

Studentenes fritekstkommentarer gir et mer nyansert bilde av behov og forventninger:

- Mange opplever at kvaliteten på dagens begrunnelser er for lav
- Begrunnelser anses som viktige for læring, rettssikkerhet og tillit
- Flere ønsker at begrunnelse skal gis automatisk sammen med karakter
- Det er gjennomgående skepsis til å la KI skrive begrunnelser
- Samtidig åpnes det for bruk av KI som støtteverktøy, men ikke som erstatning for menneskelig vurdering

6. Anbefalinger basert på evalueringen

På tvers av evalueringene fra vår og høst 2025 peker tilbakemeldingene i samme retning: Verktøyet har et tydelig potensial for å støtte sensorer i arbeidet med begrunnelser, men krever videreutvikling for å være tilstrekkelig presist, robust og tidsbesparende i praksis.

Nedenfor presenteres sentrale anbefalinger for videre utvikling.

6.1 Standardiser lengde, tone og formål

Flere sensorer opplevde at de KI-genererte utkastene var for omfattende og for formativt orientert, sammenlignet med behovet for en kort, summativ begrunnelse.

Dette kommer blant annet til uttrykk i utsagn som:

- «Problemet var at det var for mye tekst»

- «Begrunnelsene var altfor omfattende»
- «Utkastet er ... en formativ vurdering ... i stedet for en begrunnelse for en karakter»

Det anbefales å etablere en tydelig standard for begrunnelsens lengde, tone og formål. En kort, saklig og mer nøktern fremstillingsform vil kunne redusere behovet for etterarbeid og bidra til mer konsistente begrunnelser.

6.2 Innfør en fast mal for begrunnelse

Flere sensorer etterlyser en mer enhetlig struktur og formulering:

- «Skulle ønske det kunne legges inn en mal for begrunnelse slik at alle ble like»
- «Jeg kunne ønske at verktøyet brukte mer enhetlige formuleringer»

Det anbefales å utvikle en fast mal for begrunnelser, med:

- tydelig struktur
- fast rekkefølge på elementer
- kortfattede og standardiserte formuleringer

Dette vil kunne bidra til økt konsistens, enklere kvalitetssikring og mer effektiv bruk.

6.3 Forenkle styring av utkastet (prompt-støtte)

Flere sensorer opplever at kvaliteten på utkastene i stor grad avhenger av hvor godt de klarer å formulere instruksjoner til verktøyet.

Dette illustreres gjennom utsagn som:

- «Utkastene ble etter hvert bra når det ble promptet riktig»
- «Jeg brukte mye tid på å prøve ut ulike instruksjoner»

Det anbefales å redusere behovet for individuell prompting ved å tilby:

- forhåndsdefinerte valg (f.eks. lengde, nivå, formål)
- forslag til standardiserte instruksjoner
- enkle justeringsmuligheter i grensesnittet

Dette vil gjøre verktøyet mer tilgjengelig og redusere tidsbruk.

6.4 Reduser risiko for feil og uønsket påvirkning fra karakter

Evalueringen viser eksempler på at modellen i for stor grad lar den oppgitte karakteren styre innholdet i begrunnelsen, samt tilfeller av faktiske feil.

Eksempler fra tilbakemeldingene:

- «Verktøyet synes å skjele for mye til den oppgitte karakteren»
- «Da jeg endret karakteren til F, kommenterte KI feil som ikke var der»
- «Opplevde at KI genererte direkte feil»

Det anbefales å utvikle mekanismer som:

- i større grad forankrer begrunnelsen i den faktiske besvarelsen
- reduserer risiko for hallusinasjoner
- gjør det enklere for sensor å identifisere og kontrollere usikre eller svakt funderte påstander

6.5 Avklar rollefordeling: KI som støtteverktøy, ikke erstatning

Tilbakemeldingene peker tydelig på at sensorer ønsker å beholde kontrollen over begrunnelsene.

Dette uttrykkes blant annet gjennom:

- «Så lenge jeg som sensor har kontroll på det som sendes ut»
- «Vi skal ALLTID gjennomgå svar fra AI»
- «Det er problematisk om ... begrunnelsen blir sendt automatisk»

Samtidig uttrykkes det et klart potensial dersom kvaliteten forbedres.

Det anbefales derfor å videreutvikle verktøyet som et støtteverktøy for sensorer – ikke som et fullt automatisert system. Sensorens rolle som faglig ansvarlig og kvalitetssikrer bør ligge fast.

7. Oppdateringer og tiltak som er iverksatt som følge av evalueringene

Som følge av tilbakemeldingene fra evalueringene er det gjennomført flere justeringer og forbedringer av verktøyet.

Følgende tiltak er implementert:

- Språkmodellen er justert for å generere kortere og mer konsise utkast.
- En brukerdeklarasjon på norsk og engelsk inkluderes nå automatisk, med mindre bruker aktivt velger å fjerne den.
- Det er utviklet en dedikert front-end som gir større fleksibilitet i videreutviklingen av brukergrensesnittet.
 - Dette inkluderer blant annet mulighet for å veksle mellom norsk og engelsk.
- Maksimal lengde på dokumenter som kan behandles er økt betydelig.

I tillegg har verktøyet fått en mer varig teknisk løsning:

- Verktøyet er nå tilgjengelig via en stabil URL
- Tilgangsstyring håndteres gjennom Feide-pålogging, noe som sikrer at løsningen er tilgjengelig for ansatte innenfor trygge rammer

Videre tiltak til vurdering

På bakgrunn av evalueringene kan det også være aktuelt å vurdere ytterligere tiltak.

Ett konkret forslag er å utvikle en egen ressurs som gir brukerne tilgang til:

- forslag til strukturerte og kvalitetssikrede prompter
- veiledning i hvordan ulike parametere (f.eks. lengde og nivå) kan justeres

Samtidig bør behovet for en slik ressurs avklares nærmere i dialog med fagmiljøene, for å sikre at tiltaket svarer på reelle behov hos brukerne.

8. Konklusjon

Evalueringen av begrunnelsesprosjektet ved UiA viser at bruk av generativ kunstig intelligens har et tydelig potensial som støtte i arbeidet med karakterbegrunnelser.

På tvers av testperiodene opplever sensorer at verktøyet kan bidra til struktur, effektivisering og økt bevissthet om begrunnelsenes innhold. Samtidig viser både sensor- og studenttilbakemeldinger at kvalitet, presisjon og faglig kontroll er avgjørende for tillit til løsningen.

Studentenes tilbakemeldinger understreker betydningen av begrunnelser som læringsverktøy, men peker også på en tydelig skepsis til bruk av KI i vurderingssammenheng. Dette stiller krav til hvordan slike verktøy utvikles, implementeres og kommuniseres.

Samlet sett peker funnene mot at KI bør videreutvikles som et støtteverktøy for sensorer, ikke som en erstatning for faglig vurdering. Videre arbeid bør særlig rette seg mot å forbedre kvalitet, redusere behov for etterarbeid og sikre tydelige rammer for ansvarlig bruk.

Universitetet i Agder. (2025). *Bruk av kunstig intelligens i begrunnelsesskriving. En oversiktsrapport.*
https://www.uia.no/intranett/arbeidssotte/ki-ved-uia/oversiktsrapport_ki-i-begrunnelser.pdf

Universitets- og høyskoleloven. (2024). *Lov om universiteter og høyskoler* (LOV-2024-03-08-9). Lovdata.
https://lovdata.no/dokument/NL/lov/2024-03-08-9/KAPITTEL_11#KAPITTEL_11