4 / 2026

RAPPORT

Store språkmodeller i analyseprosesser

Erfaringer fra NOKUTs tematiske analyser i tredje runde med periodisk oppfølging

2026

NOKUT

NOKUT – Nasjonalt organ for kvalitet i utdanningen – er et statlig forvaltningsorgan under Kunnskapsdepartementet.

NOKUT har et eget styre og er faglig uavhengig i oppgaver som gjelder akkreditering, tilsyn og evalueringer for å vurdere kvaliteten i høyere utdanning og høyere yrkesfaglig utdanning. I tillegg har NOKUT forvaltningsoppgaver på vegne av departementet.

Formålet med NOKUTs virksomhet er å føre tilsyn med kvaliteten i høyere utdanning og høyere yrkesfaglig utdanning, og å stimulere til kvalitetsutvikling som sikrer et høyt internasjonalt nivå i utdanningstilbudene ved institusjonene. NOKUTs formål og oppgaver er forankret i universitets- og høyskoleloven og fagskoleloven.

NOKUT skal bidra til at samfunnet har tillit til kvaliteten i norsk høyere utdanning og høyere yrkesfaglig utdanning. Gjennom arbeidet skal NOKUT søke å bistå institusjonene i kvalitetsarbeidet deres.

NOKUT bruker sakkyndige i akkrediteringer, tilsyn, evalueringer og prosjekter.

Du kan lese mer om arbeidet vårt på nokut.no.

Tittel	Store språkmodeller i analyseprosesser
Forfatter(e)	Gustavo Guajardo, Jon Lanested, Philipp E. L. Friedrich
Dato	15.01.2026
Rapportnummer	4-2026
ISSN-nr	1892-1604

Forord

NOKUTs tredje runde med periodisk oppfølging med systematisk kvalitetsarbeid i perioden 2017–2024 har resultert i til sammen 48 tilsyn og tilhørende tilsynsrapporter. Dette har gitt oss anledning til å se nærmere på fellestrekk, felles utfordringer og gode praksiser i institusjonenes systematiske kvalitetsarbeid de siste åtte årene.

I arbeidet med å organisere, analysere og verifisere datamaterialet og funnene har vi fått verdifull støtte fra en intern referansegruppe i NOKUT. Gruppen har bestått av kolleger med solid erfaring fra den tredje tilsynsrunden: Ane B. Lillehammer, Eva Refsdal, Frøydis Maurtvedt, Hedvig Maria Bergem, Hege Brodahl, og samt Gustavo Guajardo som har bistått med spisskompetanse i bruk av kunstig intelligens i analysearbeid.

Rapportene har som mål å støtte videre utvikling av systematisk kvalitetsarbeid i sektoren. Vi håper funnene kan gi et nyttig grunnlag for refleksjon, læring og videre kvalitetsutvikling.

Innhold

Sammendrag

Denne rapporten undersøker hvordan en stor språkmodell, nærmere bestemt *NotebookLM* utviklet av *Google DeepMind*, kan brukes som støtteverktøy i analyseprosesser innen ekstern kvalitetssikring av høyere utdanning. Datagrunnlaget for analysen er 48 tilsynsrapporter fra NOKUTs tredje runde med periodisk oppfølging av høyere utdanning i Norge. Analysen legger særlig vekt på generelle utfordringer og gode praksiser i intern kvalitetssikring og kvalitetsarbeid. Ved å bruke *NotebookLM* til å analysere rapportene har vi fått innsikt i hva modellen identifiserer som de største utfordringene og gode praksiser i institusjonenes kvalitetsarbeid. Sammenlignet med en manuell analyse viser *NotebookLM* en styrke i å strukturere store datamengder og identifisere mønstre, mens den manuelle analysen gir en dypere forståelse av kontekst og nyanser. Kombinasjonen av analyser støttet av kunstig intelligens og manuell analyse fremstår derfor som den mest treffsikre og helhetlige tilnærmingen. Funnene viser at *NotebookLM* har potensial som analytisk støtte, men at bruken blant annet krever et strukturert datagrunnlag, presise instruksjoner (*prompts*), kompetanse i bruk av kunstig intelligens, faglig kompetanse på saksfeltet og menneskelig kvalitetssikring.

1 Innledning

Denne delrapporten har som formål å undersøke hvordan store språkmodeller kan brukes til å analysere datagrunnlaget fra den tredje runden av NOKUTs periodiske oppfølging av institusjonene innen høyere utdanning i Norge, og hvordan en slik tilnærming står i forhold til tradisjonelle, manuelle analysemetoder.[1] Analysen bygger utelukkende på offentlig tilgjengelige data i form av NOKUTs offisielle tilsynsrapporter, som ikke inneholder personsensitive opplysninger. Hovedrapportens problemformulering omfatter tre spørsmål, hvorav det tredje lyder som følger:

- ***Hvordan kan NOKUT anvende kunstig intelligens i analyser av datagrunnlaget fra den tredje runden med periodisk oppfølging?***

I denne delrapporten presenterer vi altså ikke en fullstendig gjennomgang av funnene fra den tredje runden med periodisk oppfølging, men fokuserer i stedet på en refleksjon rundt erfaringene med å bruke språkmodellen som et verktøy i analysen av NOKUTs tilsynsrapporter. Delrapporten inngår i en større tematisk analyse bestående av tre sammenhengende rapporter (se tabell 1). Det overordnede formålet med analysen er å videreutvikle NOKUTs metode for periodisk oppfølging, støtte institusjonenes interne kvalitetsarbeid og gi myndighetene, sektoren og samfunnet bedre innsikt i utfordringer og utviklingsmuligheter i intern kvalitetssikring av høyere utdanning.

Med utgangspunkt i den pågående teknologiske utviklingen vurderer vi det i tillegg som viktig å utforske hvordan nye teknologiske verktøy som store språkmodeller, heretter også omtalt som «modell» eller «kunstig intelligens (KI)», kan bidra til å effektivisere og forbedre analyse- og evalueringsprosesser, særlig i eksterne kvalitetssikringsorganer. Samtidig antar vi at slike modeller også kan bidra til nye perspektiver på kvalitetsarbeid i høyere utdanning.

Vi mener derfor at denne rapporten er særlig relevant for andre kvalitetssikringsorganer som vurderer å ta i bruk KI i sine analyseprosesser. For en helhetlig forståelse av det systematiske kvalitetsarbeidet ved norske universiteter og høyskoler, anbefaler vi også å lese de to øvrige rapportene i denne tematiske analysen.

Tabell 1. Oversikt rapportstruktur

NOKUTs tematiske analyse	
Rapport A	Tematisk analyse av NOKUTs tredje runde med periodisk oppfølging – hovedrapport
Rapport B	Tematisk analyse av NOKUTs tredje runde med periodisk oppfølging – høyskoler med akkrediterte studietilbud og uten selvakkrediteringsmyndighet
Rapport C	**Store språkmodeller i analyseprosesser – erfaringer fra NOKUTs tematiske analyser i tredje runde med periodisk oppfølging**

[1] Vi har brukt den lisensierte versjonen av *Microsoft Copilot* til å korrigere og språkvaske rapportene i denne tematiske analysen, for å sikre god språklig kvalitet, men også for å teste denne funksjonen i praksis.

2 Datagrunnlaget og analyseprosessen

2.1 Datagrunnlaget

Den primære datakilden for denne analysen har vært tilsynsrapportene fra NOKUTs tredje runde med periodisk oppfølging. I alt 48 rapporter på totalt ca. 1 800 sider, som dekker perioden 2017–2024, har blitt gjennomgått av språkmodellen for å analysere institusjonenes kvalitetsarbeid, slik det er beskrevet og vurdert i tilsynsrapportene. Hvert av de ni tilsynskravene er vanligvis vurdert på to til tre sider i en tilsynsrapport.[2] For hvert tilsyn oppnevnte NOKUT en sakkyndig komite som er ansvarlig for vurderingene i tilsynsrapportene. Vurderingene tar utgangspunkt i dokumentasjonen som institusjonen har sendt inn i forbindelse med tilsynene, med informasjon fra institusjonsbesøk som sekundærkilde. I runde tre har NOKUT krevd dokumentasjon både på institusjonsnivå og for utvalgte studietilbud. Dokumentasjonen på institusjonsnivå skulle gi et helhetlig bilde av hvordan institusjonen overordnet arbeider systematisk med å sikre og videreutvikle kvaliteten i utdanningene. Dokumentasjonen på studietilbudsnivå skulle gi innsikt i hvordan kvalitetsarbeidet faktisk praktiseres i studietilbudene, og fungerer som konkrete eksempler på bruken av kvalitetssystemet. I vurderingene legges det særlig vekt på hvordan kvalitetsarbeidet ved institusjonen etterlever de spesifikke kravene, når det gjelder både egne rutiner og praksis. Se vedlegg 2 i denne rapporten for en liste med alle institusjonene som er vurdert i tilsynsrapportene.

2.2 Bruken av KI-modellen *NotebookLM*

Den KI-assisterte analysen av tilsynsrapportene ble utført ved hjelp av *Notebook LM*. *Notebook LM* er en modell utviklet av *Google DeepMind* for kontekstuell informasjons-analyse og kunnskapsutvikling. Det er en KI-modell som bruker *Google* sin *Gemini*-modell til å hjelpe brukere med å analysere store mengder tekstbasert informasjon. Denne modellen fungerer som en KI-drevet assistent for dem som jobber med komplekse tekstsamlinger, og hjelper til med å hente ut nøkkelinformasjon, oppsummere dokumenter og finne sammenhenger på tvers av tekster.

Notebook LM er en modell som analyserer og trekker ut informasjon fra spesifikke kilder brukeren gir den. For at modellen skal kunne besvare spørsmål, må den først få tilgang til et datagrunnlag. Dette datagrunnlaget kan bestå av opplastede dokumenter, lenker til nettsider eller videoer. Modellen går gjennom dette materialet og baserer alle svar utelukkende på den informasjonen den finner der. Det betyr at den ikke trekker inn kunnskap fra eksterne databaser eller internett generelt, slik som tradisjonelle språkmodeller som *ChatGPT-4o*, *Gemini* eller *Claude* ofte gjør. *Notebook LM* bygger på *Retrieval-Augmented Generation* (*RAG*), men i motsetning til mange andre *RAG*-løsninger har den ikke tilgang til en stor, ekstern kunnskapsbase. Den arbeider kun med det innholdet brukeren gir den tilgang til. På den måten har brukeren full kontroll over hvilken

[2] Se vedlegg 1 for mer informasjon om ordlyden i kravene til systematisk kvalitetsarbeid.

informasjon modellen bruker som grunnlag for svarene sine. Hvis man for eksempel stiller spørsmålet «Hvordan koker man et egg?», vil modellen bare kunne svare dersom denne informasjonen finnes i de opplastede dokumentene eller lenkene som brukeren har delt med modellen. Hvis denne informasjonen ikke finnes i de opplastede filene, vil modellen heller ikke forsøke å generere et svar basert på gjetning eller generell kunnskap. Det betyr at muligheten for at modellen finner på eller dikter opp et svar, er lavere enn hos andre «vanlige» modeller. Et ytterligere nyttig aspekt ved *Notebook LM* er at modellen inneholder en referanse til kilden der den har funnet svaret. Denne funksjonen bidrar til å lette arbeidet med å kvalitetssikre svarene som modellen gir.

Selv om noen versjoner av *Gemini*-modellen støtter kontekstvinduer på opptil én million *tokens*[3] (noe som i teorien tillater at svært store tekstmengder kan behandles samtidig) betyr ikke dette at *NotebookLM* alltid analyserer hele dokumentet på én gang. Som nevnt over bruker verktøyet i praksis *RAG*-metoden, der det kun henter og behandler relevante tekstbiter ved behov, for å gi raskere og mer presise svar. Dette gjør det mulig å håndtere større datamengder, men nøyaktigheten synker dersom dokumentene er dårlig strukturerte, redundante eller inneholder motstridende informasjon. Ifølge *Google*s hjelpesider fungerer *NotebookLM* mest presist med et samlet dokumentvolum på 100 til 200 sider, forutsatt at tekstene er godt strukturerte og konsistente (Google Support 2025). Derfor kan det i enkelte tilfeller oppstå feil, også fordi modellen ikke forstår innholdet på samme måte som et menneske. Modellen genererer svar basert på sannsynlige tekstmønstre og kan derfor noen ganger feiltolke eller kombinere informasjon på en måte som fører til upresise slutninger. Menneskelig kvalitetssikring av svarene er derfor viktig, særlig i faglige spørsmål og i situasjoner som trenger svært presise svar.

2.3 Metodeutvikling og systematisering av datagrunnlaget

En sentral forutsetning for å oppnå meningsfulle og relevante resultater med modeller som *NotebookLM* er utformingen av selve forespørselen, den såkalte *prompten,* heretter også omtalt som *instruks*.

En *prompt* er instruksjonen som gis til språkmodellen for å få frem bestemte typer svar eller analyser. I en analyse av tilsynsrapporter kan dette for eksempel være en forespørsel om å identifisere hvordan en institusjon beskriver sitt kvalitetsarbeid, hvordan studentenes tilbakemeldinger integreres i kvalitetsarbeidet, eller hvordan tilbakemeldingssløyfer er organisert. Hvordan denne instruksen formuleres, har stor betydning for modellens respons, med hensyn til både relevans, dybde og nyansering (Brown et al., 2020; Reynolds & McDonell, 2021).

[3] *Tokens* er små tekstbiter som språkmodeller bruker for å forstå og prosessere tekst. En *token* kan være et ord, en del av et ord eller tegnsetting. For eksempel deles setningen «Dette er en test.» opp i ca. 5 *tokens*. Modeller har en grense for hvor mange *tokens* de kan håndtere samtidig (dette kalles «kontekstvinduet»). Selv om noen modeller støtter opptil 1 million *tokens*, henter verktøy som *NotebookLM* kun inn relevante tekstutdrag i stedet for å lese alt på én gang.

Begrepet *prompt- eller instruksdesign* refererer til det systematiske arbeidet med å utvikle og forbedre slike forespørsler. I en vitenskapelig analyseprosess, der presisjon og tolkning er sentrale metodiske hensyn, blir instruksdesign en avgjørende del av prosessen. Tidligere studier har vist hvordan små variasjoner i formuleringer kan gi vesentlige forskjeller i svarenes kvalitet og relevans (Zhou et al., 2022). Spesielt i analyser av komplekse og kontekstavhengige tekster, slik som tilsynsrapporter, kan nøye utformede instrukser gjøre det mulig å trekke ut innsikt som ellers ville krevd manuell koding.

I de følgende avsnittene viser vi hvordan vi har brukt *NotebookLM* til å analysere tilsynsrapportene, og hvordan vi har arbeidet med instruksdesign som en del av metodeutviklingen. Vi redegjør for ulike typer instrukser, fra åpne og utforskende spørsmål til strukturerte og instruksjonsbaserte tilnærminger, og diskuterer hvordan disse påvirket hvilke innsikter som kunne hentes ut.

2.3.1 Utforskningsfasen

I starten av analyseprosessen stilte vi korte og spesifikke spørsmål istedenfor lengre og mer omfattende spørsmål/instrukser. Med denne tilnærmingen ønsket vi å redusere risikoen for falsk informasjon og samtidig øke muligheten for modellen til å finne frem til relevante og nøyaktige svar basert på datagrunnlaget.

Vi oppdaget flere utfordringer vi måtte løse før vi kunne gå videre i prosessen. Først og fremst innså vi at antallet tilsynsrapporter (48) og det totale antallet sider (ca. 1 800) i utgangspunktet var for høyt til at modellen kunne gå gjennom dem på én gang og hente ut all informasjon en instruks krevde. En årsak til dette er at rapportene ikke er helt like. Selv om de alle følger en overordnet struktur med tilnærmet like kapitteloverskrifter, bærer de preg av å være skrevet av ulike sakkyndige komiteer og saksbehandlere i NOKUT. Disse små variasjonene kan utgjøre en utfordring for en KI-modell som skal finne spesifikk informasjon i en omfattende rapport. Særlig når samme type informasjon er plassert litt ulike steder i rapportene, kan det bli vanskelig for *Notebook LM* (eller andre store språkmodeller) å finne alt som bør inngå i et svar. Dette samsvarer med de begrensningene *Google* selv har pekt på, der de anbefaler å bruke *NotebookLM* med et moderat antall sider og dokumenter for å sikre presise og relevante svar.

En ytterligere utfordring vi støtte på, var at institusjonene kan bli omtalt med fullt navn, som «Universitet i Oslo», eller med en forkortelse, som «UiO». Det samme gjaldt filnavnene. Noen av filene hadde institusjonens fulle navn, mens andre bare hadde forkortelsen. Dette førte til at modellen ble «forvirret» og enten brukte feil forkortelse for en institusjon eller diktet opp en forkortelse som ikke fantes.

2.3.2 Forenkling og systematisering av datagrunnlaget

I lys av disse utfordringene ble det tydelig at vi har behov for enda bedre strukturerte tilsynsrapporter, med kapitler eller avsnitt som (så langt det lar seg gjøre) har akkurat samme tittel og oppbygning på tvers av rapportene. En annen utfordring var at innholdet,

selv om det overordnet sett er konsistent, tidvis tematiseres under ulike krav i de forskjellige rapportene. For eksempel kan «tilbakemeldingssløyfer» i institusjonenes kvalitetsarbeid omtales under flere krav i én rapport, mens det samme innholdet behandles under til dels andre krav i en annen. Dette skyldes at diskusjonen av slike aspekter ofte er kontekstavhengig og kan variere fra institusjon til institusjon. Samtidig gjør dette det vanskeligere å veilede modellen presist til relevante avsnitt og å sammenligne innholdet på tvers av tilsynsrapportene.

Vi valgte derfor en løsning som delte opp innholdet og gjorde det enklere for modellen å identifisere og sammenligne relevant informasjon mellom rapportene. På bakgrunn av dette brukte vi også en annen språkmodell, *ChatGPT-4o,* for å stille den spørsmål om hvordan vi kan støtte *Notebook LM* i analyseprosessene, samt for å utforske mulige løsninger som kan sikre høy kvalitet i modellens svar. En løsning som ble foreslått av *ChatGPT-4o*, var å organisere tilsynsrapportene i separate mapper (såkalte *Notebooks*) i *NotebookLM*, basert på en valgt tematisk kategori, og deretter stille de samme spørsmålene per mappe. Vi valgte å bruke institusjonenes akkrediteringsmyndighet (også kalt akkrediteringsnivå) som overordnet prinsipp for en grov inndeling av tilsynsrapportene. Siden det finnes fire institusjonskategorier – 1) universiteter, 2) vitenskapelige høyskoler, 3) akkrediterte høyskoler og 4) høyskoler med akkrediterte studietilbud, men uten institusjonsakkreditering – opprettet vi fire tilsvarende mapper i *NotebookLM*.[4] Selv om vi i etterkant analyserte hver rapport separat, bidro den inndelte mappestrukturen til å begrense konteksten modellen måtte forholde seg til, og vi reduserte dermed risikoen for feil. For eksempel er det lettere for modellen å identifisere fellesutfordringer og fellestrekk i institusjonelt kvalitetsarbeid på tvers av institusjonstyper, fordi analysen gjennomføres sekvensielt innenfor grupper med likhetstrekk. Dette legger til rette for sammenligning av funn mellom gruppene fremfor at modellen skal håndtere hele materialet samtidig. I litteraturen omtales slike strategier ofte som *chunking*, der større datamengder deles opp i mindre, håndterbare enheter for å redusere kontekstbelastning og støtte språkmodellens analyse. I dette prosjektet skjedde inndelingen primært på dokument- og kontekstnivå, gjennom tematisk gruppering av rapporter i separate mapper, snarere enn ved oppdeling av enkelttekster i mindre tekstbiter (Liu et al. 2023). Formålet med denne inndelingen var altså ikke å analysere rapportene på gruppenivå, men å forbedre modellens arbeidsbetingelser.

Basert på disse vurderingene utarbeidet vi så et opplæringsdokument med to instrukser (se boks 1). For hver institusjon/tilsynsrapport ønsket vi at modellen skulle trekke ut følgende informasjon:

- institusjonskategori
- akkrediteringsmyndighet
- krav i studiekvalitetsforskriften § 2-1 (2) (med tilhørende instruks)

[4] Se rapport A og B for mer informasjon om de ulike institusjonskategoriene og akkrediteringsnivåene.

- krav i studietilsynsforskriften § 4-1 (3) (med tilhørende instruks)

Vi valgte å ta med kravene i studiekvalitetsforskriften § 2-1 (2) og studietilsynsforskriften § 4-1 (3) med to tilhørende instrukser, fordi analysen i de andre rapportene (A og B) viste at dette var de to kravene institusjonene hadde størst utfordringer med å oppfylle i runde tre.[5] Modellen skulle derfor konsentrere seg om disse kravene for å kunne lære å trekke ut overordnede og strukturelle utfordringer i institusjonelt kvalitetsarbeid.

I tillegg til tilsynsrapportene tok vi inn to ytterligere støttedokumenter i datagrunnlaget som *Notebook LM* kunne benytte. Det ene var NOKUTs veileder for systematisk kvalitetsarbeid publisert i 2022 («støttedokument 1»), som inneholder NOKUTs tolkning av kravene som stilles til universiteter og høyskolers systematiske kvalitetsarbeid, og en generell veiledning til hva slags type dokumentasjon som kan være aktuell når NOKUT gjennomfører tilsyn med kvalitetsarbeid ved en institusjon (NOKUT 2022). Det andre dokumentet («støttedokument 2») har vi utarbeidet spesielt for denne analysen, og det inneholder en systematisk oversikt over navn og forkortelser på alle høyere utdanningsinstitusjoner i Norge.

Boks 1. Struktur for opplæringsdokumentet

- **Institusjonskategori:** *universitet, vitenskapelig høyskole, akkreditert høyskole og høyskole uten institusjonsakkreditering*
    - ***NotebookLMs* svar: […]**
- **Akkrediteringsnivå:** *detaljert beskrivelse av institusjonens akkrediteringsmyndighet*
    - ***NotebookLMs* svar: […]**
- **Krav i studiekvalitetsforskriften § 2-1 (2) med tilhørende instruks:** Bruk dokumentet nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022 som grunnlag for å forstå innholdet i krav § 2-1 (2). Med utgangspunkt i dette kravet, hvilke hovedutfordringer nevnes i rapportene knyttet til oppfyllelse av § 2-1 (2)?
    - ***NotebookLMs* svar: […]**
- **Krav i studietilsynsforskriften § 4-1 (3) med tilhørende instruks:** Bruk dokumentet nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022 som grunnlag for å forstå innholdet i krav § 4-1 (3). Med utgangspunkt i dette, hvilke hovedutfordringer nevnes i rapportene når det gjelder oppfyllelse av § 4-1 (3)?
    - ***NotebookLMs* svar: […]**

[5] § 2-1 (2) i studietilsynsforskriften handler om at institusjonene skal gjennomføre periodiske evalueringer av studietilbudene sine, med bidrag fra arbeids- eller samfunnsliv, studenter og eksterne sakkyndige. Resultatene skal være offentlige. Kravet er videreført i universitets- og høyskoleforskriften § 1-8 (6). Kravet § 4-1 (3) handler om at institusjonene skal ha ordninger for systematisk å kontrollere at alle studietilbud tilfredsstiller gjeldende krav til akkreditering av studietilbud. Dette kravet er videreført i universitets- og høyskoleforskriften § 1-8 (5).

2.3.3 Utvikling av instrukser (prompts)

Basert på de svarene vi fikk fra *NotebookLM* i opplæringsdokumentet, spesielt på grunnlag av de to tilhørende instruksene for studiekvalitetsforskriften § 2-1 (2) og studietilsynsforskriften § 4-1 (3), videreutviklet vi instruksene ved hjelp av *ChatGPT-4o*. Vi informerte *ChatGPT-4o* om at vi arbeidet med *Notebook LM*, og siden den allerede hadde kjennskap til denne modellen, var det ikke nødvendig med ytterligere forklaring. Utviklingen av instruksene foregikk som en dialog med *ChatGPT 4o*, der vi ga modellen informasjon om hva *Notebook LM* hadde tilgang til, og hvilken type informasjon vi ønsket å hente ut fra dokumentene.

En generell anbefaling fra *ChatGPT-4o* var å gi *Notebook LM* tydelig veiledning når man formulerer en instruks. Når det gjaldt kravene til systematisk kvalitetsarbeid, anbefalte *ChatGPT-4o* at vi enten henviste direkte til de stedene i datagrunnlaget der kravene er beskrevet (for eksempel støttedokument 1), eller at vi oppsummerte eller siterte kravet i selve instruksen.

En mindre utfordring vi støtte på i begynnelsen med *Notebook LM*, var språket. Selv om modellen behersker flere språk enn engelsk, fikk vi likevel svar på engelsk. Dette skjedde til tross for at datagrunnlaget utelukkende bestod av dokumenter på norsk, og at instruksene også var skrevet på norsk. For å løse dette begynte vi å legge til uttrykket «svar på norsk» først i hver instruks. Språkutfordringene ved modellen ser ut til å fungere bedre nå. Vi har nylig erfart at modellen ikke lenger trenger ekstra veiledning for å bruke norsk fra starten.

I neste fase ble det oppdaterte opplæringsdokumentet, som nå inneholdt *NotebookLMs* svar, analysert ved hjelp av nye instrukser for å gjennomføre en mer omfattende og dyptgående analyse på tvers av institusjonene. For å oppnå dette delte vi det oppdaterte opplæringsdokumentet med *ChatGPT-4o* og informerte modellen om at vi ønsket å analysere det oppdaterte opplæringsdokumentet videre. Vi instruerte *ChatGPT-4o* om å foreslå to til fire nye instrukser som kunne gi ytterligere innsikt i felles utfordringer og strukturelle svakheter i institusjonelt kvalitetsarbeid, samt å trekke ut det som eventuelt kan anses som gode praksiser. Vi fikk fire forslag, men ett av dem var ikke relevant, ettersom det forutsatte at rapportene inneholdt beskrivelser av utfordringer fra institusjonene – noe de ikke gjorde siden de kun inneholder de sakkyndiges og NOKUTs beskrivelser. Vi informerte *ChatGPT-4o* om dette, og den foreslo deretter tre nye instrukser. Til slutt valgte vi å bruke de to første instruksene fra det opprinnelige forslaget samt to av de nye, som handlet om felles utfordringer, strukturelle svakheter og god praksis i institusjonenes kvalitetsarbeid (se tabell 2).

Tabell 2. Endelige instrukser

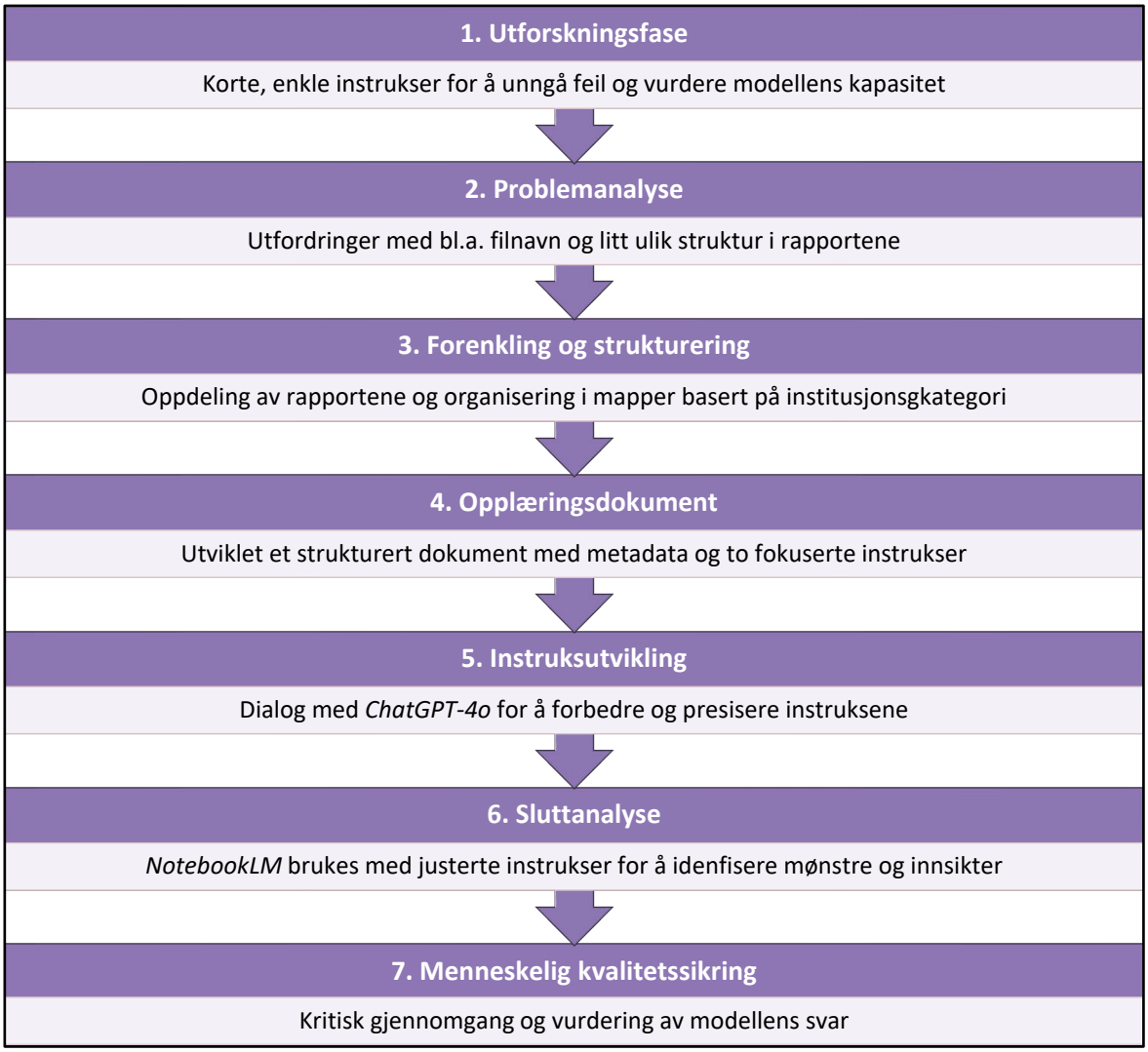
Nr.	Tema	Spørsmål
1	**Utfordringer med § 2-1 (2)**	– Hva er de mest gjentakende utfordringene institusjonene møter i oppfyllelsen av krav § 2-1 (2)? Er det forskjeller mellom ulike institusjonskategorier?
2	**Utfordringer med § 4-1 (3)**	– Hva er de mest gjentakende utfordringene institusjonene møter i oppfyllelsen av krav § 4-1 (3)? Er det forskjeller mellom ulike institusjonskategorier?
3	**Felles utfordringer og strukturelle svakheter**	– Hvilke utfordringer går igjen uavhengig av institusjonskategori, og hva kan dette si om felles strukturelle svakheter i kvalitetsarbeidet?
4	**Gode praksiser i kvalitetsarbeid i høyere utdanning**	– Hva kjennetegner godt kvalitetsarbeid ved høyere utdanningsinstitusjoner, basert på analysen av kildene? Hvilke tegn og indikatorer tyder på at en institusjon arbeider systematisk og målrettet med kvalitet, både i utdanningstilbudet og i organisasjonskulturen?

Instruksene i tabell 2 var de siste vi benyttet i analyseprosessen, blant annet fordi vi på det tidspunktet var fornøyde med kvaliteten på svarene vi fikk fra *NotebookLM*. Dette innebærer imidlertid ikke at vi oppfattet svarene som endelige eller publiserbare uten videre redigering og faglig vurdering.

Selv om vi ikke går i dybden på funnene her, kan vi likevel nevne at modellen identifiserte flere gjennomgående temaer knyttet til kvalitetsarbeid. Blant disse var utfordringer med **å systematisere kvalitetsarbeidet** og **uklare ansvarsområder og rapporteringslinjer**. I tillegg pekte modellen på at gode praksiser kjennetegnes av bruk av **kvalitetsresultater i strategisk styring**, **tydelig definerte roller og forankring i ledelse og styre** samt **koblinger mellom kvalitetsarbeid og områder som undervisning, forskning og læringsmiljø.** Disse funnene bidro til å synliggjøre hvordan modellen kan brukes til å avdekke mønstre i store tekstmengder, men også hvordan menneskelig vurdering fortsatt er avgjørende for tolkning, for kontekstualisering og for å sikre faglig presisjon i analysen.

Figur 1 under viser og oppsummerer hele prosessen for utviklingen av analysen og systematisering av datagrunnlaget knyttet til opplæringen av *Notebook LM*, med de ulike fasene som vi har vært gjennom.

Figur 1. Analyseutvikling og systematisering av datagrunnlaget



3 Diskusjon og metodiske refleksjoner

Våre erfaringer med *Notebook LM* viser at modellen kan bidra til å trekke frem innsikter, identifisere mønstre og gjøre taus kunnskap mer eksplisitt. Samtidig stiller dette arbeidet krav til struktur, metode og faglig skjønn. I det følgende presenterer vi noen refleksjoner over forskjeller og likhetstrekk mellom manuell og KI-assistert analyse, og drøfter fem sentrale aspekter som har betydning for hvordan kunstig intelligens kan brukes som støtte i analyseprosesser. Disse aspektene belyser både muligheter og begrensninger ved KI-verktøy i møte med komplekse vurderinger. Det dreier seg om struktur og datakvalitet, standardisering og faglig skjønn, modellens rolle og verdi i analysearbeidet, hvilke ferdigheter som kreves for å bruke teknologien hensiktsmessig, og etiske hensyn og tillit i bruken av KI.

3.1 Manuell versus KI-assistert analyse – noen refleksjoner

Felles for de rent manuelle analysene og analysene hvor vi har brukt KI som en hjelper i analysearbeidet, er at de peker på mange av de samme utfordringene i institusjonenes kvalitetsarbeid. Samtidig mener vi at det også er noen forskjeller mellom hva den manuelle analysen og den KI-assisterte analysen faktisk bidrar med. Den manuelle analysen åpner for en dypere forståelse av kontekst, prosess og intensjon. Vi har kunnet tolke språklige nyanser, vurdere institusjonenes modenhet og trekke inn vurderinger av hvordan kvalitetssystemene faktisk fungerer i praksis. Vi har erfart sakene på nært hold, ikke bare slik de fremstår i tilsynsrapportene. Dette står i kontrast til en utenforstående aktør, eller i dette tilfellet en maskin, som ikke har samme kontekstforståelse eller erfaringsgrunnlag.

NotebookLM har vist styrke i strukturering, oppsummering og identifisering av mønstre. Modellen har vært god til å hente ut eksplisitte elementer fra dokumentene, knytte funn til konkrete forskriftskrav og systematisere indikatorer på godt kvalitetsarbeid. Den har også bidratt til å tydeliggjøre felles utfordringer på tvers av institusjoner og utarbeide forslag til generaliserbare praksiser. Dette har vært nyttig i arbeidet med å se helheten og mønstre i store tekstmengder, og det har gitt oss et nytt blikk på hvordan funn kan kategoriseres og formidles. KI-analysen har også gjort oss mer bevisste på behovet for bedre datastruktur og systematikk, noe vi drøfter nærmere lenger ned i kapittelet. Samtidig ser vi at modellen kan trekke bastante slutninger, i likhet med mennesker, særlig når det gjelder årsakssammenhenger eller implisitte forhold som ikke nødvendigvis er faglig holdbare uten ytterligere vurdering.

Det største potensialet, slik vi ser det, ligger i kombinasjonen mellom manuell og KI-assistert analyse. Når vi bruker KI som støtteverktøy i analysen, for eksempel til å identifisere mønstre, strukturere store datamengder og foreslå kategorier, kan vi frigjøre tid og ressurser til andre oppgaver. På den måten blir KI et supplement, ikke en erstatning, og bidrar til å styrke kvaliteten og effektiviteten i det samlede analysearbeidet (Davison et al. 2024).

3.2 Sentrale aspekter i analyseprosessen

3.2.1 Struktur og datakvalitet

Notebook LM har i stor grad klart å identifisere relevant informasjon i tilsynsrapportene. Likevel er presisjonen i svarene (som forventet) sterkt avhengig av kvaliteten på og oppbyggingen av datagrunnlaget og utformingen av *promptene*. Selv om tilsynsrapportene i seg selv har høy faglig kvalitet, var noe av informasjonen enten uklart formulert, spredt over flere deler av rapportene eller vanskelig tilgjengelig for modellen grunnet uensartet struktur på tvers av rapportene.

Et sentralt teknisk aspekt ved bruk av store språkmodeller er derfor hvordan dokumenter deles opp i mindre tekstbiter for at modellen lettere skal kunne prosessere dem (Hitch 2024; Nguyen-Trung 2025). Dette skjer automatisk og tar ikke hensyn til tematisk sammenheng. Det medfører at informasjon som er fragmentert, ustrukturert eller kontekstavhengig, lettere kan bli feiltolket eller oversett. Vår erfaring er at kortere, tydeligere og mer målrettede tekster gir bedre resultater. En tydeligere struktur med enhetlige maler, klare overskrifter og bruk av tematiske markører vil bidra til å hjelpe både mennesker og KI-modeller med å forstå, analysere og sammenligne innhold på tvers av rapporter.

3.2.2 Standardisering og faglig skjønn

I NOKUTs tilsyn har vi erfart at både standardisering og faglig skjønn spiller viktige og utfyllende roller i vurderingen av institusjonenes kvalitetsarbeid. Variasjon i hvordan saksbehandlere og sakkyndige vurderer og dokumenterer informasjon, viser behovet for at vi bør ta noen ytterligere grep for økt standardisering på tvers av rapportene. Dette kan bidra til at det blir lettere å manøvrere på tvers av rapportene og legge et nødvendig metodisk grunnlag for at KI-modeller som *NotebookLM* skal kunne støtte analysearbeidet på en presis og konsistent måte.

Samtidig mener vi at økt standardisering av vurderingsarbeidet i våre tilsynsprosesser ikke bare vil understøtte teknisk konsistens, men også det faglige skjønnet som utøves i konteksten av institusjonenes egenart og kompleksitet. Vurderingsarbeidet til NOKUT og de sakkyndige i tredje runde har vist at fleksibilitet og tilpasning til lokale (institusjonelle) forhold kan fungere godt, uten at det nødvendigvis står i motsetning til for rigide metodiske rammer. Snarere vil tydelige metodiske rammer og begrepsmessig avklaring styrke det faglige skjønnet og sikre at vurderingene forblir konsistente, også når det kreves tilpasning til de ulike institusjonene. *NotebookLM* og lignende KI-verktøy stiller riktignok høyere krav til struktur og systematikk (Nguyen-Trung, 2025), noe vi anser som et insentiv til å videreutvikle standardiseringen og den faglige friheten i samspill, til gjensidig nytte.

3.2.3 Rollen til KI-modeller: assistent med analytisk verdi

Notebook LM har vist evne til å «lese mellom linjene» og trekke frem helhetlige innsikter som ikke nødvendigvis er eksplisitt formulert i rapportene. I én konkret forespørsel (*prompt* 3) ble modellen bedt om å identifisere felles utfordringer og strukturelle svakheter på tvers av institusjoner. Her viste den imponerende evne til å integrere informasjon fra ulike deler av datagrunnlaget, abstrahere over tematiske mønstre og peke på utfordringer i kvalitetsarbeidet. I tillegg klarte modellen å identifisere mulige underliggende årsaker til disse utfordringene. Slike analyser understreker KI-modellens potensial som assistent i analysearbeid; et verktøy som kan støtte, forsterke og strukturere, men ikke erstatte menneskelig fagkompetanse og skjønnsmessige vurderinger (Hitch, 2024).

Det oppstod noen feil som illustrerer begrensningene ved bruk av språkmodeller. For eksempel forvekslet modellen forkortelser og institusjonsnavn, noe som i flere tilfeller førte til usanne slutninger. Dette underbygger behovet for manuell kontroll og veiledning, ikke bare i selve analysen, men også i det forberedende arbeidet med datagrunnlaget.

3.2.4 Kompetanseutvikling i bruk av KI

Effektiv bruk av store språkmodeller forutsetter at saksbehandlere utvikler ny kompetanse. Det handler ikke bare om teknisk innsikt, men også om forståelse for hva KI kan brukes til, hvilke begrensninger KI har, og hvordan man kan legge til rette for gode analyser. KI-modeller kan på sikt bidra til mer effektiv bruk av menneskelige ressurser, blant annet gjennom automatisering av enkelte oppgaver. Men dette forutsetter kompetanseutvikling og en bevisst strategi for hvordan teknologien bør og kan innlemmes i eksisterende arbeidsprosesser.

I denne sammenhengen har vi også erfart at det kan være nyttig å bruke flere KI-modeller i samspill, for eksempel ved å benytte *ChatGPT-4o* til å utvikle og justere presise *prompter* til *NotebookLM*. Dette gir mulighet for en mer målrettet og effektiv utnyttelse av datagrunnlaget, der ulike modeller utfyller hverandre med sine styrker. Metodisk sett reiser dette spørsmål om arbeidsdeling, kontroll og transparens: Hvilken modell har hvilken rolle, og hvordan dokumenterer vi samspillet mellom dem? Utfordringen ligger særlig i å sikre konsistens og sporbarhet i analysene, ettersom ulike modeller opererer med ulike kapasiteter, kontekstvinduer og tolkningsrammer. Samtidig åpner det for et betydelig potensial: Ved å bruke én modell som sparringspartner og metodestøtte kan vi skjerpe både faglig presisjon og analytisk struktur i samhandlingen med den modellen som utfører selve analysen. Dette gir et nytt metodisk rom, men fordrer bevissthet, dokumentasjon og kritisk refleksjon i alle ledd.

3.2.5 Etiske hensyn og tillit til KI-genererte analyser

Bruk av KI i analysearbeid reiser uunngåelig spørsmål om tillit, ansvar og faglig integritet (Davison 2024; Hitch 2024; Nguyen-Trung 2025). Hvor mye kan vi stole på modellens svar? Hvordan sikrer vi at vurderingene som produseres, er faglig gyldige og etterprøvbare? Vi mener det er avgjørende at KI-resultater alltid valideres av mennesker med relevant kompetanse. Selv ved høy treffsikkerhet vil det fortsatt være behov for et sterkt faglig skjønn, ikke minst for å forstå kontekst, vurdere relevans og sette analysene inn i en helhetlig sammenheng. KI-modeller kan effektivisere bearbeiding av informasjon og bidra til analytisk støtte, noe som igjen kan frigjøre tid.

Et annet, relatert aspekt handler om datavern og rettigheter knyttet til bruk av data i analyseprosesser (Nguyen-Trung, 2025). Selv om terskelen for å bruke tilgjengelige data og dokumenter kan oppleves som lavere når man benytter KI-verktøy, bør de samme akademiske og vitenskapelige prinsippene fortsatt gjelde som ved manuell, ikke-KI-assistert analyse. I denne analysen har vi derfor utelukkende brukt NOKUTs offentlig tilgjengelige tilsynsrapporter som datagrunnlag, som ikke inneholder personsensitive opplysninger.

Det er også viktig å være oppmerksom på de miljømessige sidene ved KI-teknologi. KI-systemer krever store mengder strøm og energi (Verdecchia et al., 2023), noe som naturlig reiser spørsmål om hvorvidt ressursbruken står i forhold til den merverdien KI faktisk tilfører, for eksempel i analysearbeid. Selv om vi fortsatt befinner oss i en utforskende fase når det gjelder bruken av KI, der vi undersøker hvilke typer oppgaver teknologien egner seg for (som analyse, oppsummering eller språkvask) mener vi at bruken hele tiden må styres av både miljøhensyn og vurderinger av faktisk nytte. I tillegg er det nødvendig å være oppmerksom på teknologiavhengighet som et potensielt sårbarhetsområde, særlig i forbindelse med ansvarlig bruk av KI. En økende avhengighet av slike verktøy kan svekke selvstendig faglig vurdering og kritisk refleksjon. Foreløpig aksepterer vi premisset at KI som arbeidsverktøy skal utforskes, men det bør alltid ligge en kost-nytte-vurdering bak.

Bruken av KI må derfor styres av en åpen diskusjon om verdier og ansvar. Det gjelder både hvordan NOKUT bruker slike verktøy, og hvordan institusjoner og sakkyndige eventuelt tar dem i bruk. Samspillet mellom aktørene må være basert på åpenhet, forståelse av teknologiske begrensninger og respekt for faglig autonomi. I en fremtid der språkmodeller blir stadig mer sofistikerte, vil det være desto viktigere å beskytte fagligheten og sørge for at mennesket fortsatt har det siste ordet i vurderingsprosesser.

4 Konklusjon og anbefalinger

Bruken av *Notebook LM* i denne tematiske analysen har vist at språkmodeller kan gi treffsikre, reflekterte og tidvis overraskende svar, særlig når de får gode arbeidsbetingelser gjennom presise spørsmål og et godt strukturert datagrunnlag. Modellen evner å abstrahere fra enkeltformuleringer og trekke ut overordnede innsikter, men samtidig er den svært sensitiv for små variasjoner i spørsmålsformuleringen. Dette stiller høye krav til utformingen av *promptene*.

Et annet viktig funn er betydningen av et godt strukturert datagrunnlag. Modellen analyserer tekst i avgrensede informasjonsenheter, noe som innebærer at informasjon som er spredt over flere steder i datagrunnlaget eller som er uklart formulert, kan lettere gå tapt. Vår erfaring gjennom dette analysearbeidet har vist at kortere, tydeligere og mer målrettede tekster gir bedre resultater.

Under arbeidet ble det også tydelig at ensartet terminologi og konsekvent navngivning av filer er viktig for modellens arbeidsbetingelser. Inkonsekvente forkortelser og varierende filnavn skapte forvirring for modellen, noe som krevde manuell opprydding av oss. Dette illustrerer også hvordan bruk av KI-modeller kan virke disiplinerende på egen datastruktur og organisering.

Et interessant trekk ved *Notebook LM* er dens såkalte *emerging abilities*, altså uventede ferdigheter som kan komme til uttrykk i komplekse analyser. Dette kan være både nyttig og utfordrende, særlig i tematiske analyser av tilsynsrapporter, hvor helhetlig informasjon kan være fragmentert og tverrgående analyser er nødvendige. Her har *Notebook LM* vist potensial, men også noen tydelige fallgruver, som feiltolkning av institusjonsforkortelser, sammenblanding av informasjonskilder eller påstander og tolkninger som mennesker i saksbehandlerrollen ikke nødvendigvis gjør.

For at modeller som *Notebook LM* skal kunne brukes forsvarlig og effektivt i analyseprosessene, spiller saksbehandlere som har god kompetanse innen det som analyseres, derfor en avgjørende rolle gjennom kvalitetssikring og veiledning i analyseprosessen. Ettersom språkmodeller ikke forstår innholdet på samme måte som mennesker, men i stedet baserer seg på statistiske mønstre, er det nødvendig med faglig dømmekraft og kritisk vurdering.

Samtidig må det understrekes at kunstig intelligens er i en rivende utvikling. Selv i løpet av den perioden denne analysen er gjennomført (2025), har det skjedd betydelige fremskritt i både modellkapasitet, funksjonalitet og tilgjengelige verktøy. Det er derfor grunn til å forvente at mulighetene for bruk av KI i analysearbeid vil øke markant innen kort tid. Dette innebærer også at vi i NOKUT er åpne for å teste og ta i bruk andre modeller enn *NotebookLM* dersom disse viser seg å være mer hensiktsmessige. Bruken av KI bør derfor forankres i tydelige interne retningslinjer som regulerer når og hvordan slike verktøy kan og skal brukes, særlig med tanke på personvern, datagrunnlag, IT-sikkerhet, transparens, etterprøvbarhet og bærekraft, og at det gjøres risikoanalyser av aktuelle verktøy før de tas i bruk.

Avslutningsvis mener vi at KI-modeller har potensial til å styrke og effektivisere «tradisjonelle» analyser, men at dette forutsetter en bevisst og systematisk innsats i forberedelsene og gjennomføringen. Uten strategisk styring og kompetanseheving på dette området kan bruk av KI i analysearbeidet snarere bidra til uklarhet og feiltolkninger enn til kvalitativt bedre innsikter og beslutningsgrunnlag.

Referanser

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Davison, R.M., Chughtai, H., Nielsen, P., Marabelli, M., Iannacci, F., van Offenbeek, M., Tarafdar, M., Trenz, M., Techatassanasoontorn, A.A., Díaz Andrade, A. and Panteli, N. (2024). The ethics of using generative AI for qualitative data analysis. *Inf Syst J, 34*: 1433-1439.

Google Support (2025). https://support.google.com/notebooklm?sjid=12491595086394054000-EU#topic=16164070, 15. mai 2025.

Hitch D. (2024). Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future? *Qual Health Res. 2024 juni 34*(7):595-606.

IEEE Computer Society Team (2025). AI and Resource Consumption: Examining the Environmental Impact. Tilgjengelig på nett: https://www.computer.org/publications/tech-news/research/ai-and-the-environment?utm_source=chatgpt.com, 15. juni 2025.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2023). Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172*.

Nguyen-Trung, K. (2025). ChatGPT in thematic analysis: Can AI become a research assistant in qualitative research?. *Qual Quant* (2025).

NOKUT 2022. VEILEDNING. Systematisk kvalitetsarbeid i høyere utdanning: *Veiledning til kravene i universitets- og høyskoleloven, studiekvalitetsforskriften og studietilsynsforskriften* . https://www.nokut.no/siteassets/tilsyn-og-kvalitetsarbeid/nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022.pdf

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (s. 1-7).

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(4), e1507.

Zhou, C., He, J., Ma, X., Berg-Kirkpatrick, T., & Neubig, G. (2022). Prompt consistency for zero-shot task generalization. *arXiv preprint arXiv:2205.00049*.

Vedlegg

Vedlegg 1. Kravene til systematisk kvalitetsarbeid i runde tre

Paragraf	Lovtekst	Rettskilde
§ 1-6	Universiteter og høyskoler skal ha et tilfredsstillende internt system for kvalitetssikring som skal sikre og videreutvikle kvaliteten i utdanningen. Studentevalueringer skal inngå i systemet for kvalitetssikring.	*universitets- og høyskoleloven*
§ 4-3 (5)	Institusjonens arbeid med læringsmiljøet skal dokumenteres og inngå som en del av institusjonens interne system for kvalitetssikring etter § 1-6.	*universitets- og høyskoleloven*
§ 2-1 (2)	Institusjonene skal gjennomføre periodiske evalueringer av studietilbudene sine. Representanter fra arbeids- eller samfunnsliv, studenter og eksterne sakkyndige som er relevante for studietilbudet, skal bidra i evalueringene. Evalueringsresultatene skal være offentlige.	*studiekvalitets-forskriften*
§ 4-1 (1)	Institusjonens kvalitetsarbeid skal være forankret i en strategi og dekke alle vesentlige områder av betydning for kvaliteten på studentenes læringsutbytte.	*studietilsyns-forskriften*
§ 4-1 (2)	Kvalitetsarbeidet skal være forankret i institusjonens styre og ledelse på alle nivåer. Institusjonen skal gjennom kvalitetsarbeidet bidra til å fremme en kvalitetskultur blant ansatte og studenter.	*studietilsyns-forskriften*
§ 4-1 (3)	Institusjonen skal ha ordninger for systematisk å kontrollere at alle studietilbud tilfredsstiller kravene i forskrift om kvalitetssikring og kvalitetsutvikling i høyere utdanning og fagskoleutdanning § 3-1 til § 3-3 og kapittel 2 i denne forskrift.	*studietilsyns-forskriften*
§ 4-1 (4)	Institusjonen skal systematisk innhente informasjon fra relevante kilder for å kunne vurdere kvaliteten i alle studietilbud.	*studietilsyns-forskriften*
§ 4-1 (5)	Kunnskap fra kvalitetsarbeidet skal brukes til å utvikle kvaliteten i studietilbudene og avdekke eventuelt sviktende kvalitet. Sviktende kvalitet skal rettes opp innen rimelig tid.	*studietilsyns-forskriften*
§ 4-1 (6)	Resultater fra kvalitetsarbeidet skal inngå i kunnskapsgrunnlaget ved vurdering og strategisk utvikling av institusjonens samlede studieportefølje.	*studietilsyns-forskriften*

Vedlegg 2. Oversikt institusjoner i runde tre

Gruppe	Institusjon	Tilsynsperioden
Pilotprosjekt	Arkitektur- og designhøgskolen (AHO) MF vitenskapelige høyskole (MF) Norges Handelshøyskole (NHH) Norges musikkhøgskole (NMH) VID vitenskapelige høgskole (VID)	2017–2018
Gruppe 1	Høgskolen i Innlandet (HINN) Høgskulen på Vestlandet (HVL) Nord universitet (NORD) Universitetet i Sørøst-Norge (USN)	2018/2019
Gruppe 2	Norges miljø- og biovitenskapelige universitet (NMBU) OsloMet – storbyuniversitetet (OsloMet) Universitetet i Agder (UiA) Universitetet i Stavanger (UiS)	2019/2020
Gruppe 3	Høgskolen i Molde (HiMolde) Høgskulen i Volda (HVO) Høgskolen i Østfold (HiØ) Høyskolen Kristiania (HK) Samisk høgskole/Sámi allaskuvla (SH/SA)	2019/2020
Gruppe 4	Handelshøyskolen BI (BI) Kunsthøgskolen i Oslo (KHiO) Norges idrettshøgskole (NIH)	Høst 2020–vår 2021
Gruppe 5	Ansgar høyskole (AHS) Fjellhaug Internasjonale Høgskole (FIH) Lovisenberg diakonale høgskole (LDH) NLA Høgskolen (NLA)	Vår 2021–høst 2021
Gruppe 6	Dronning Mauds Minne Høgskole for barnehagelærerutdanning (DMMH) Forsvarets høgskole (FHS) Høyskolen for ledelse og teologi (HLT) Politihøgskolen (PHS)	Høst 2021–vår 2022
Gruppe 7	Norges teknisk-naturvitenskapelige universitet (NTNU) Universitetet i Bergen (UiB) Universitetet i Oslo (UiO) UiT – Norges Arktiske Universitet (UiT)	Vår 2022–høst 2022
Gruppe 8	Atlantis Medisinske Høgskole (AMH) Høyskolen Barratt Due (HBD) Bergen Arkitekthøgskole (BAS) Oslo Nye Høyskole (ONH) Høgskulen for grøn utvikling (HGUt) Høyskolen for yrkesfag (HØFY) Høyskolen for dansekunst (HFDK) Kriminalomsorgens høgskole og utdanningssenter KRUS (KRUS)	Høst 2022–høst 2024

Lillehammer Institute of Music Production and Industries (LIMPI)
Noroff University College (NUC)
Norsk barnebokinstitutt (NBI)
Norsk Gestaltinstitutt Høyskole (NGI)
Skrivekunstakademiet (SKA)
NSKI Høyskole (NSKI)
Steinerhøyskolen

Abstract

This report examines how a large language model, specifically *NotebookLM* developed by Google DeepMind, can be used as a support tool in analytical processes within external quality assurance of higher education. The empirical basis for the analysis consists of 48 review reports from the third cycle of periodic follow-up of higher education institutions conducted by the Norwegian Agency for Quality Assurance in Education (NOKUT). The analysis focuses in particular on recurring challenges and examples of good practice related to internal quality assurance systems and institutional quality work. By applying *NotebookLM* to the analysis of the reports, the study explores which challenges and practices the model identifies as most salient in institutional quality work. In comparison with a manual analysis, *NotebookLM* demonstrates a particular strength in structuring large volumes of material and identifying overarching patterns, while the manual analysis provides a more nuanced understanding of context and complexity. A combined approach, integrating AI-supported analysis with manual qualitative analysis, therefore appears to offer the most robust and comprehensive analytical framework. The findings indicate that *NotebookLM* has clear potential as an analytical support tool. At the same time, effective use of the model requires, among other factors, a well-structured data corpus, precise instructions (*prompts*), user competence in working with artificial intelligence, subject-matter expertise, and systematic human quality assurance.

DRAMMENSVEIEN 288 | POSTBOKS 578,1327 LYSAKER | T: 21 02 18 00 | NOKUT.NO