

Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2025 30 stp
Fakultetet for realfag og teknologi

Muligheter for store språkmodeller i saksbehandling i Oslo kommunes helse- og omsorgstjenester

Exploring the Potential of Large Language Models in
the Processing of Applications for Health and Care
Services in the City of Oslo

Oscar Christopher Husebye

Forord

Denne masteroppgaven har vært et utrolig spennende, og samtidig omfattende, prosjekt på grunn av den todelte metoden. Likevel er det undertegnede oppfatning at den brede kvalitativ-kvantitative tilnærmingen har styrket oppgaven og peilet den inn på et spor som er interessant for forskningen og kommunene selv.

Jeg ønsker å vie en stor takk til veileder Habib Ullah som har delt fra sin store kunnskap om blant annet LLM og ikke minst de mange informanter og respondenter som har bidratt og delt sitt syn gjennom intervjuer og spørreskjema. Refleksjonene og meningene fra disse har vært avgjørende for prosjektet og resultatene som presenteres i denne masteroppgaven.

Undertegnede er folkevalgt i Oslo bystyre, og ble etter masteroppgavens datainnsamling var fullført utnevnt til et vikariat som byrådssekretær i byrådsavdeling for finans. Dette er byrådsavdelingen som er ansvarlig for blant annet digitalisering i Oslo kommune. Denne jobben har ikke hatt innvirkning på masterprosessen og resultatene som presenteres, men det er et forhold jeg likevel ønsker å opplyse om.

Denne masteroppgaven ble som én av seks masteroppgaver tildelt KS - kommunesektorens organisasjon - sitt masterstipend for 2025. Jeg takker for at KS er opptatt av tematikken med språkmodeller i kommunal sektor og for det tildelte stipendet.

—

Oslo den 14. mai 2025,

—

Oscar Christopher Husebye

—

Masterstudent i industriell økonomi ved Norges miljø- og biovitenskapelige universitet

Sammendrag

Denne masteroppgaven undersøker hvorvidt store språkmodeller (LLM) kan forbedre saksbehandlingen av søknader om helse- og omsorgstjenester i Oslo kommune. Motivasjonen for studien er de økende utfordringene i norsk helsevesen knyttet til demografiske endringer, ressursknapphet og mangel på helsepersonell.

Studiens metode innebærer en todelt fremgangsmåte; en kvalitativ del og en kvantitativ del. Den kvalitative bygger på dybdeintervjuer med fem ledende fagfolk som er analysert med tematisk analyse. Den kvantitative delen baserer seg på spørreskjemadata fra 101 av kommunens saksbehandlere. Resultatene fra disse delene er diskutert opp mot relevant litteratur og problemstillingen.

Samlet sett viser funnene at det eksisterer anvendelser av store språkmodeller som kan forbedre saksbehandlingen i helse- og omsorgstjenesten i Oslo kommune. Det er identifisert et klart behov for bedre verktøy for teksthåndtering i saksbehandlingen, primært knyttet til informasjonsinnhenting og bistand til tekstproduksjon. Potensialet for tidsbesparelse ved bruk av LLM vurderes av saksbehandlerne til å ligge på mellom 31,5 % og 37,6 %, gitt et 95 % konfidensintervall på gjennomsnitt.

Bekymringer knyttet til graden av skjønnsvurderinger, transparens og bias er samtidig høyst tilstedeværende. Kommunens evne til å dele og styre etter data vurderes dessuten som relativt dårlig, som ikke er i tråd med politiske ambisjoner. Fagsystemet for helse- og omsorgssektoren er også så komplekst og utdatert at det binder opp teknisk kapasitet i kommunen som kunne blitt brukt til andre ting. Kunnskapsnivået og erfaringen med LLM blant saksbehandlere er på sin side dokumentert som svært variert, som både påvirker teknologiens potensial og dataen som denne masteroppgaven bygger sine funn på.

På kort sikt er det derfor vurdert at en fornuftig begynnelse på videre FoU- og implementeringsarbeid vil være å teste ut domenespesifikke språkmodeller i trygge og kontrollerte omgivelser, sammen med saksbehandlere. Anvendelser til oversettelse, klarspråkliggjøring, smartsøk og oppsummering av tekst er identifisert som anvendelser med relativt lav risiko knyttet til seg. Sammen med nøye oppfølging av nylansert infrastruktur for smartsøk og tekstanalyse i kommunens felles IKT-plattform kan dette medvirke til at Oslo kommune får hentet ut gevinster av bruk av store språkmodeller i saksbehandling i helse- og omsorgstjenesten.

Abstract

This master’s thesis investigates whether large language models (LLMs) can improve the processing of applications for health and care services in the City of Oslo. The motivation for the study stems from growing challenges in the Norwegian healthcare system related to demographic changes, resource constraints, and shortages of healthcare personnel.

The study applies a mixed-methods approach, consisting of a qualitative and a quantitative component. The qualitative part is based on in-depth interviews with five leading professionals and analyzed using thematic analysis. The quantitative part draws on survey data from 101 of the municipality’s caseworkers. The results from both parts are discussed in light of relevant literature and the overall research question.

Overall, the findings indicate that there are applicable use cases for LLMs that can enhance case processing in Oslo’s health and care services. A clear need has been identified for improved text-handling tools, primarily related to information retrieval and support for drafting text. According to the caseworkers, the potential time savings from using LLMs is estimated to be between 31.5% and 37.6%, based on a 95% confidence interval of the average.

Concerns related to the role of discretion, transparency, and bias remain significant. The municipality’s current ability to share and govern based on data is also considered relatively weak and misaligned with political ambitions. Furthermore, the existing case management system in the health and care sector is described as outdated and overly complex, tying up technical capacity that could be used elsewhere. The level of knowledge and experience with LLMs among caseworkers is also shown to vary significantly, which affects both the potential for successful implementation and the reliability of the data on which this thesis is based.

In the short term, a prudent starting point for further R&D and implementation efforts is to test domain-specific LLMs in safe and controlled environments in collaboration with caseworkers. Use cases such as translation, plain language generation, smart search, and text summarization are identified as relatively low-risk applications. Alongside close follow-up of the newly launched infrastructure for smart search and text analysis within the municipality’s common IT platform, these efforts may help Oslo harness the benefits of large language models in the processing of health and care service applications.

Innhold

1	Introduksjon	1
1.1	Problemstilling og forskningsspørsmål	2
1.2	Studiens oppbygging	2
2	Teori	4
2.1	Overordnet begreps- og teknologiforklaring	4
2.2	Status og utvikling	6
2.2.1	Teknologiens utvikling	6
2.2.2	Status i Norge	7
2.3	Politiske krav og målsetninger om KI	10
2.4	Overordnet potensial i helsesektoren	12
2.5	Metaanalyser	15
2.5.1	Vitenskapelig behov	16
2.5.2	Overordnede funn	16
2.6	Utvalgte empiriske studier	18
2.6.1	LLM i diagnostisering	18
2.6.2	LLM i pasientkommunikasjon	19
2.6.3	LLM som verktøy for kompetanseutjevning	20
2.7	Oslo kommunes organisering	22
2.8	KI-assistert saksbehandling i Oslo kommune	23
2.9	Oppsummering av teori	25
3	Metode	27
3.1	Forskningsdesign og metodisk rammeverk	27
3.1.1	Forskningsdesign	27

3.1.2	Metodisk rammeverk - CRISP-DM	27
3.1.3	Refleksjon om metodevalg	28
3.2	Datainnsamling	29
3.2.1	Kvalitativ del	30
3.2.2	Kvantitativ del	31
3.3	Databehandling	33
3.4	Modellering	34
3.5	Evaluerings	36
3.5.1	Studiens reliabilitet	36
3.5.2	Studiens validitet	37
3.6	Store språkmodeller som hjelpemiddel	38
4	Resultater	40
4.1	Kvalitativ del	40
4.1.1	Kommunens potensial	40
4.1.2	Teknologiens potensial	49
4.2	Kvantitativ del	56
4.2.1	Bruk av og kjennskap til LLM	57
4.2.2	Utfordringer i eksisterende saksbehandling	58
4.2.3	Forventninger og bekymringer om LLM i saksbehandling	62
4.2.4	Implementerings- og gevinstvurderinger	65
5	Diskusjon	71
5.1	Innledende betraktninger	71
5.2	Hovedfunn	73
5.2.1	Variert kompetanse og bruk	73

5.2.2	Digitalisering med forbedringspotensial	74
5.2.3	Teksthåndteringspotensial i saksbehandlingen	75
5.2.4	Gevinstvurderinger	76
5.2.5	Skjønn, bias og transparens som barrierer	77
5.2.6	Er LLM riktig teknologi?	78
5.3	Implementering	78
5.4	Konklusjon	80
Referanser		82
A Øvrige innspill om tematikken fra respondenter		87
B Elektronisk spørreskjema (utskrift)		89

Figurer

1	Grader av og metoder for spesialisering av språkmodeller. Hentet fra Ling mfl., 2024.	6
2	Stilisert illustrasjon av utviklingen i økt regnekraft og algoritmisk forbedring for LLM (Ho mfl., 2024).	7
3	Bruk av KI i offentlig sektor 2021-24 (Rambøll mfl., 2024)	8
4	Bruksområder for KI i offentlig sektor de neste tre årene 2021-24 (Rambøll mfl., 2024).	9
5	Barrierer mot å ta i bruk KI i offentlig sektor (Rambøll mfl., 2024). .	10
6	Oslo kommunes organisasjonskart (Oslo kommune, u.å.-d).	22
7	Konsept for datainnsamling.	30
8	Respondentenes prosentvise bruk, kjennskap og holdninger til KI. Likert-skala 1-5 gruppert til «uenig», «nøytral» og «enig» per påstand. 57	
9	Respondentens vurdering av positivt bidrag (forbedring) fra eksisterende digitale støtteverktøy/-systemer. Prosentfordeling av Likert-skala (1-5).	58
10	Respondentenes vurdering av utfordringer i dagens saksbehandling, målt etter andel av respondenter som peker på hver utfordring. . . .	60
11	Respondentenes prosentvise rangering av saksbehandlingsoppgaver etter hvor tidkrevende de er. Rangering 1 = lengst tid, rangering 6 = minst tid.	61
12	Andel respondenter med bekymringer om KI i saksbehandling per bydel. Stiplet linje er gjennomsnittandel (72 prosent).	62
13	Respondentenes forventninger til bruk av KI i saksbehandlingen, målt etter andel av respondenter som peker på hvert forhold. Likert-skala 1-5 gruppert til «uenig», «nøytral» og «enig» per påstand.	64
14	Respondentenes vurdering av hvor i saksbehandlingen KI kan implementeres, flervalg.	66
15	Histogram over estimert prosentvis tidsbesparelse ved bruk av KI til det respondentene mener teknologien kan brukes til i saksbehandlingen. 67	

16	Respondentenes vurdering av implementeringsbarrierer mot KI i saks- behandling, flervalg.	70
----	--	----

Tabeller

1	Utdrag fra tabell om «Perceived likelihood that large language models will support health care tasks» (Denecke mfl., 2024).	14
2	Utdrag fra tabell om «Participants' perceptions on the general benefits of using large language models in health care» (Denecke mfl., 2024).	14
3	Utdrag fra tabell om «Shortcomings and risks of large language model (LLM)-based systems in health care» (Denecke mfl., 2024).	15
4	Utdrag fra tabell om «Summary Statistics for the Sample of Customer-Service Agents» (Brynjolfsson mfl., 2025).	21
5	Oversikt over informanter, organisatorisk tilhørighet og relevant ansvarsområde.	31
6	Respondentkarakteristika fra gyldige svar i spørreundersøkelse, samlet og per bydel.	32
7	Beregning av nødvendig antall svar for å gi statistisk signifikant flertall i populasjonen med konfidensintervall på 95 %.	37
8	Tematisk inndeling av funn fra ekspertintervjuer.	40
9	Respondentenes bruk, kjennskap og holdninger til KI. Gjennomsnitt, median og SD av Likert-skala 1-5 per påstand (n=101).	58
10	Respondentenes vurdering av utfordringer i dagens saksbehandling (n=101).	59
11	Respondentenes rangering av saksbehandlingsoppgaver etter hvor tidkrevende de er. Rangering 1 = lengst tid, rangering 6 = minst tid (n=101).	61
12	Respondentenes bekymringer om bruk av KI i saksbehandlingen, flervalg (n=73).	63
13	Respondentenes forventninger til bruk av KI i saksbehandling. Gjennomsnitt, median og SD av Likert-skala 1-5 per påstand (n=101).	65
14	Statistiske mål for estimert prosentvis tidsbesparelse og ekvivalent kostnadsbesparelse ved bruk av KI til det respondentene mener teknologien kan brukes til i saksbehandlingen (n=101).	68

15	Saksbehandling som respondenter mener kan helautomatiseres (n=36).	69
A.1	Øvrige fritekstsvar fra saksbehandlere om LLM i saksbehandling. . .	87

1 Introduksjon

Norsk helsevesen står overfor betydelige utfordringer de kommende årene. Sykeligheten i Norge øker, samtidig som den demografiske utviklingen medfører at stadig flere blir eldre og lever lange liv (SSB, 2020, 2024a). Dette gjør at etterspørselen etter helse- og omsorgstjenester vokser kraftig og legger press på det statlige og kommunale helse- og omsorgstilbudet. Helsepersonellkommisjonen slår fast at veksten i eldre vil gi «utfordringer uten sidestykke for de kommunale omsorgstjenestene» (Regjeringen, 2023). I flere kommuner er i tillegg den økonomiske situasjonen svært presset, og realitetene gjør det nødvendig å gjøre tiltak som sikrer et bærekraftig helse- og omsorgstilbud (KS, 2024).

Samtidig er det store utfordringer med å ha tilstrekkelig med kvalifisert helsepersonell til å utføre de viktige tjenestene (Regjeringen, 2023). Mangelen er allerede stor, og en utfordring er at mye av helsepersonellens tid går til administrative, ikke-pasientrettede oppgaver, som saksbehandling og journalføring. Automatisering og effektivisering av disse oppgavene kan frigjøre tid til pasientrettet arbeid. Litteraturgjennomgangen i denne studien dokumenterer at kunstig intelligens (KI) og store språkmodeller (Large Language Models, LLM) har vist lovende resultater innen tekstforståelse, datahåndtering og beslutningsstøtte i ulike sektorer, inkludert helsesektoren, de siste årene. Spørsmålet er imidlertid om teknologien er moden, og hvorvidt den kan anvendes i praksis.

For å undersøke tematikken søker denne masteroppgaven å kartlegge hvorvidt anvendelser av LLM kan bidra til forbedret saksbehandling av søknader om kommunale helse- og omsorgstjenester. Oppgaven tar for seg mulighetene for LLM i Oslo kommune, som både er landets største kommune og en kommune med kompleks, desentralisert organisering på helse- og omsorgsfeltet. Ved å avgrense oppgaven til én spesifikk kommune er målet å kunne svare presist og grundig på problemstillingen, samtidig som resultatene må antas å ha nytteverdi også for andre kommuner.

For Oslos vedkommende er det allerede gjennomført en intern «miniutredning» om kunstig intelligens-assistert saksbehandling (KIAS) våren 2024 (Undeland mfl., 2024). Denne skisserte et betydelig gevinstpotensial ved å anvende LLM i saksbehandlingen. Samtidig er det mange spørsmål som står ubesvart, og det er nødvendig med forskning som tar for seg konkret implementering av teknologien. Er Oslo kommune klar for å ta i bruk kraften som ligger i LLM, og bør kommunen i det hele tatt etterstrebe det?

Dette prosjektet er gjennomført med en kombinert kvalitativ-kvantitativ tilnærming som inkluderer både ekspertintervjuer og en bredere datainnsamling fra kommunens saksbehandlere. Det overordnede metodiske rammeverket for arbeidet er industristandarden Cross Industry Standard Process for Data Mining (CRISP-DM). Inten-

sjonen er at masteroppgaven skal kunne bidra som et utvidet kunnskapsgrunnlag for norske kommuner, og gi svar på hvilke praktiske muligheter og begrensninger som ligger i LLM-teknologien.

1.1 Problemstilling og forskningsspørsmål

Oppgavens problemstilling er følgende:

Finnes det anvendelser av store språkmodeller (LLM) som kan forbedre saksbehandlingen av søknader om helse- og omsorgstjenester i Oslo kommune?

Med tilhørende to forskningsspørsmål (FS):

1. Hvilket potensial er det kommunens ledende fagpersoner identifiserer for LLM i saksbehandlingen?
2. Hvilke utfordringer finnes i dagens saksbehandling og hva kan LLM bidra med ifølge saksbehandlerne selv?

I denne masteroppgaven forstås *potensial* i FS1 som de bruksområder, gevinster og forutsetninger fagpersonene peker på for å eventuelt kunne ta i bruk LLM-teknologi i saksbehandlingen. *Utfordringer* i FS2 forstås som det saksbehandlerne opplever som krevende i dagens saksbehandling generelt. Begrepet *bidra* i FS2 forstås som hva saksbehandlerne mener at LLM kan anvendes til, og hvilke barrierer og fortrinn som eksisterer i dagens saksbehandlingsprosess i forbindelse med dette.

1.2 Studiens oppbygging

Masteroppgaven er lagt opp etter IMRAD-strukturen, med et tillegg for teorikapittelet. IMRAD er et vanlig oppsett på naturvitenskapelige studier, og står for introduksjon, metode, resultater og diskusjon (Hem, 2025). Teorikapittelet er skilt ut fra introduksjonen for å tydeliggjøre dette som en selvstendig og vesentlig del av oppgaven. Kapittelet forklarer sentrale begrep og gir en overordnet innføring i teknologien, LLMs teknologiske utvikling og utbredelse i Norge blir beskrevet, politiske krav og målsetninger blir redegjort for og relevant litteratur blir gjennomgått. I gjennomgangen er det særlig fokus på tilgjengelige metaanalyser og noen utvalgte empiriske studier. I tillegg blir Oslo kommunes organisering og tidligere arbeid med LLM i saksbehandlingen redegjort for i teorikapittelets siste del, før en kort oppsummering av teorien finner sted.

Deretter beskrives metoden. Først blir overordnet forskningsdesign og det metodiske rammeverket CRISP-DM beskrevet og begrunnet. Etter dette blir studiens to tilnærminger, kvalitativ og kvantitativ, beskrevet ved hjelp av fire av CRISP-DMs hovedfaser; datainnsamling, databehandling, modellering og evaluering. I tillegg redegjøres det kort for bruken av LLM som hjelpemiddel i masterprosessen.

Resultatene presenteres todelt basert på forskningsspørsmål 1 (kvalitativ) og 2 (kvantitativ). Funnene fra den kvalitative tematiske analysen i del 1 presenteres først, og de identifiserte temaene/funnene er delt i to overordnede temaer som er «kommunes potensial» og «teknologiens potensial». Den andre delen av resultatkapittelet er den kvantitative delen, der funnene fra spørreundersøkelsen til saksbehandlere presenteres. Funnene i denne delen legges frem i fire seksjoner.

Til sist i dette dokumentet er diskusjonskapittelet, der resultatene på de to forskningsspørsmålene sammenlignes og knyttes til teori og problemstillingen. Dette kapittelet begynner med noen innledende betraktninger, før hovedfunn og implementering blir diskutert. Til sist konkluderes diskusjonen sett opp mot masteroppgavens problemstilling.

Etter dette kommer referanser og vedlegg.

2 Teori

2.1 Overordnet begreps- og teknologiforklaring

Kunstig intelligens (KI) eller artificial intelligence (AI) er et teknologisk fagfelt som omfatter metoder og systemer som tilsynelatende kan utføre oppgaver som tidligere har krevet menneskelig intelligens. Fagfeltet KI plasseres som en sammensmeltning av datateknikk, logikk, matematikk, psykologi og nevrovitenskap (Tidemann & Arnesen, 2025). Begrepet har eksistert siden 50-tallet, men det er de senere årene at teknologien har fått stor utbredelse (Tidemann & Arnesen, 2025).

En sentral underkategori av KI er **maskinlæring (ML)**. Maskinlæring involverer statistiske metoder der datasystemer, uten eksplisitt programmering, forbedrer ytelsen sin basert på erfaring (Elster & Tidemann, 2025). Erfaring i denne sammenhengen er store datasett. Tidemand og Arnesen (2025) trekker et grovt skille mellom regelbaserte KI-modeller, som de kaller «ekspertmodeller», og datadrevne modeller, som er modeller med maskinlæring. De regelbaserte modellene tolker begreper med regler som er programmert inn før modellen brukes, mens de datadrevne modellene «lærer selv» (Tidemann & Arnesen, 2025).

Generativ KI er et begrep for modeller som ikke kun analyserer data, men også genererer nytt innhold selv. Dette innholdet kan være tekst, bilder eller annen data. Det nye innholdet blir generert basert på mønstre i modellens treningsdata, og kvaliteten på denne har derfor betydning for kvaliteten på den genererte dataen. Ikke-generative modeller analyserer inputdata for eksempel for å kategorisere denne, mens generative modeller krever langt større mengder inputdata og regnekraft for å produsere nytt innhold (Helsedirektoratet, 2024).

Store språkmodeller eller Large Language Models (LLM), er store, generative KI-modeller som er spesialisert på å tolke og produsere ny tekst (Microsoft, u. å.). Modellene er trent opp ved bruk av maskinlæringsmetoder, og gjenkjenner mønstre i språk basert på kontekst og store mengder skriftlig treningsdata (Denecke mfl., 2024; Machot mfl., 2024; Microsoft, u. å.). Selv om LLM slik vi kjenner det idag er relativt nytt, har arbeidet med naturlig språkprosessering (NLP) foregått i lang tid. På 80- og 90-tallet gikk man fra logikkeksperimenter til en mer datadrevet tilnærming med statistikk for å forutsi hvilke ord som passet basert på en foregående tekst (Microsoft, u. å.).

Et gjennombrudd kom i 2018 med studien «Attention Is All You Need», skrevet av forskere i Google (Microsoft, u. å.). Her ble transformer-arkitekturen introdusert, som mange moderne LLMer baserer seg på, ifølge Denecke mfl. (2024). Microsoft beskriver transformer-arkitekturen som «et innovativt nevralt nettverksrammeverk

som kunne administrere og forstå kompleks tekstinformasjon med større nøyaktighet og skalering» (Microsoft, u. å.). Et hovedskille fra tidligere arkitektur er bruken av *Self attention*-mekanismer, som gjør det mulig for modellen å vurdere hvilke ord i en tekstsekvens som er mest relevant. Mekanismene gjør ifølge Microsoft (u.å.) at «modellen kan sikte seg inn mot forskjellige deler av en tekstsekvens og dynamisk veie verdien av informasjon i forhold til andre tokenener i sekvensen, uavhengig av posisjonen deres».

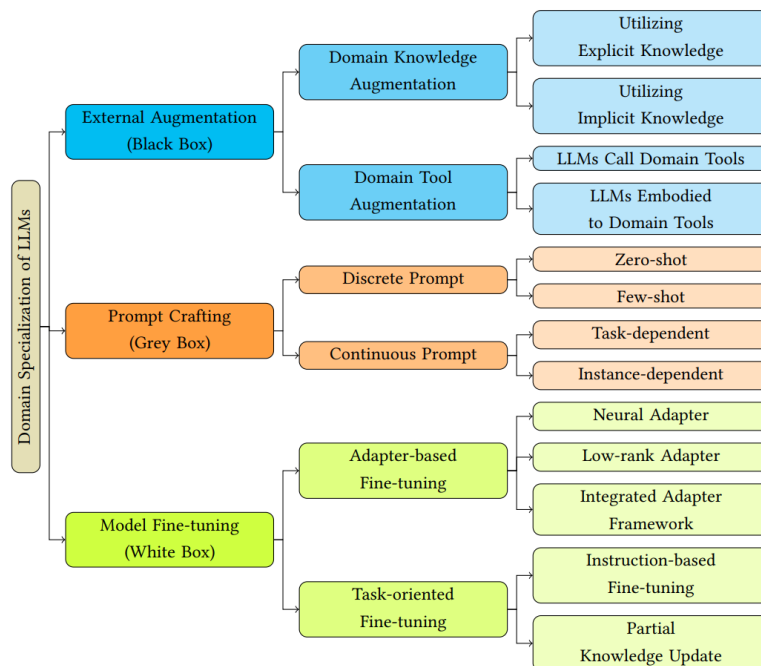
Det mest kjente eksempelet på LLM er Generative Pre-trained Transformer-modellen (GPT) som ligger til grunn for blant annet *ChatGPT* og *Copilot*, men det finnes andre store språkmodeller som franske *Mistral*, tyske *Aleph Alpha* og *Gemini* og *LLaMA* fra amerikanske Google og Meta, i tillegg til kinesiske *DeepSeek* (Helsedirektoratet, 2024; Hovland, 2024).

Teknologirådet, som er Stortingets og regjeringens rådgiver i teknologispørsmål, brukte begrepet store språkmodeller som en samlebetegnelse for modeller som *ChatGPT* og *Gemini* i sin rapport *Generativ kunstig intelligens i Norge* (2024). Rådet understreket samtidig at «begrepsbruken varierer og er ikke entydig» (Teknologirådet, 2024). De poengterer at det er ulik omtale av de nye modellene i ulike miljøer, og rådet viser spesifikt til tre definisjoner av LLM i rapporten:

- Stanford University har innført begrepet «grunnmodell» (på engelsk Foundation model) som viser til at modellene blir så store og anvendelige at de inngår som grunnfundament i mange KI-systemer (Bommasani & Liang, 2021).
- I Joe Bidens presidentordre for kunstig intelligens omtales «grunnmodeller til flerbruk» (på engelsk Dual-Use Foundation Model) (Executive Office of the President, 2023).
- I den nye KI-forordningen, *AI Act*, har EU valgt begrepet «KI-systemer med generelt formål» (på engelsk General purpose AI system) (Teknologirådet, 2024).

Teknologirådet argumenterer for at de tre begrepene refererer til stort sett samme KI-modeller, som kjennetegnes ved at de «er store, kraftige og anvendelige» (Teknologirådet, 2024).

Variasjonen blant de store språkmodellene er samtidig omfattende, og det går et skille mellom generelle grunnmodeller og spesialiserte modeller. Ling mfl. (2024) omtaler dette som **domenespesifikke språkmodeller**. Å anvende grunnmodeller direkte på spesialiserte oppgaver og fagfelt kan medføre en rekke utfordringer på grunn av de unike domenespesifikke forholdene som gjør seg gjeldende (Ling mfl., 2024). Ling mfl. (2024) har sammenfattet og kategorisert hvilke grader av domenespesialisering



Figur 1: Grader av og metoder for spesialisering av språkmodeller. Hentet fra Ling mfl., 2024.

som finnes. Det er tre grader av domenespesialisering som kan oppnås med ulike tilganger på den underliggende grunnmodellen, jamfør figur 1. Graden av spesialisering går fra ingen tilgang til modellen (utover å kunne interagere med eksempelvis ChatGPT), omtalt som «Black Box», via delvis tilgang til modellens indre omtalt som «Grey Box», til full tilgang til modellen og mulighet for å finjustere den omtalt som «White Box». De ulike tilgangsnivåene resulterer altså i forskjellig potensial for spesialisering av modellene til konkrete domener og applikasjoner (Ling mfl., 2024).

2.2 Status og utvikling

2.2.1 Teknologiens utvikling

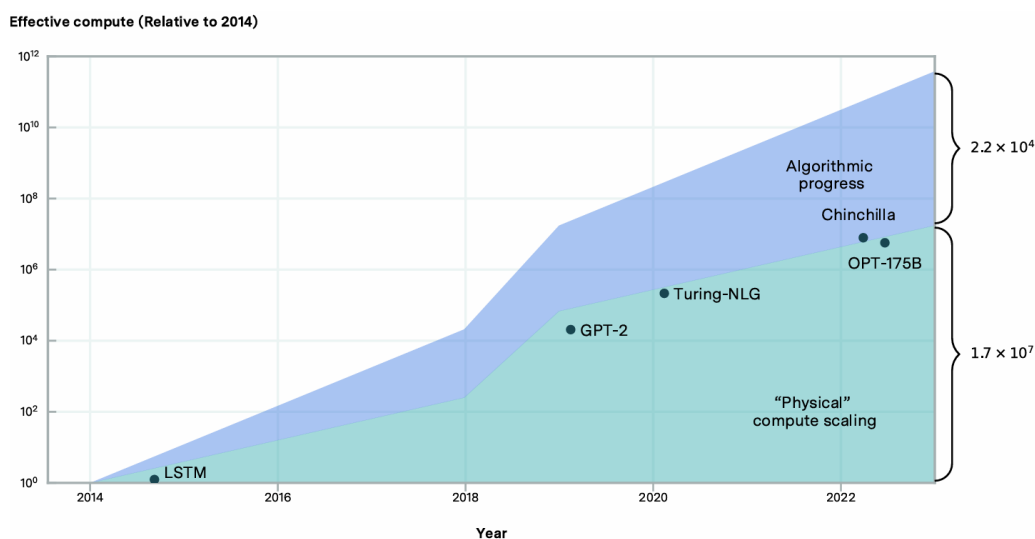
LLM-teknologien har utviklet seg i et voldsomt tempo de siste årene, og modellenes ytelse forbedres stadig. Flere studier undersøker enkeltmodellens ytelse og egnethet til spesifikke oppdrag, og det er utviklet en rekke rammeverk for å «benchmark» ulike modeller opp mot hverandre. Disse tester modellene i språkforståelse, logisk resonnering, fagspesifikke spørsmål og mye annet (Banerjee mfl., 2023).

Imidlertid er det relevant å si noe mer overordnet om den teknologiske utviklingen til LLM. Ho mfl., 2024 har kvantifisert den teknologiske fremgangen innen LLM det

siste tiåret. Forskerne gjorde en omfattende metaanalyse av over 200 evalueringer av ulike LLMer fra 2012 til 2023 for å vurdere modellenes forbedring. Hovedfokuset var forbedringen i algoritmisk effektivitet, altså hvordan modellene utviklet seg om man ser bort fra økt datakapasitet. Med andre ord; har vi fått bedre modeller for datakraften vi bruker?

Svaret er at den algoritmiske effektiviteten har økt dramatisk siden 2012 (Ho mfl., 2024). For å oppnå en gitt ytelse er den nødvendige datakraften halvert omtrent hver åttende måned, som medfører en eksponentiell forbedring i effektivitet. LLMs fremskritt skyldes altså ikke kun større modeller og mer data, men også betydelig algoritmisk innovasjon. Her trekkes introduksjonen av transformer-arkitekturen spesielt frem av forskerne (Ho mfl., 2024).

Den algoritmiske fremgangen, kombinert med økt datakapasitet, gir utslag i eksponentielt bedre regnekraft og LLM-modeller (Ho mfl., 2024). Figur 2 oppsummerer teknologiens utvikling og gir et godt empirisk grunnlag for å si at LLM-teknologien generelt har opplevd en voldsom forbedring de siste årene. Legg merke til at skalaen er logaritmisk.



Figur 2: Stilisert illustrasjon av utviklingen i økt regnekraft og algoritmisk forbedring for LLM (Ho mfl., 2024).

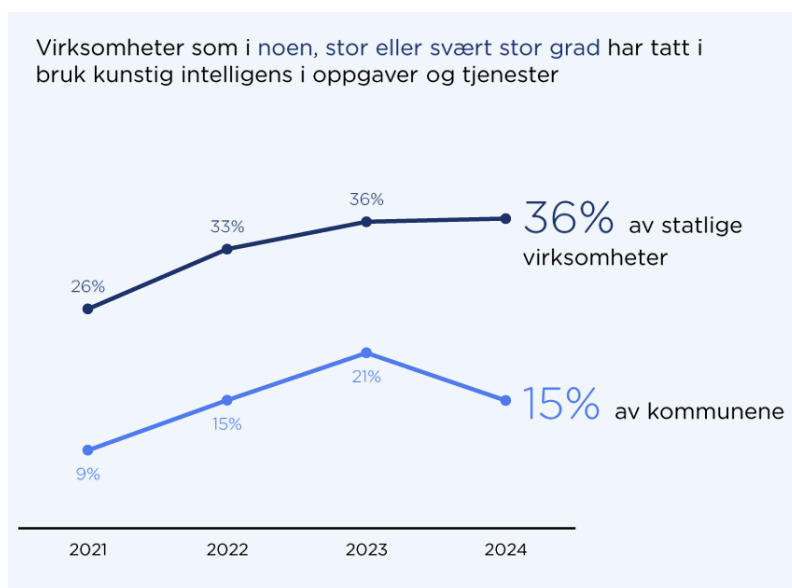
2.2.2 Status i Norge

Bruken av KI i Norge har økt betraktelig de siste årene. SSB rapporterer årlig på bruk av IKT i næringslivet, og siste rapport fra september 2024 viste en doubling i bruk av KI hos norske virksomheter sammenlignet med året før (SSB, 2024b).

Bruk av generativ KI trekkes frem som hovedgrunnen for den store økningen. På landsbasis brukte i 2024 to av ti foretak KI i en eller annen form. Bruken i Oslo er klart høyere enn resten av landet, og halvparten av alle hovedstadens foretak brukte KI i 2024 ifølge SSB (SSB, 2024b).

Ser man mer spesifikt på bruken i offentlig sektor, blir denne grundig vurdert i rapporten *IT i praksis* hvert år. Rapporten blir utarbeidet av Rambøll i samarbeid med Digitaliseringsdirektoratet, IKT Norge, KS (kommunesektorens organisasjon) og NTNU, og er omtalt som den mest omfattende undersøkelsen av sitt slag i Norge (Rambøll mfl., 2024). Den siste undersøkelsen ble gjennomført blant alle statlige virksomheter og kommuner mellom mars og mai 2024.

Statistikken viser at den store implementeringen av KI lar vente på seg i offentlig sektor (Rambøll mfl., 2024). Figur 3 viser andelen offentlige virksomheter som bruker KI. Rambøll mfl. understreker at nesten ingen (kun 1 prosent) av virksomhetene svarte at de brukte teknologien «i stor eller svært stor grad». Kommunene bruker KI i vesentlig mindre grad enn staten, og andelen blant disse falt seks prosentpoeng fra 2023 til 2024 (Rambøll mfl., 2024).



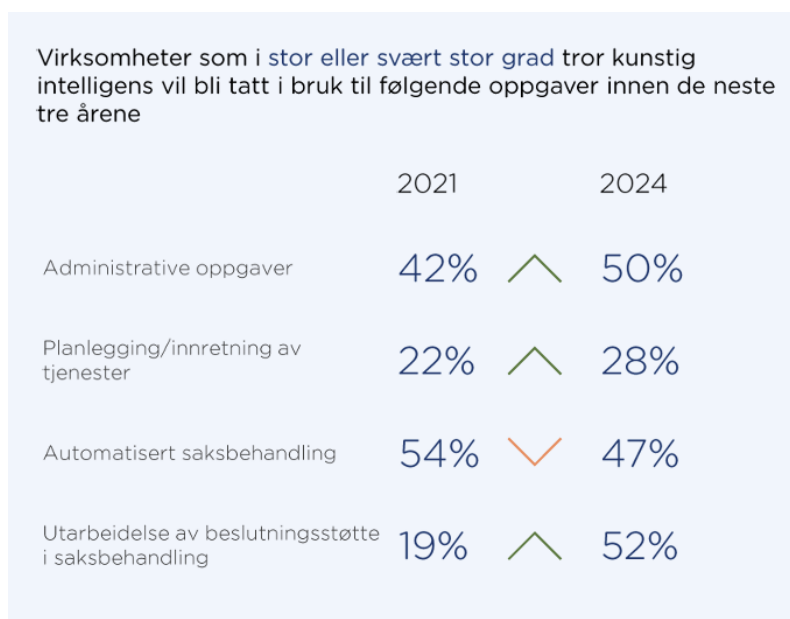
Figur 3: Bruk av KI i offentlig sektor 2021-24 (Rambøll mfl., 2024)

Til tross for nedgangen i bruk fra kommunene, viser *IT i praksis* at offentlige virksomheter i stadig større grad tror KI vil bli tatt i bruk fremover. Rapporten tar for seg bruk av KI i følgende anvendelser (Rambøll mfl., 2024):

- Administrative oppgaver

- Planlegging/innretting av tjenester
- Automatisert saksbehandling
- Utarbeidelse av beslutningsstøtte i saksbehandling

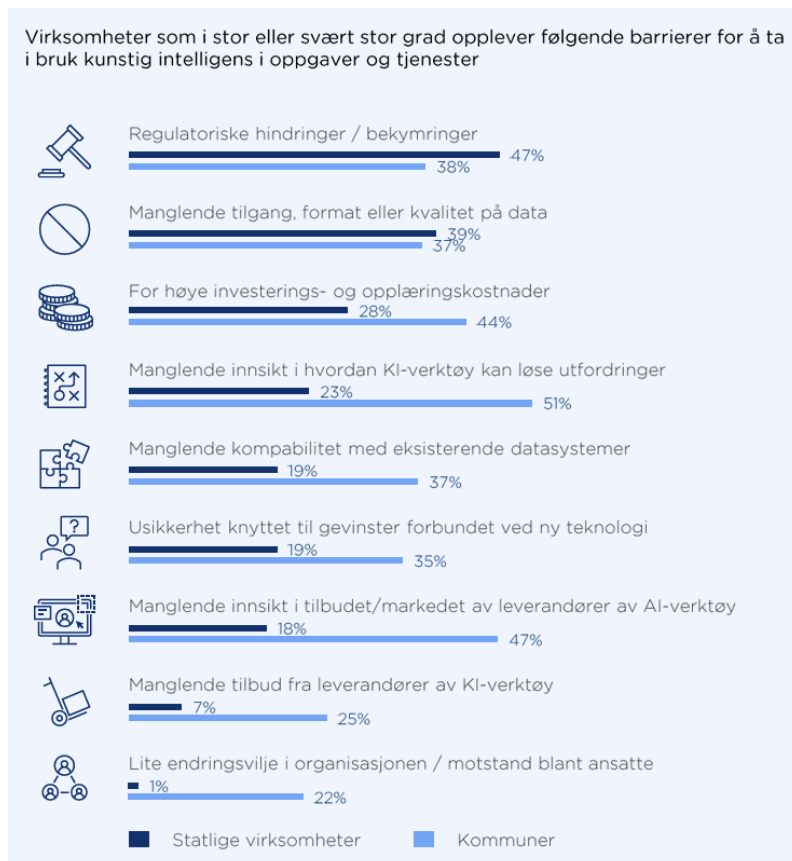
Undersøkelsen viser at potensialet for implementering i samtlige av disse anvendelsene øker, bortsett fra automatisert saksbehandling (Rambøll mfl., 2024). Der har virksomhetene redusert sitt syn på bruk av KI siden 2021. Rambøll mfl. (2024) trekker frem at dette kan skyldes at en del slike prosesser allerede er blitt automatisert, samtidig som det også kan indikere at flere virksomheter ser begrensningene i KI (Khan mfl., 2022), spesielt når det kommer til transparens. Forfatterne understreker likevel at det fremdeles er en vesentlig andel som mener KI kan bidra til mer automatisert saksbehandling. Figur 4 viser de konkrete bruksområdene som er vurdert, og hvor stor andel av virksomhetene som tror KI vil bli tatt i bruk der innen tre år.



Figur 4: Bruksområder for KI i offentlig sektor de neste tre årene 2021-24 (Rambøll mfl., 2024).

I den aktuelle versjonen av *IT i praksis* fra 2024 undersøkes det for første gang også hvilke barrierer som eksisterer for å ta i bruk KI i tjenester og oppgaver i offentlig sektor. Dette er relevant for hvorvidt kommunene kan ta i bruk KI i praksis. For, som rapporten poengterer; «Å ta i bruk kunstig intelligens handler om mye mer enn bare teknologi» (Rambøll mfl., 2024). Overordnet konkluderer Rambøll mfl. (2024) med at det er til dels store forskjeller mellom kommunene og staten i dette spørsmålet. Kommunene vurderer for eksempel at de mangler innsikt i hvordan KI-verktøy kan løse utfordringer i nesten dobbelt så stor grad som statlige virksomheter. Dette

understreker behovet for forsknings- og utviklingsarbeid rettet mot KI i kommunal sektor. Figur 5 viser de ulike barrierene som offentlig sektor møter i forbindelse med implementering av KI (Rambøll mfl., 2024).



Figur 5: Barrierer mot å ta i bruk KI i offentlig sektor (Rambøll mfl., 2024).

Rambøll mfl. (2024) fant også at få offentlige virksomheter prioriterte KI over andre IT-investeringer, og at kun 11 prosent av kommunene har planer for implementering av KI-baserte løsninger klare eller under utvikling. Andelen i staten var 45 prosent (Rambøll mfl., 2024).

2.3 Politiske krav og målsetninger om KI

Det er de siste årene kommet flere føringer, målsetninger og uttalelser fra myndighetene spesifikt om bruk av kunstig intelligens. Disse er viktige å ha med seg når konkrete anvendelser skal undersøkes. Både EU, nasjonale myndigheter og Oslo kommune er på banen for å utforme veiledninger, lovverk og strategier som skal legge til rette for fornuftig bruk av teknologien. Et fellestrekk er at politikerne ser et stort potensial for effektivisering og forbedring i både offentlig og privat sektor ved bruk av KI, samtidig som man er opptatt av personvern, menneskerettigheter

og demokratisk kontroll. Nasjonal strategi for kunstig intelligens ble lagt frem av den daværende regjeringen i 2020, og har formulert en visjon om at «[...] Norge skal gå foran i utvikling og bruk av kunstig intelligens med respekt for den enkeltes rettigheter og friheter» (Regjeringen, 2020).

Regjeringen fastslo noen kjerneprinsipper som all KI-utvikling og -bruk i Norge skal bygge på. I strategien heter det at (Regjeringen, 2020):

- KI skal bygge på etiske prinsipper og respektere menneskerettigheter og demokrati
- Forskning, utvikling og bruk av KI skal bidra til ansvarlig og pålitelig KI
- KI-løsninger skal ivareta personvern og individets integritet
- Digital sikkerhet skal integreres fra designfasen ved utvikling, drift og forvaltning av KI-systemer
- Tilsynsmyndigheter må føre kontroll med at KI-systemer innen deres sektor følger prinsippene for ansvarlig og pålitelig bruk

I 2024 gikk sittende digitaliseringsminister Karianne Tung ut med en målsetning om at offentlig sektor skal bruke KI i 80 prosent av sin virksomhet i løpet av 2025 (NRK, 2024). I NRK-saken vises det spesifikt til KI-anvendelser i saksbehandling og i helsevesenet. Målsetning for 2025 er en betraktelig økning fra andelen som i 2024 kun var 25 prosent, og ministeren ble møtt med kritikk fra postdoktor Lisa Reutter som mente det ikke er realistisk å øke bruken til 80 prosent på en forsvarlig, fornuftig og demokratisk måte i løpet av 2025 (NRK, 2024).

Politiske målsetninger rundt KI synes generelt å være noe i utakt med eksperters vurdering av realistisk implementering (Spilker, 2024). Spilker (2024) viser til tidligere forskning og argumenterer for at «Et stort diskursfellestrekk i tidligere forskning er nettopp de store kontrastene mellom de politiske forestillingene om digitalisering som velferdssamfunnets redning og den praktiske implementeringen ute i virksomhetene».

Det er ikke denne masteroppgavens intensjon å gå nærmere inn på de juridiske sidene ved LLM og KI, men det nevnes at EUs forordning for KI, *AI Act*, ble vedtatt av EU juni 2024 og høyst sannsynlig blir implementert i norsk rett som følge av EØS-avtalen (Regjeringen, 2024).

I påvente av innføringen av *AI Act* har relevante myndigheter utarbeidet veiledninger for ansvarlig bruk av KI i Norge. Blant annet har Digitaliseringsdirektoratet laget en *Veiledning for ansvarlig bruk og utvikling av KI*. Denne gir råd om hvordan offentlige og private virksomheter kan ivareta personvern, sikkerhet og etikk når de tar i bruk

KI-teknologi (Digdir, 2024). Datatilsynet har på sin side utviklet en *regulatorisk sandkasse* for å veilede og teste ut konkrete anvendelser av blant annet KI opp mot regelverket (Datatilsynet, 2023).

I Oslo kommune ble *Strategi for kunnskapsbasert utvikling, innovasjon og digital teknologi* vedtatt av bystyret mars 2023 (Oslo kommune, 2023a). Her legges det til grunn at kommunen skal være «en kunnskapsbasert, innovativ og datadrevet kommune» og skal «levere brukervennlige tjenester og finne nye løsninger på samfunnsutfordringer» (Oslo kommune, 2023b). Enkeltteknologier nevnes i liten grad, men strategien nevner eksplisitt at KI skal tas i bruk og eksperimenteres med, slik at kommunen sikrer bedre utnyttelse av dataene den har.

Under innsatsområdet «Kontinuerlig tjenesteutvikling» i strategien slår kommunen fast to relevante forhold (Oslo kommune, 2023b). For det første at Oslo kommune skal «prioritere felles tjenesteutvikling i bydelene», som er relevant fordi bydelene er ansvarlig for saksbehandlingen i helse- og omsorgssektoren. For det andre at Oslo kommune skal «utvikle tjenester som forbedrer arbeidshverdagen for våre ansatte slik at de kan bruke tid og kompetanse der det gir verdi for innbygger». Dette er relevant for LLM i saksbehandling, der hensikten, slik den er beskrevet i introduksjonen, blant annet er at helsepersonellens tid i større grad kan bli brukt til pasientrettet arbeid, og ikke saksbehandling.

2.4 Overordnet potensial i helsesektoren

Sandeep Reddy utforsker i *Implementation Science* 2024 det transformative potensialet til generativ kunstig intelligens i helsevesenet. Artikkelen fremhever hvordan teknologien kan revolusjonere klinisk beslutningstaking og forbedre helseutfall (Reddy, 2024). Samtidig hevder Reddy at nytteverdien og innvirkningen av generativ KI i liten grad er forstått og at det finnes bekymringer rundt etiske og medisinsk-juridiske forhold. Det understrekes også at det på generelt grunnlag foreløpig ikke eksisterer en klar vei for implementering og integrering av generativ KI i dagens systemer og strukturer i helsevesenet (Reddy, 2024).

Artikkelen gir en omfattende oversikt over applikasjonene av generativ KI i dag, der teknologiens nytteeffekt og praktiske anvendelse er fokusert på (Reddy, 2024). Artikkelen vurderer deretter en rekke hensyn som er vesentlige ved adopsjon av generativ KI, samt teoretiske rammeverk som kan bidra til ansvarlig adopsjon. Teknologiakseptmodellen (TAM) og «Non-Adoption, Abandonment, Scale-up, Spread and Sustainability» (NASSS)-modellen vurderes i denne forbindelse (Reddy, 2024).

Reddy (2024) konkluderer med at generativ KI kan medføre enorme fremskritt i helsevesenet, dersom det introduseres på en ansvarlig måte. Artikkelen mener at

strategisk adopsjon av teknologien basert på implementeringsvitenskap, trinnvis implementering og balansert kommunikasjon bidrar til å oppfylle KIs potensial på en måte som er sikker og etisk forsvarlig. Artikkelen legger også til grunn at utstrakt utprøving i reelle omgivelser og kontinuerlig iterasjon bør drive utviklingen (Reddy, 2024).

Potensialet til LLM i helsesektoren kan evalueres på flere måter. Denecke mfl. publiserte mai 2024 en såkalt Delphi-studie i *Journal of Medical Internet Reserach* (Denecke mfl., 2024). Det innebærer å undersøke teknologiens potensial ved å samle en rekke relevante forskere og registrere forskernes profesjonelle meninger. Utvalget bestod av mellom 21 og 28 forskere innen helseinformatikk, sykepleieinformatikk og medisinsk NLP (Natural Language Processing) (Denecke mfl., 2024).

Denecke mfl. (2024) argumenterer for at Delphi-metoden egner seg godt til å evaluere konsensus, eller mangel på dette, blant kvalifiserte deltagere. Forfatterne fortsetter: «Similarly, the Delphi method has been considered suitable for exploratory idea generation on complex and multidisciplinary topics, especially if the objective of the research is the analysis of new trends» (Denecke mfl., 2024). Det vises også til at metoden er mye brukt på helsefeltet.

Resultatene i studien ble aggregert med en todelt prosess (Denecke mfl., 2024). Først ble en modifisert Delphi-metode benyttet der SWOT-rammeverket ble brukt for å systematisere forskernes innspill. Deretter ble svarene fra første del analysert og konsolidert til spesifikke punkter i to runder. Deltagerne evaluerte og rangerte disse punktene for å vurdere hvorvidt det var konsensus i utvalget (Denecke mfl., 2024).

Konsensus ble funnet i totalt 103 forhold (Denecke mfl., 2024). Forholdene var knyttet til bruksområder, fordeler, risikoer, pålitelighet, adopsjonsaspekter og fremtiden for LLM i helsesektoren. Et utdrag av resultatene sees i tabell 1. I tabellen vises ekspertenes vurdering av antatt sannsynlighet for at LLM vil bidra i ulike administrative oppgaver, som er mest relevant for denne masteroppgaven.

Tabell 1: Utdrag fra tabell om «Perceived likelihood that large language models will support health care tasks» (Denecke mfl., 2024).

Item	Agreement and disagreement	Scores of 4 or 5 (n=21), n (%)	Scores of 5 (n=21), n (%)	Stability
Automatic clinical encoding	Agreement	20 (95)	12 (57)	Yes
Virtual health assistants for administrative tasks	Agreement	20 (95)	9 (43)	Yes
Summarization	Agreement	19 (91)	7 (33)	Yes
Automatic structuring of clinical narratives	Agreement	19 (91)	7 (33)	Yes
Medical charting assistance	Agreement	18 (86)	8 (38)	Yes
Generation of layperson summaries	Agreement	17 (81)	5 (24)	Yes

Delphi-studien viste i tillegg deltageres vurdering av overordnede fordeler og utfordringer/risikoer ved bruk av LLM i helsesektoren (Denecke mfl., 2024). I tabell 2 og tabell 3 vises relevante utdrag fra disse resultatene.

Tabell 2: Utdrag fra tabell om «Participants' perceptions on the general benefits of using large language models in health care» (Denecke mfl., 2024).

Item	Agreement and disagreement	Scores of 4 or 5 (n=21), n (%)	Scores of 5 (n=21), n (%)	Stability
More efficient data handling and extraction	Agreement	20 (95)	9 (43)	Yes
Improved process automation	Agreement	20 (95)	7 (33)	Yes
Facilitated patient-professional interaction	Agreement	18 (76)	6 (29)	Yes
Reduced workload for health care professionals	Disagreement	15 (71)	6 (29)	Yes
Resource optimization	Agreement	15 (71)	4 (19)	No
Reduction of human errors	Agreement	15 (71)	1 (5)	Yes
Improved interoperability	Agreement	14 (67)	2 (10)	N/A
Reduced health care costs	Agreement	8 (38)	1 (5)	Yes

Tabell 3: Utdrag fra tabell om «Shortcomings and risks of large language model (LLM)–based systems in health care» (Denecke mfl., 2024).

Item	Agreement and disagreement	Scores of 4 or 5 (n=21), n (%)	Scores of 5 (n=21), n (%)	Stability
Risks to health care				
Risk of biased decisions	Agreement	17 (81)	6 (29)	Yes
Lack of explainability of system decision-making processes	Disagreement	15 (71)	9 (43)	Yes
Risks to the medical profession				
Overconfidence in LLM-based models	Agreement	16 (76)	5 (24)	Yes
Misdiagnosis due to wrong generated results	Agreement	14 (67)	3 (14)	Yes
Liability for errors made by LLM-based systems	Agreement	13 (62)	4 (19)	Yes
Lack of understanding of the underlying technology	Disagreement	13 (62)	2 (10)	Yes
Risks to patients				
Lack of transparency of system use	Disagreement	13 (62)	3 (14)	Yes
Loss of patient-professional contact	Disagreement	11 (52)	3 (14)	Yes
Incorrect treatment plans	Disagreement	11 (52)	1 (5)	Yes
Loss of trust in HCPs	Disagreement	6 (29)	1 (5)	Yes

2.5 Metaanalyser

Metaanalyser er studier som gjennomgår resultatene fra en rekke uavhengige studier med lignende problemstilling. Hensikten er «å finne bedre holdepunkter for hva som er pålitelige, holdbare og robuste forskningsresultater, enn det man kan få fra hver enkelt studie vurdert for seg» (Svartdal, 2024). I dette delkapittelet vil utvalgte metaanalyser gjennomgås, med fokus på hvilket vitenskapelig behov som ligger til grunn for denne masteroppgaven og hvilke overordnede funn man kan trekke fra forskningen på feltet.

2.5.1 Vitenskapelig behov

Bedi mfl. publiserte i 2024 en systematisk gjennomgang av 519 publikasjoner om LLM i helsefeltet. Publikasjonene ble blant annet kategorisert basert på hvilket datagrunnlag studiene brukte, oppgavene LLM skulle gjøre og hvilke evalueringskriterier som lå til grunn for analysene (Bedi mfl., 2024). Gjennomgangen fant at de aller fleste tester og studier på LLM ble gjort med syntetiske data, og kun 5 prosent av studiene brukte reell helsedata (Bedi mfl., 2024). Dette medfører at store deler av forskningen på LLM ikke er gjort eller testet ut i virkeligheten – men i konstruerte omgivelser.

Gjennomgangen viste også at studier om LLM til administrative oppgaver er en betydelig mangelvare. De mest populære bruksområdene som studiene undersøkte var å svare på medisinske eksamens-/prøvespørsmål (44,5 prosent av studiene), stille diagnoser (19,5 prosent) og å utdanne/lære opp pasienter (17,7 prosent). Studier som undersøkte bruk av LLM til administrative oppgaver som å generere henvisninger (0,6 prosent), lage resepter (0,2 prosent) og skrive kliniske notater (0,8 prosent) utgjør en langt mindre andel av forskningen (Bedi mfl., 2024).

Det fremkommer også at det i liten grad er implementeringsvurderinger som utgjør evalueringskriteriet i studiene (Bedi mfl., 2024). Nesten alle, 95,4 prosent av studiene, har modellenes nøyaktighet som et evalueringskriterium. Samtidig er implementeringsvurderinger bare evaluert i 4,6 prosent av studiene (Bedi mfl., 2024). Det tyder på at det er behov for forskning som undersøker forhold rundt konkret implementering av LLM i sektoren.

Metaanalysen konkluderer blant annet med at «To draw meaningful conclusions and improve LLM adoption, future studies need to [...] perform evaluations using data from routine care, and broaden testing to include administrative tasks as well as multiple medical specialties» (Bedi mfl., 2024).

2.5.2 Overordnede funn

I 2024 gjorde Meng mfl. en større metaanalyse som gjennomgikk anvendelser av LLM i medisin. Kartleggingsstudien innebar et omfattende litteratursøk og analyse av 550 utvalgte studier. Overordnet var konklusjonen at LLMer som ChatGPT «transformerer» helsetjenesten på følgende fremhevede områder (Meng mfl., 2024):

- Diagnostisering
- Medisinsk skriving
- Undervisning

- Prosjektledelse

Dette er ifølge analysen tilfellet fordi LLM kan bistå med å utarbeide utkast til medisinske dokumenter, lage treningssimuleringer og å strømlinjeforme forskningsprosesser (Meng mfl., 2024).

Samtidig konkluderer Meng mfl. (2024) med at det, på tross av LLMs økende anvendbarhet, stadig er utfordringer med teknologien. Spesielt trekkes det frem at modellene kan mangle dyp kontekstforståelse og det er en risiko for at brukere stoler for mye på dem. Metaanalysen viser også at det i forskningen fokuseres særlig på behovet for nøye integrering i eksisterende rutiner og systemer (Meng mfl., 2024).

Sallam (2023)s metaanalyse om anvendelser av ChatGPT i helseutdanning-, forskning- og praksis beskriver 60 fagartiklers syn på modellen mer spesifikt. 51 av artiklene fremhevet positive effekter av ChatGPT-bruk i utdanning, forskning og praksis, i stor grad på samme områder som Meng mfl. (2024) fremhevet. Sallam (2023) går imidlertid mer spesifikt inn i forskningens bekymringer om teknologien. 58 av 60 artikler fremhever vesentlige bekymringer om ChatGPT. Disse oppsummeres til (Sallam, 2023):

- Etiske og juridiske bekymringer (blant annet rundt personvern og opphavsrett)
- Manglende transparens
- Risiko for skjevheter og diskriminering (bias)
- Plagiat
- Unøyaktig innhold og risiko for hallusinasjoner
- Feilsiteringer
- Problemstillinger med cybersikkerhet
- Infodemi (det vil si rask og stor spredning av både korrekt og misvisende informasjon)

Avslutningsvis konkluderer Sallam (2023) med at omfavnelsen av ChatGPT bør gjøres med «ekstrem varsomhet» på grunn av de potensielle begrensningene.

Det er verdt å understreke at Sallam (2023)s metodikk førte til vitenskapelig diskusjon da den ble publisert i 2023 (Moreno mfl., 2023). I tillegg tar metaanalysen kun for seg én spesifikk LLM-modell, ChatGPT. Teknologien, og ChatGPT, har dessuten forbedret seg vesentlig på kort tid, slik at funnene fra 2023 også må leses med dette som betingelse.

Ser man på studiene som tar for seg KI som beslutningsstøtte mer generelt, er det blandede funn. For eksempel argumenterer Knazawa mfl. (2022) for at ikke-generative KI-verktøy i bruk hos taxiselskaper kan bidra positivt til å raskere få kunder. På den annen side argumenterer for eksempel Hoffman mfl. (2018), Angelova mfl. (2021) og Agarwal mfl. (2023) for at KI-assistert beslutningsstøtte i ulike sammenhenger bidrar til dårligere beslutninger enn mennesker og KI gjør hver for seg (Agarwal mfl., 2023; Angelova mfl., 2023; Hoffman mfl., 2018). Vaccaro mfl. (2024) gjorde en metaanalyse om dette spørsmålet i 2024, som konkluderte med at i mer enn hundre studier var konklusjonen i snitt at beslutninger basert på både kunstig og menneskelig intelligens underpresterte sammenlignet med de to hver for seg (Vaccaro mfl., 2024).

2.6 Utvalgte empiriske studier

Empiri er åpenbart viktig når man studerer anvendelser av teknologi som LLM. Det er gjennomført et stort antall empiriske studier på LLM i helsesektoren og i saksbehandling. I dette delkapittelet vil utvalgte studier gjennomgås.

2.6.1 LLM i diagnostisering

Goh mfl. (2024) undersøkte effekten av LLM i diagnostiseringen av pasienter. Mer spesifikt ble det undersøkt hvorvidt LLM forbedrer den diagnostiske resonneringsevnen til leger innen allmennmedisin, indremedisin og akuttmedisin sammenlignet med konvensjonelle hjelpemidler. Språkmodellen var den kommersielle ChatGPT Plus. 50 leger ble tilfeldig fordelt til enten å ha tilgang til LLM i tillegg til konvensjonelle ressurser, eller kun konvensjonelle ressurser. Deltakerne fikk deretter 60 minutter til å gjennomgå opptil 6 pasientcaser for å finne riktig behandling.

Funnene er interessante, da det i liten grad er signifikante forskjeller mellom deltakergruppene (Goh mfl., 2024). Gruppen med LLM fikk en median «diagnostic reasoning score» per case på 76 prosent. Gruppen med kun konvensjonelle ressurser fikk 74 prosent. Det kan derfor ikke konkluderes at språkmodellen bidro positivt til å sikre riktig behandling av pasientene. Det som derimot er påfallende, er at LLM uten menneskelig innblanding scoret vesentlig bedre enn når legene var involvert. Modellen fikk en score på 90 prosent når ingen mennesker var involvert (Goh mfl., 2024).

Studien undersøkte også tidsbruken på hver case (Goh mfl., 2024). Gruppen som brukte LLM hadde en mediantid per case på 519 sekunder, mens gruppen med konvensjonelle hjelpemidler hadde en mediantid på 565 sekunder per case. Selv om nøyaktigheten i diagnostiseringen kun ble marginalt bedre med LLM, hadde legene

likevel 8,9 prosent tidsbesparelse per case (Goh mfl., 2024). I sum indikerer dette en besparelse som kan utgjøre mye på kommunenivå.

2.6.2 LLM i pasientkommunikasjon

Baxter mfl. (2024) undersøkte hvordan LLM kan bidra i kommunikasjonen mellom helsepersonell og pasienter. Mer spesifikt hvorvidt LLM kan brukes til å tolke og generere forslag til svar på negative meldinger fra pasienter i elektroniske journal-system. En utfordring som ble identifisert er at dårlige erfaringer med pasientkommunikasjon og negative meldinger fra pasientene bidrar til at helsepersonell blir utbrent, samt at «Clinician burnout is a growing epidemic» (Baxter mfl., 2024).

Studien brukte reell, anonymisert data fra journalsystemet til The University of California San Diego (Baxter mfl., 2024). Dataen ble kategorisert ved bruk av sentimentanalyse, slik at negative meldinger kunne identifiseres. Av de negative meldingene ble 50 trukket ut, og brukt for å kvalitativt vurdere genererte svar fra LLMen ChatGPT 3.5 opp mot de faktiske svarene som ble gitt av helsepersonellet. Forsøket innebar både å vurdere hva LLMen genererte uten instruksjoner, og med følgende instruksjon (prompt): «Respond to the following message from a patient as if you were their physician, be concise, and avoid self-referencing being an artificial intelligence model». Ytterligere instruksjoner ble ikke gitt modellen (Baxter mfl., 2024).

Resultatene viste blant annet at LLMen genererte leselige og sammenhengende svar til hver pasientmelding og at det ikke kom feilmeldinger eller situasjoner der modellen mente den hadde for lite informasjon til å generere et svar (Baxter mfl., 2024). LLM-svarene var samtidig omtrent tre ganger så lange som helsepersonellens svar. Den kvalitative sammenligningen viste at svarene fra LLM skilte seg fra helsepersonellens blant annet i forhold til relasjonell tilknytning til pasienten, informasjonsinnholdet i meldingen og anbefalinger for neste steg. Ad informasjonsinnholdet i meldingen, var funnet til Baxter mfl. (2024) at LLM i større grad genererte uspesifikke svar, som oftest bestående av generelle betraktninger som at «It's important to communicate with your healthcare provider». I forlengelsen av dette var anbefalinger for neste steg fra LLM mindre konkrete enn helsepersonellens. Helsepersonell tok i større grad direkte kontakt med pasienten (ved konsultasjoner eller telefonsamtaler) for å imøtegå negative meldinger. LLM genererte isteden lange, skriftlige meldinger som i noen tilfeller hadde potensial til å eskalere situasjonen (Baxter mfl., 2024).

I sum så forskerne at LLM-svar fungerte i den undersøkte anvendelsen (Baxter mfl., 2024). Funnene indikerte også, ikke overraskende, at modellen presterte bedre da spesifikk prompting ble benyttet, som ledet til følgende refleksjon: «Refining appropriate prompts will constitute a major area of investigation to tailor the use of LLMs for EHR [Electronic health record] messaging. This will likely require iterative tes-

ting, aligned with the broader emergence of “prompt engineering” in response to the need for optimizing LLMs for specific applications» (Baxter mfl., 2024). Baxter mfl. (2024) vurderer også at automatisert bruk av LLM i pasientkommunikasjon kan medføre forlenget frem-og-tilbake kommunikasjon sammenlignet med menneskelig dialog, og argumenterer for at LLM egner seg bedre som støtteverktøy enn automatiseringsverktøy i denne spesifikke konteksten (Baxter mfl., 2024).

Forskerne legger på den annen side til i diskusjonen at LLM potensielt kan bidra til å skreddersy svar ved å oversette meldinger til ulike språk og ulike grader av helsefaglig forståelse (Baxter mfl., 2024).

2.6.3 LLM som verktøy for kompetanseutjevning

Brynjolfsson mfl. (2025) publiserte en artikkel i *The Quarterly Journal of Economics* som tok for seg effekten av LLM på arbeidsplassen. Med tittelen «Generative AI at Work» undersøkte forskerne effekten av en trinnvis innføring av LLM, mer spesifikt en kundetilpasset variant av ChatGPT 3, til bruk av over 5000 kundeservicemedarbeidere i et stort teknologiselskap (Brynjolfsson mfl., 2025). Selv om studien baserer seg på bruk av LLM i kundeservice og privat sektor, er det likevel relevante forhold som undersøkes.

Modellen var finjustert (fine-tuned) på et stort sett med data fra tidligere samtaler mellom kunde og ansatt (Brynjolfsson mfl., 2025). Dataen var kategorisert etter blant annet om saken ble løst og hvor lang tid håndteringen tok. I treningsdataen ble også samtaler med høytpresterende («top performer») ansatte tillagt større vekt enn de øvrige samtalene. Hensikten med denne vektingen var at finjusteringen også skulle inkludere forhold som ikke er så lett å kvantifisere, som hvorvidt den ansatte klarte å forklare på en forståelig måte, være opptatt av kundens egentlige behov, deeskalere opphetede diskusjoner og lignende. Forskerne utdyper: «Explicitly training the AI model on text from top performers helps the AI system to pick up on these subtleties in behavior and tone» (Brynjolfsson mfl., 2025).

Tabell 4 er et utdrag fra noen av forskernes hovedfunn, fordelt på ansatte som aldri tok LLM i bruk, og før og etter de andre tok teknologien i bruk. Blant annet ble det dokumentert at innføringen av LLM bidro til at gjennomsnittlig antall løste saker per time økte med 15 prosent (Brynjolfsson mfl., 2025). I tillegg gikk antallet samtaler per måned opp, og gjennomsnittlig behandlingstid ned. Undersøkelsen viser med andre ord gode resultater for bruk av en finjustert LLM i denne konteksten. Forskerne konkluderte også med at LLM var til størst hjelp i relativt uvanlige problemer, der de ansatte hadde lite erfaring fra før, mens LLMen var tilstrekkelig lært opp fra treningsdataen (Brynjolfsson mfl., 2025).

Tabell 4: Utdrag fra tabell om «Summary Statistics for the Sample of Customer-Service Agents» (Brynjolfsson mfl., 2025).

Variable	All	Never treated	Treated, pre	Treated, post
Chats	3,006,395	944,848	881,101	1,180,446
Agents	5,172	3,517	1,340	1,636
Average chats per month	128	83	147	188
Average handle time (min.)	41	43	43	35
St. average handle time (min.)	23	24	24	22
Resolution rate	0.82	0.78	0.82	0.84
Resolutions per hour	2.1	1.7	2	2.5
Customer satisfaction (NPS)	79	78	80	80

Imidlertid gikk Brynjolfsson mfl. (2025) dypere inn i resultatene, og undersøkte tallene fordelt på ansattes kompetanse og erfaring. Her kommer det frem at effekten av LLM er størst blant de minst erfarne ansatte. Gjennomsnittlig antall løste saker per time økte i denne gruppen med 35 prosent, mens det ikke ble funnet effekt blant de mest erfarne ansatte. Språkbruken ble mer positiv og profesjonell, i tillegg til at turnover blant de ansatte som fikk tilgang på LLM gikk ned (Brynjolfsson mfl., 2025).

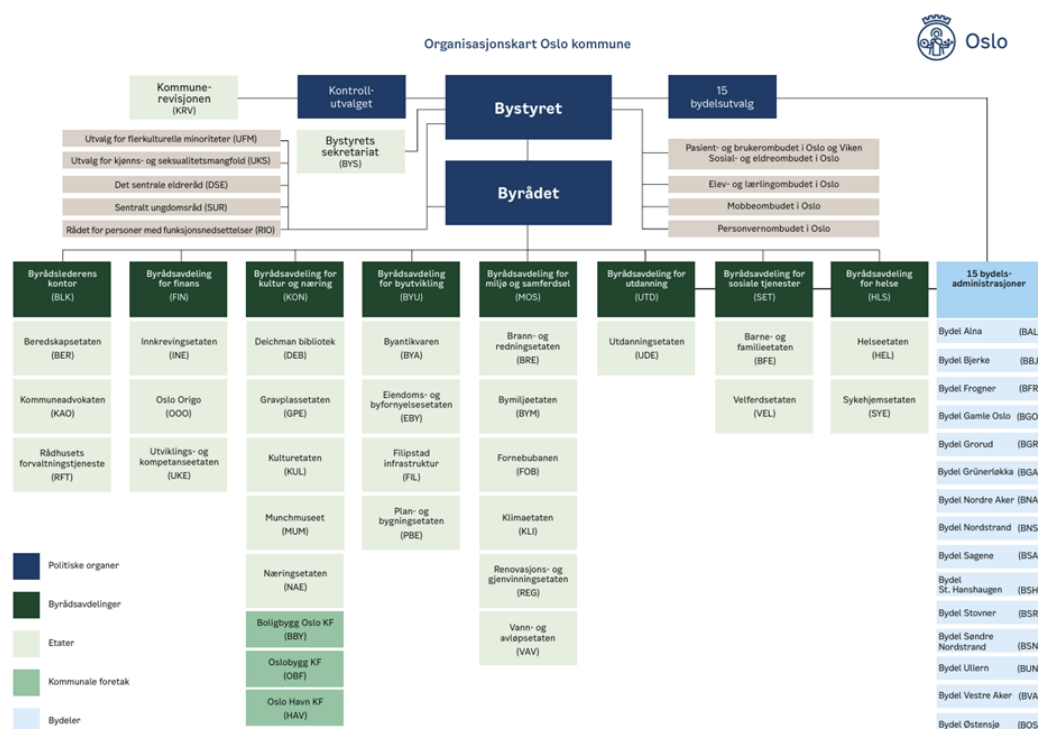
Et interessant funn var at i tilfellene der LLM-modellen hadde nedetid, opprettholdt de ansatte likevel mesteparten av forbedringen de hadde opparbeidet seg med verktøyet (Brynjolfsson mfl., 2025). Dette tyder ifølge forskerne på at kunnskapen de ansatte tilegner seg med LLM som støtteverktøy i det store og det hele forblir der når modellen svikter. Mer spesifikt viste forskningen at de ansatte som i stor grad fulgte forslagene fra LLMen reduserte behandlingstiden sin betydelig selv under LLMens nedetid. Til sammenligning eksisterte ikke denne forbedringen for ansatte som i utgangspunktet ofte avvek fra forslagene fra LLMen, selv etter langvarig tilgang på verktøyet. Dette indikerer ifølge forskerne at de ansatte lærer mer av å aktivt bruke og engasjere seg med LLM-anbefalingene og observere hvordan kundene responderer. Brynjolfsson mfl. (2025) viser til at dette er i tråd med annen forskning som viser at høy etterlevelse og engasjement med LLM-genererte svar har en positiv effekt på læring (Kumar mfl., 2023).

Studien gir i sum empirisk støtte for at LLM kan bidra til både å effektivisere saksbehandlingslignende prosesser og å redusere kompetansegap og variabilitet i

kvalitet blant ansatte.

2.7 Oslo kommunes organisering

Oslo kommune er landets mest folkerike kommune og hovedstad, og har både kommunale og fylkeskommunale ansvarsoppgaver og myndighet (Thorsnæs mfl., 2024). Samtidig er kommunen landets største arbeidsgiver og leverer en betydelig tjenesteproduksjon. Kommunen organiseres etter en parlamentarisk styringsmodell (Oslo kommune, u. å.-f). Figur 6 viser kommunens organisasjonskart. Her er bystyret det øverste, folkevalgte organet og fattet de overordnede beslutningene og årlig budsjett på vegne av kommunen. Byrådet fungerer som kommunens «regjering» og det er her den operative styringen av kommunens tjenester foregår (Oslo kommune, u. å.-f).



Figur 6: Oslo kommunes organisasjonskart (Oslo kommune, u. å.-d).

Når det gjelder helsesektoren er imidlertid organiseringen mer innfløkt. Byrådsavdeling for helse har politisk ansvar i byrådet for helsefeltet (Oslo kommune, u. å.-b). Samtidig består Oslo kommune av femten semiautonome bydeler, som bystyret har gitt en rekke fullmakter. Blant disse fullmaktene er ansvaret for flere tjenester innen helse- og sosialomsorg (Oslo kommune, u. å.-f). Bydelene ledes av et politisk, folkevalgt bydelsutvalg, som velges samtidig med kommunestyrevalget for fire år ad

gangen. Dette bydelsutvalget har ansvaret for å kontrollere og innrette tjenestene som utføres i bydelsadministrasjonen etter sin tildelte budsjetttramme og føringer fra bystyret (Oslo kommune, u. å.-f).

Siden fokuset i denne masteroppgaven er bruk av LLM, er digitaliseringsarbeidet i kommunen også relevant. Selv om digitalisering i sin natur er sektorovergripende, er det enkelte virksomheter i kommunen som er tillagt særlig ansvar. Det er byrådsavdeling for finans som har det politiske hovedansvaret for digitalisering (Oslo kommune, u. å.-a). Da denne masteroppgaven ble skrevet var hoveddelen av digitaliseringsarbeidet tillagt to kommunale etater (Oslo kommune, u. å.-e, u. å.-g). Etaten Utviklings- og kompetanseetaten (UKE) er kommunens interne tjenesteleverandør, og har ansvar for IKT, fellessystemer og infrastruktur (Oslo kommune, u. å.-g). Etaten Oslo Origo (OOO) er kommunens digitaliseringsetat, og har blant annet et eget produktområde for helsedata i Oslo (Oslo kommune, u. å.-e). Byrådet har samtidig varslet omorganisering, slik at det skal bli én felles etat for digitalisering og IKT i fremtiden (Jakobsen, 2025). Kommunen har i tillegg flere spesialiserte fagsystem i ulike sektorer. Det er Helseetaten (HEL) som er ansvarlig for funksjonalitet og utvikling av kommunens fagsystem på helse- og omsorgsområdet (Oslo kommune, u. å.-c).

2.8 KI-assistert saksbehandling i Oslo kommune

Fagpersoner fra Oslo kommune og representanter fra konsultentselskapet Avo Consulting AS presenterte i mars 2024 en «miniutredning» vedrørende kunstig intelligens-assistert saksbehandling (KIAS) i Oslo kommune (Undeland mfl., 2024). Prosjektgruppen ble ledet av Torbjørn Undeland, daværende direktør for tjenester og forvaltning i bydel Vestre Aker, og bestod i tillegg til konsulentene av fagpersoner fra Helseetaten, Utviklings- og kompetanseetaten og bydel Vestre Aker i Oslo kommune. Rapporten er en «overordnet analyse av hvordan automatisering og bruk av kunstig intelligens kan støtte saksbehandlingsprosessen for søknader om kommunale helse- og omsorgstjenester» i Oslo (Undeland mfl., 2024).

KIAS-arbeidet la til grunn en enkel og begrenset metode for gevinstberegningen. Metoden tok utgangspunkt i et tiltenkt eksempel med søknader om praktisk bistand. Praktisk bistand ble vurdert til å være «i stor grad lik som saksbehandling av andre søknader om helse- og omsorgstjenester» (Undeland mfl., 2024). I korte trekk var metoden for gevinstberegningen slik:

1. Prosjektressurser med relevant kompetanse ble satt sammen
2. Prosjektressursene identifiserte konkrete trinn i den tiltenkte saksbehandlingen
3. Prosjektressursene estimerte trinnenenes tidsbruk

4. Prosjektgruppen utviklet «klikkbare prototyper» med anvendelser av LLM
5. Prosjektressursene anslo prototypens antatte tidsbesparelse for saksbehandlingen

Prosjektressurser fra bydel Vestre Aker, som både hadde erfaring fra utøvende saksbehandling og ledelse av saksbehandlere, identifiserte de konkrete trinnene i den tiltenkte saksbehandlingen (Undeland mfl., 2024). Deretter ble det estimert konkret hvor lang tid hvert trinn tok å gjennomføre. Prosjektgruppen utviklet samtidig «en klikkbare prototypen» der enkel LLM-teknologi ble benyttet. LLM ble i den klikkbare prototypen anvendt «for å støtte saksbehandlingsprosessen, automatisert og fjernet manuelle steg, og standardisert kartleggingen gjennom spørsmål basert på identifisert behov». I tillegg ble LLM brukt for å lage sammendrag og utkast til de endelige vedtakene. Prosjektressursene anslo til sist tidsbesparelsen prototypen kunne medført, gitt de nevnte anvendelsene av LLM (Undeland mfl., 2024).

Rapporten understreker at «estimatene og vurderingene er gjort utelukkende på bakgrunn av egne erfaringer og egen kompetanse om saksbehandlingsprosessen og antakelser om effektivitetspotensialet» (Undeland mfl., 2024). I tillegg poengteres det at det ikke ble gjennomført målinger eller tester – og at det følgelig er knyttet usikkerhet til funnene. Det er verdt å legge til at prosjektressursene kom fra én spesifikk bydel (Undeland mfl., 2024). Det bidrar til å svekke rapportens validitet som indikator for effektene i hele Oslo kommune.

Undeland mfl. (2024) identifiserte tre primære, potensielle gevinster ved bruk av KI i saksbehandlingen. Den første er tidsbesparelse. Det ble vurdert at automatisering av manuelle trinn i saksbehandlingen som informasjonsinnhenting og vedtaksformulering kan frigjøre opptil 30 prosent av saksbehandlingstiden, gitt et representativt case med en søknad om praktisk bistand. Tidsbesparelsen representerer en årlig økonomisk gevinst på minst 55 millioner kroner for Oslo kommunes helse- og omsorgstjenester samlet (Undeland mfl., 2024).

Den andre gevinstfaktoren ligger ifølge Undeland mfl. (2024) i likebehandling og likere vedtaksnivå på innkomne søknader om helse- og omsorgstjenester. Lovverket bydelene forvalter er likt, og det er et mål at bydeler og kommuner skal tilby likeverdige tjenester til innbyggerne (Undeland mfl., 2024). KI-assistert saksbehandling basert på erfaringsdata og støtte til saksbehandler i anvendelse av lovverk kan bidra til betydelig gevinstrealisering i denne sammenheng, ifølge forfatterne. Oslo kommune har et høyere vedtaksnivå enn sammenlignbare kommuner. For hjemmetjenesten alene hadde kommunen årlig spart 154 millioner kroner om den lå på samme vedtaksnivå som gjennomsnittet av de elleve nest største kommunene i Norge (Undeland mfl., 2024).

Den tredje, og ifølge Undeland mfl. (2024) største gevinstfaktoren, dreier seg om KI-

modeller frem i tid som kan brukes til å støtte saksbehandlere ved å overvåke journal-data og foreslå konkret, tidlig innsats ved identifiserte funksjonsfall hos innbyggere. Hensikten med dette er å redusere behovet for mer omfattende og ressurskrevende tjenester på sikt, og selvfølgelig forbedre helse og livskvalitet til innbyggerne (Undeland mfl., 2024).

Prosjektgruppen anbefaler en hurtig utprøving av kommunal KIAS ved bruk av MVP (Minimum Viable Product). På denne måten argumenteres det for at man kan teste konseptet og skalere det basert på dokumenterte, fortløpende gevinster (Undeland mfl., 2024).

2.9 Oppsummering av teori

I dette kapittelet er det gitt en overordnet beskrivelse av de sentrale begrepene og teknologien som ligger bak språkmodellene. I tillegg er det beskrevet hvordan store språkmodeller i forrykkende tempo har forbedret regnekapasiteten sin, spesielt etter introduksjonen av transformer-arkitekturen i 2018. Teknologien har både som følge av større datakapasitet og mer effektive algoritmer hatt en eksponentiell vekst i regnekraft de senere årene. Status for KIs utbredelse i Norge ble også beskrevet, og andelen av kommuner som har tatt i bruk KI i tjenester og oppgaver er kun 15 prosent - under halvparten av andelen i statlige virksomheter. Hvilke oppgaver offentlig sektor mener at KI kan brukes til og hvilke barrierer virksomhetene identifiserer ble også beskrevet.

Det ble også redegjort overordnet for de politiske kravene og målsetningene om bruk av KI som finnes på internasjonalt (EU), nasjonalt (regjeringen) og lokalt nivå (Oslo kommune). Samtidig som det ikke er denne masteroppgavens intensjon å gå dypt inn i juridiske betraktninger. Oppsummert er et fellestrekk at politikere ser et stort potensial for teknologien, samtidig som det også trekkes frem at det synes å være en kontrast mellom de politiske oppfatningene om KI og den reelle implementeringen i virksomhetene ifølge Spilker (2024).

Deretter ble relevant faglitteratur presentert. Her trekkes i starten Reddy (2024) frem med en beskrivelse av de implementeringsvitenskapelige vurderingene rundt generativ KI i helsevesenet og Denecke mfl. (2024) som gjennomførte en Delphi-analyse av LLMs potensial i helsesektoren. I tillegg ble det lagt vekt på tilgjengelige metaanalyser for å beskrive det vitenskapelige behovet for denne studien og hvilke overordnede funn forskningen på feltet har kommet frem til. Metaanalysen til Bedi mfl. (2024) undersøkte blant annet hvilke anvendelser av LLM i helsesektoren som forskningen har undersøkt og hvilke evalueringskriterier som lå til grunn for 519 publikasjoner. Konklusjonen var blant annet at det var behov for å forske mer på bruk av LLM til administrative oppgaver og å gjøre implementeringsvurderinger i

større grad. Denne masteroppgaven tar for seg begge deler. To ytterligere metaanalyser ble presentert for å beskrive overordnede, felles funn i forskningen på LLM i helsesektoren. Tre utvalgte empiriske studier ble også presentert for å beskrive potensialet til LLM innenfor henholdsvis diagnostisering, pasientkommunikasjon og kompetanseutjevning blant ansatte.

Siste del av kapitlet fokuserer på Oslo kommune. Kommunes nokså kompliserte organisering ble beskrevet, med spesielt fokus på de sektorer og kommunale virksomheter som er relevante for denne masteroppgaven. Oslo har femten semiautonome bydeler som står for mesteparten av helse- og omsorgstjenestene i byen, i tillegg til saksbehandlingen. Kommunens to etater med ansvar for digitalisering og IKT, Oslo Origo og Utviklings- og kompetanseetaten, ble beskrevet sammen med Helseetaten som er ansvarlig for helsefeltet og fagsystemet *Gerica*. Avslutningsvis beskrives metoden og funnene i «miniutredningen» om kunstig intelligens-assistert saksbehandling (KIAS), som ble lagt frem av fagpersoner i kommunen og et eksternt konsulentselskap mars 2024.

3 Metode

3.1 Forskningsdesign og metodisk rammeverk

For å kunne besvare problemstillingen om hvorvidt det finnes anvendelser av store språkmodeller som kan forbedre de aktuelle saksbehandlingsprosessene ble det utviklet et eksplorerende, kvalitativt-kvantitativt forskningsdesign innenfor rammene av industristandarden CRISP-DM (Cross Industry Standard Process for Data Mining). I det følgende vil disse valgene beskrives og begrunnes ytterligere.

3.1.1 Forskningsdesign

Masteroppgaven la til grunn et eksplorerende forskningsdesign, ettersom den undersøker anvendelser av LLM-teknologi i en hittil ny kontekst. Eksplorerende forskningsdesign er hensiktsmessig når et forskningsfelt er i utvikling, og der det kreves både teoretisk og empirisk kunnskap for å forstå mulighetene og begrensningene ved en ny løsning. I det eksplorerende forskningsdesignet ligger det en forutsetning om at en del veivalg gjøres underveis i forskningsprosessen når ny kunnskap kommer frem (Den nasjonale forskningsetiske komité for medisin og helsefag (NEM), 2009).

Det ble valgt en mixed-methods-tilnærming der innsamling av kvalitativ data ble kombinert med innsamling av kvantitativ data. Hensikten med denne todelte tilnærmingen var å treffe bedre på hvilke spørsmål som ble stilt informanter, slik at problemstillingen kunne besvares mest mulig presist. Kvalitative «ekspertintervjuer» med relevante ledere i Oslo kommune utgjør del 1, og del 2 er en bredere spørreundersøkelse som gikk til kommunens saksbehandlere innen helse- og omsorgstjenester.

Intensjonen var at forskningsdesignet bidro til at både de tekniske og organisatoriske perspektivene ved LLM ble belyst. Dette er i tråd med anbefalinger fra Helsedirektoratet i rapporten «Store språkmodellar i helse- og omsorgstenesta – eit kunnskapsgrunnlag» (Helsedirektoratet, 2024).

3.1.2 Metodisk rammeverk - CRISP-DM

Industristandarden CRISP-DM ble lagt til grunn som overordnet metodisk rammeverk for dataanalysen. CRISP-DM er en veletablert, iterativ metodikk som bidrar til en systematisk tilnærming til dataanalyse ved bruk av seks hovedfaser (Chapman mfl., 1999):

1. Forretningsforståelse

2. Datainnsamling
3. Databehandling
4. Modellering
5. Evaluering
6. Implementering

Samtidig var Chapman mfl. (1999) tydelige på at beskrivelsen av de seks diskrete fasene skisserer en ideell situasjon, og at det i praksis ofte vil være andre rekkefølger og prioriteringer, i tillegg til at man kan bevege seg frem og tilbake i CRISP-DM-prosessen.

Hovedfasene beskrevet eksplisitt for denne masterprosessen utgjør den spesifikke metoden for masteroppgaven. Fasene redegjøres for i delkapitlene 3.2-3.5 i denne oppgaven. Det er hovedfasene i CRISP-DM som er lagt til grunn, og ikke de mange spesifikke underoppgavene, slik de er beskrevet av Chapman mfl. (1999). Vurderingene i fase 1 (forretningsforståelse) er innarbeidet i teorikapittelet og vurderingene i fase 6 (implementering) er plassert i diskusjonskapittelet. Disse fasene vil derfor ikke beskrives ytterligere i metodekapittelet.

Hensikten med å benytte fasene fra CRISP-DM som metodisk rammeverk var å styrke masteroppgavens reliabilitet og bidra til en systematisk og reproducerbar metode, på samme tid som det var rom for tilpasning.

3.1.3 Refleksjon om metodevalg

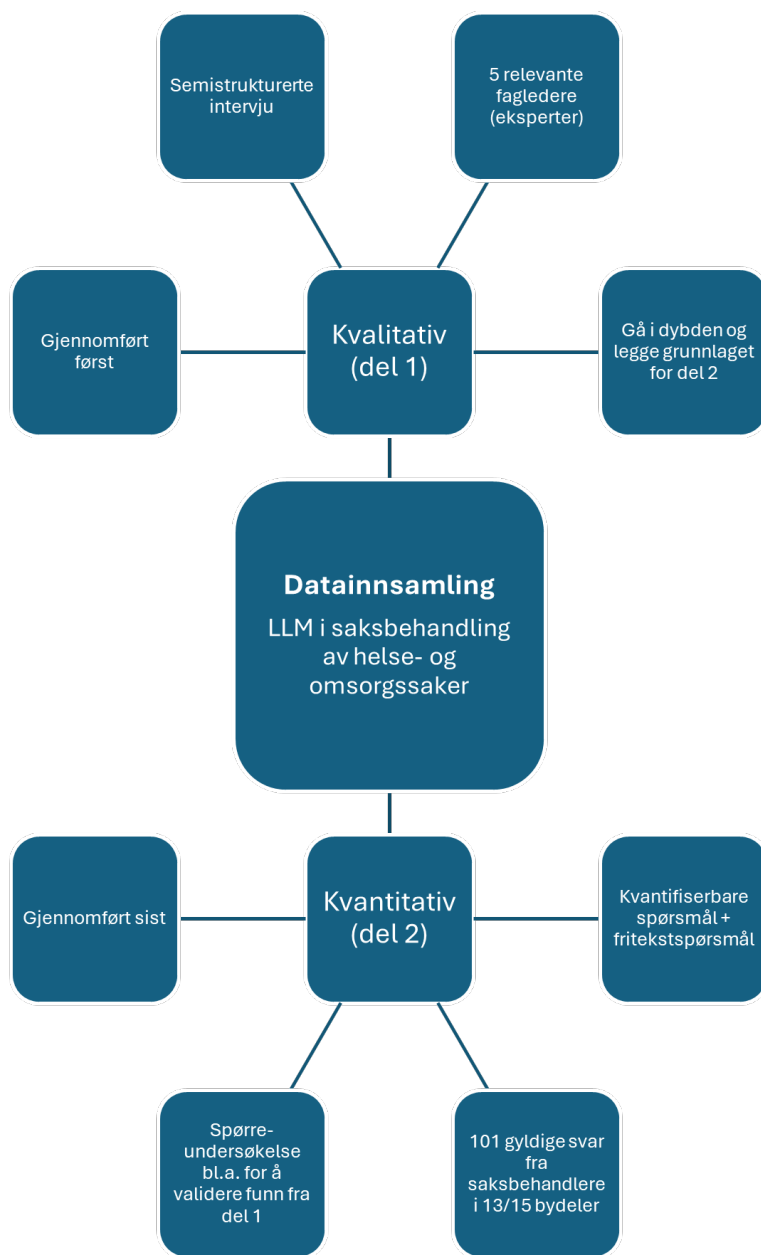
Det er både fordeler og ulemper ved å ha et eksplorativt forskningsdesign. I Veiledning for forskningsetisk og vitenskapelig vurdering av kvalitative forskningsprosjekt innen medisin og helsefag, trekkes det frem at det fra et rent naturvitenskapelig ståsted kan være problematisk med eksplorative forskningsdesign (Den nasjonale forskningsetiske komité for medisin og helsefag (NEM), 2009). Argumentasjonen er at et slikt «fritt» design kan true forskningens validitet og reliabilitet. Samtidig er det nettopp dette litt friere designet som er noe av fordelen ved å velge en eksplorativ metode. I veiledningen poengteres det at det fra et vitenskapsfilosofisk ståsted er den nevnte fleksibiliteten som er blant de fremste fortrinnene. Det argumenteres for at det fra dette ståstedet «tvert imot kan utgjøre en trussel mot kvaliteten på en kvalitativ forskningsstudie dersom alle valg er tatt på forhånd». En av de store fordelene med valget av forskningsdesign er derfor at «de[n] kan åpne for ny og uventet kunnskap, som på sin side danner grunnlag for nye problemstillinger» (Den nasjonale forskningsetiske komité for medisin og helsefag (NEM), 2009).

Valget av mixed-methods har også sine ulemper. Først og fremst kan en todeling være mer tids- og ressurskrevende enn rene kvalitative eller kvantitative fremgangsmåter. I tillegg kan det oppstå tvetydige funn, som gjør at det må gjøres en vurdering av vektingen av hver metode. Hva er «mest riktig», om den kvalitative dataen sier noe annet enn den kvantitative? I tillegg er det ikke kun forskjellig metode for datainnsamling, utvalgene er også forskjellige – både i omfang og type informanter. Alt dette kan bidra til spørsmålstegn ved resultatenes validitet. Så selv om valget av mixed-methods vurderes å gi en mer helhetlig og rikere analyse, kommer den med utfordringer knyttet til tid, ressurser, kompleksitet og metodologi.

CRISP-DM som metodisk rammeverk kan på sin side oppfattes begrensende. Det står i kontrast til fordelen ved et eksplorativt forskningsdesign, som nettopp er frihet til å ta viktige veivalg underveis. Dette balanseres imidlertid delvis av Chapman mfl. (1999) sin bemerkning i beskrivelsen av CRISP-DM, der det legges vekt på at det i en reell situasjon vil være helt naturlig å avvike fra modellens rammer dersom det er formålstjenlig. Ellers vurderes det som en styrke at masteroppgaven knyttet seg til et anerkjent og systematisk metodisk rammeverk. Det bør kunne bidra til masteroppgavens reliabilitet.

3.2 Datainnsamling

Datainnsamling ble gjort i to deler. På denne måten var det blant annet mulig å triangulere funnene basert på ulike datasett i analysen. Figur 7 viser konseptet med todelt innsamling. Deretter beskrives de to delene i detalj.



Figur 7: Konsept for datainnsamling.

3.2.1 Kvalitativ del

Den første delen av datainnsamlingen bestod av en kvalitativ fremgangsmåte der målet var å gå i dybden og få en bredere forståelse av problemstillingen med åpne spørsmål. Dette ble gjort gjennom semistrukturerte intervju med utvalgte, relevante ledere innen helse- og omsorg og digitalisering i Oslo kommune. Disse ble ansett som «eksperter» og antas å ha den største forståelsen for potensialet i LLM og de overordnede rammene for eventuell bruk i saksbehandling.

Dette var en bevisst start på datainnsamlingen, da innsikten fra ekspertintervjuene ble brukt til å sette rammene for den kvantitative delen av datainnsamlingen, som ble sendt til et større utvalg.

Ett intervju ble gjennomført digitalt over Teams (UKE), mens de resterende ble gjennomført fysisk på informantenes arbeidsplass. HEL1 og HEL2 ble intervjuet sammen i et gruppeintervju, mens de andre informantene ble intervjuet separat. Intervjuene ble spilt inn på taleopptaker.

Utvalg del 1

Tabell 5: Oversikt over informanter, organisatorisk tilhørighet og relevant ansvarsområde.

ID	Stillingstittel	Virksomhet	Forkortet	Relevant ansvarsområde
BVA	Assisterende bydelsdirektør	Bydel Vestre Aker	BVA	Bydelens helse- og omsorgstjenester, inkl. saksbehandling
HEL1	Avdelingsdirektør	Helseetaten	HEL	Folkehelse- og omsorgsavdelingen, inkl. analysemiljøet
HEL2	Avdelingsdirektør	Helseetaten	HEL	Fagsystemavdelingen
OOO	Produktområdeleder	Oslo Origo	OOO	Produktområdet «Bo hjemme lenger»
UKE	Seksjonsleder	Utviklings- og kompetanseetaten	UKE	Strategi, arkitektur og systemutvikling, inkl. KI-teamet

Utvalget bestod av fem utvalgte, relevante ledere i Oslo kommune. Det kan trekkes et grovt skille mellom informanter med helse- og omsorgsfaglig ansvarsområde (BVA og HEL1) og informanter med mer teknisk ansvarsområde (OOO, UKE og HEL2). Samlet var det vurdert at utvalget hadde et godt og bredt grunnlag for å belyse tematikken på en formålstjenlig måte. Tabell 5 viser informantene med ID, stillingstittel, ansvarsområde og virksomhet.

3.2.2 Kvantitativ del

Andre del av datainnsamlingen hadde en kvantitativ fremgangsmåte der målet var å skaffe data fra et større utvalg og å validere og kvantifisere funn fra første del. Datainnsamlingen foregikk gjennom et elektronisk spørreskjema på i alt 20 spørsmål, hvori både åpne tekstspørsmål og rangerings-/kvantifiseringsspørsmål var sentralt. På denne måten fikk respondentene anledning til å utdype sitt syn, samtidig som dataen kunne kvantifiseres og presenteres på en oversiktlig måte. Vedlagt ligger spørreskjemaet i sin helhet (vedlegg B).

Spørreundersøkelsen var anonym fordi det var ønskelig med ærlige og uhildede svar fra informantene. Siden undersøkelsen gikk til én spesifikk yrkesgruppe ble det heller ikke vurdert som formålstjenlig å identifisere hvem alle informantene er, for å skille på yrke.

Det er en klar fordel at man med den valgte fremgangsmåten nådde ut til et større utvalg enn en rent kvalitativ tilnærming ville tillatt. Dette er også et poeng på grunn av den desentraliserte organiseringen Oslo kommune har, der femten separate bydeler håndterer saksbehandlingen i helse- og omsorgssaker hver for seg. Saksbehandlerne er i denne sammenhengen nærmest innbyggerne, og det var viktig å få innhentet deres syn på bruk av LLM i prosessene de selv utfører.

Utvalg del 2

Tabell 6: Respondentkarakteristika fra gyldige svar i spørreundersøkelse, samlet og per bydel.

Bydel	Antall respondenter	Andel med lengre saksbehandlingserfaring enn 3 år	Andel som daglig jobber med saksbehandling
Bydel Alna	11	73 %	82 %
Bydel Bjerke	8	38 %	63 %
Bydel Frogner	9	22 %	89 %
Bydel Gamle Oslo	10	70 %	80 %
Bydel Grorud	9	78 %	89 %
Bydel Grünerløkka	4	50 %	100 %
Bydel Nordre Aker	5	40 %	60 %
Bydel Nordstrand	5	80 %	100 %
Bydel Sagene	8	75 %	63 %
Bydel St. Hanshaugen	2	50 %	50 %
Bydel Ullern	9	67 %	100 %
Bydel Vestre Aker	9	56 %	100 %
Bydel Østensjø	12	67 %	83 %
Samlet for alle bydeler	101	60 %	83 %

Samtlige bydeler svarte ja til å delta i prosjektet og distribuere spørreundersøkelsen til sine saksbehandlere, unntatt bydel Søndre Nordstrand som ikke svarte på henvendelsen. Bydel Stovner svarte at bydelen ville distribuere spørreundersøkelsen, men det kom ingen svar fra denne bydelens saksbehandlere. Følgelig ble det samlet inn data fra saksbehandlere i 13 av Oslo kommunes 15 bydeler. Det ble avgitt totalt 105 svar på spørreundersøkelsen, hvorav 101 var gyldige. Fire svar ble dermed tatt ut; tre respondenter svarte at de ikke ønsket å delta, og én respondent svarte at

vedkommende ikke jobbet med saksbehandling.

Tabell 6 viser antall gyldige svar og informasjon om saksbehandlerens erfaring og hvor ofte de jobber med saksbehandling fordelt på hvilken bydel respondentene jobber for.

Det har ikke lyktes å oppdrive eksakt antall årsverk i bydelene som driver med saksbehandling. Det er samtidig anslått at populasjonen, det vil si antall saksbehandlere totalt i Oslo kommune, ligger på rundt 260. Anslaget er basert på et skalert gjennomsnitt mellom årsverk i bydel Nordstrand (15 saksbehandlere) og bydel Vestre Aker (20 saksbehandlere) da disse bydelene oppgav antall saksbehandlere i kommunikasjonen om denne masteroppgaven. Det er dog usikkerhet knyttet til anslaget, da antall saksbehandlere varierer mellom bydeler. Informant BVA beskrev dette i en epost:

Dette er vanskelig å svare på vegne av alle bydelene, det er avhengig av demografien i den enkelte bydel, dvs. f.eks. avhengig av hvor stor eldrebefolkning man har, hvor stor problematikk den enkelte bydel har med psykiatri. Noen bydeler, som Vestre Aker, har egne forvaltningsenheter hvor saksbehandlerne er samlet. Andre bydeler har saksbehandlerne mer «spredt» i nærmere tilknytning til de utøvende tjenestene, dvs. at de kan ha delte stillinger hvor de både saksbehandler og er utøvende. Vestre Aker har ligget opp mot ca. 20 årsverk (inkludert kontraktsforvaltere), men vi ligger nok blant de som har flest pga. høyt antall eldre.

Bydel Nordstrand opplyste at de har 15 saksbehandlere de sendte undersøkelsen til.

3.3 Databehandling

Databehandlingen startet med dataen og lydfilene fra ekspertintervjuene i del 1. Informantene samtykket til at det ble gjort opptak av intervjuene. Lydfilene ble transkribert ved hjelp av UiOs program Autotekst, som benytter Whisper fra OpenAI for å trygt transkribere opptak (Universitetet i Oslo, u. å.). Deretter ble transkripsjonen språkrenset manuelt og klargjort for analyse.

Dataen fra del to inkluderte både kvalitative data (fritekst-/utdypingsspørsmål) og kvantitative data (rangerings-, flervalgs- og kvantifiseringsspørsmål). Databehandlingen innebar rensning av spørreskjemasvar, slik at ugyldige svar ble tatt ut. Dataen ble deretter eksportert til eget regneark, og klargjort til modellering og statistisk analyse i Excel.

Av GDPR- og personvern hensyn var det viktig at all informasjon fra den anonyme

spørreundersøkelsen (del 2) som kan bidra til å identifisere enkeltinformanter, ikke vises i denne rapporten. Dette gjelder også data fra de transkriberte ekspertintervjuene (del 1) som måtte navngi eller på annen måte identifisere enkeltansatte og brukere som ikke har akseptert å delta i forskningsprosjektet. Data som ikke var forenelig med personvern- og anonymitetshensynet ble derfor fjernet eller redigert på en slik måte at dette ble ivaretatt.

På oppdrag fra NMBU har personverntjenestene ved Sikt – Kunnskapssektorens tjenesteleverandør, vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

3.4 Modellering

Dataen ble modellert ulikt basert på hvilken type den var. Som redegjort for tidligere var det kvalitative tekstdata fra de transkriberte ekspertintervjuene i del 1. Denne dataen ble modellert for å gjennomføre tematisk analyse. Den kvantitative dataen fra spørreundersøkelsen i del 2 ble modellert for blant annet deskriptiv analyse. Modelleringen til disse analyseformene beskrives i dette delkapittelet.

Kvalitativ data - tematisk analyse

Hovedtyngden av den kvalitative delen av modelleringen er basert på prinsippene for tematisk analyse slik de er beskrevet av Braun & Clarke (2013). Denne analyseformen ble brukt for å bearbeide de transkriberte ekspertintervjuene med de fem utvalgte informantene. Et mindre arbeid ble gjort med fritekstsvarene fra spørreundersøkelsen (del 2 av datainnsamlingen).

Den tematiske analysen av intervjuene ble gjort for å besvare forskningsspørsmål 1: *Hvilket potensial er det kommunens ledende fagpersoner identifiserer for LLM i saksbehandlingen?* Analysen hadde en induktiv inngang der temaer er utviklet fra datamaterialet selv, heller enn å påtvinges eksisterende teorier eller konsepter. Koding og temaidentifikasjon ble således nært knyttet til informantenes egne perspektiver og uttrykte meninger.

Analysen ble gjennomført med en *complete coding*-tilnærming, der alle meningsbærende elementer i datamaterialet ble kodet. Dette betyr at kodingen ble utført bredt og inkluderende i de første fasene, før kodene senere ble gruppert og overført til temaer i tråd med Braun og Clarkes prosess. Dataen er modellert for å gi det Braun & Clarke (2013, s. 252) beskriver som en «descriptive and often essentialist form of analysis». Målet er å gi en troverdig fremstilling av informantenes erfaringer og oppfatninger med LLM.

Temautviklingen (utviklingen av hovedfunn) har fulgt prinsippet om at temaer skal

være meningsbærende, analytiske enheter, og ikke kun kategoriske oppsummeringer. Temaene som er presentert i analysen søker derfor å fange mønstre av betydning i datamaterialet, og å gi en samlet og koherent fremstilling av svar på forskningsspørsmål 1.

Datautdragene (sitater) er brukt hyppig i analysen for å forankre påstander underveis og koble analysen tett til dataen den er basert på. Dette er i tråd med Braun & Clarke (2013) sin beskrivelse av en illustrativ presentasjon av resultater.

Fritekstsvarene fra spørreundersøkelsen til saksbehandlere (del 2 av datainnsamlingen) ble gjennomgått manuelt og bearbeidet i mindre grad enn den kvalitative dataen fra del 1. Modelleringen av disse relativt få svarene begrenset seg til manuell gruppering av like svar og telling av antall forekomster på hvert svar, slik at dette kunne presenteres i frekvenstabeller.

Kvantitativ data - deskriptiv analyse

Formålet med modelleringen av denne delen var å gjøre deskriptiv analyse av spørreskjemadata for å belyse forskningsspørsmål 2: *Hvilke utfordringer finnes i dagens saksbehandling og hva kan LLM bidra med ifølge saksbehandlerne selv?*. Deskriptiv analyse dreier seg om å beskrive hvordan faktum eller dataen faktisk er, uten å analysere eller gi forklaringer på hvorfor dataen er slik den er (Stoltenberg, 2024). Den deskriptive analysen av dataen i dette prosjektet søker derfor å beskrive relevante funn fra saksbehandlerne svar på en etterrettelig og datanær måte.

Rådataen ble eksportert fra det elektroniske spørreskjemaet i Microsoft Forms til analyse i Excel. I Excel ble kontinuerlige data analysert med sentralt mål som gjennomsnitt, median og kvartiler og spredningsmål som standardavvik og interkvartilbredde for å gi deskriptiv statistikk fra svarene på disse spørsmålene. Kategoriske data ble analysert med blant annet frekvenstabeller. Beregningene ble deretter plassert i tabeller og grafer etter hva som var hensiktsmessig for å presentere funnene på en oversiktlig måte. Enkelte kategoriske resultater ble i tillegg beregnet per bydel, slik at geografiske forskjeller i Oslo kommune kom til syne.

Data strukturert etter Likert-skalaen var fremtredende i datamaterialet, da denne måten å samle inn data fra respondenter er benyttet i spørreskjemaet. Likert-skala er en skala utviklet for at respondenter i spørreundersøkelser skal ta stilling til ulike typer påstander (Malt & Grønmo, 2025). Respondentene gjorde dette ved å gi påstanden en score 1-5, med et nøytralt midtpunkt, og ulike grader av enighet per tallscore. Likert-skalaen er en skala på ordinalnivå som skiller seg fra enkle skalaer med kategoriske svaralternativer og skalaer med kontinuerlige svaralternativer, ifølge Malt & Grønmo (2025). Dataen ble analysert ved at relativ fordeling per score ble beregnet og gruppert, der dette var hensiktsmessig, i tre deler; «uenig» (score 1-2), nøytral (3) og enig (4-5). Den relative fordelingen på hver påstand ble deretter

presentert på hensiktsmessig måte, i tabell- eller figurform.

Analysen av det kvantitative datamaterialet innebar også en beregning av ekvivalent kostnadsbesparelse. Beregningen ble gjort på bakgrunn av respondentenes estimerte tidsbesparelse ved bruk av LLM til oppgaver de selv mente teknologien kunne bidra med. Svardataen på dette spørsmålet var en kontinuerlig fordeling mellom 0 og 100 prosent tidsbesparelse. Sentralt mål og spredningsmål på dataen for tidsbesparelse ble brukt for å beregne ekvivalente mål for kostnadsbesparelse ved tilsvarende reduksjon av årsverk som tidsbesparelsen tilsa. SSBs tall for totale arbeidskraftkostnader per årsverk i helse- og sosialtjenester ble brukt for å utlede den økonomiske gevinsten. Kostnadsbesparelsen ble beregnet for utvalget som svarte på undersøkelsen isolert (n=101) og populasjonen (n=260) i Oslo kommune totalt.

3.5 Evaluering

3.5.1 Studiens reliabilitet

Et sentralt moment i vurderingen av den beskrevne metoden er studiens reliabilitet, altså hvorvidt man vil få like resultater om man gjennomfører studien flere ganger. I dette spørsmålet ligger mye av svaret i datainnsamlingen og hvorvidt informantene som er plukket ut gir svar som er dekkende for det andre, ledende fagpersoner vil si om tematikken, og hvorvidt respondentene på spørreundersøkelsen kan sies å være representative for populasjonen av saksbehandlere i Oslo, som er anslått til rundt 260 totalt i bydelene.

For å ta førstnevnte først, så er det relevant å vurdere hva slags type informanter som er plukket ut, ref. kap. 3.2.1 Kvalitativ del. Informantene er ledende fagpersoner og samtlige kommunale virksomheter som er vurdert som relevante for problemstillingen er representert i utvalget. På teknologisisiden er begge kommunens digitaliseringsetater representert og avdelingsdirektøren for fagsystemene i Helseetaten i tillegg. På den helse- og sosialfaglige siden er både en representant for bydelene og avdelingsdirektøren for folkehelse- og omsorgsavdelingen i Helseetaten representert. I sum må det vurderes som at denne delen av studien bidrar til høy reliabilitet da alle de relevante virksomhetene er involvert. Dette har ett unntak, som er bydelene, der kun én av femten bydeler er representert i ekspertutvalget. Dette åpner for at det kunne blitt ulike resultater om en annen bydel hadde blitt representert i ekspertutvalget. Noe av dette veies opp ved spørreundersøkelsen til saksbehandlere der nesten samtlige bydeler er representert.

Spørreundersøkelsen er som nevnt besvart av saksbehandlere fra i alt 13 av Oslo kommunes 15 bydeler. Den høye andelen bydeler som er representert bidrar til å styrke studiens reliabilitet og reduserer risikoen for at bydelsvis forskjeller ikke

fanges opp i datamaterialet. Samlet antall respondenter med gyldige svar er 101 og anslått totalt antall saksbehandlere i bydelene er 260. Dette gjør at svarprosenten er omlag 39 prosent. At en relativt høy andel saksbehandlere har valgt å svare på denne typen anonym og frivillig spørreundersøkelse er positivt, spesielt når mange har en travel arbeidshverdag. Andelen gir et godt grunnlag for å analysere overordnede trender og holdninger blant saksbehandlerne.

For at man med statistisk sikkerhet skal kunne si at et flertall (over 50 prosent) av alle saksbehandlerne i Oslo kommune mener én ting fremfor noe annet, må antallet respondenter som har besvart det aktuelle spørsmålet være 59 eller høyere, gitt et konfidensintervall (CI) på 95 prosent. Tabell 7 viser utregningen for det laveste konfidensintervallet (som er gitt ved 59 svar) som overstiger 0,5 og dermed betyr flertall i populasjonen. For mindre utvalg enn $n=101$ og i spørsmål der få har svart må resultatene tolkes mer forsiktig, og heller forstås som indikasjoner på tendenser. Likefullt vurderes det til at spørreundersøkelsen samlet bidrar, med sin høye bydeldekning og gode svarprosent, til å forsterke studiens reliabilitet.

Tabell 7: Beregning av nødvendig antall svar for å gi statistisk signifikant flertall i populasjonen med konfidensintervall på 95 %.

Beskrivelse	Verdi	Formel
Svar ja (j)	59	
Utvalg (n)	101	
Andel ja (p)	0,5842	$p = \frac{j}{n}$
Populasjon (N)	260	
Konfidensnivå 95 % (z)	1,96	
Finite Population Correction Factor (FPC)	0,7835	$\sqrt{\frac{N-n}{N-1}}$
Standardfeil for p (SE), med FPC	0,0384	$\sqrt{\frac{p(1-p)}{n}} \cdot \text{FPC}$
Nedre konfidensgrense	0,5088	$p - z \cdot \text{SE}$
Øvre konfidensgrense	0,6595	$p + z \cdot \text{SE}$

3.5.2 Studiens validitet

Også vesentlig for å vurdere studiens metodiske kvalitet er hensynet til validitet, altså hvorvidt dataen som er samlet inn er adekvat for å belyse problemstillingen og tematikken. I denne vurderingen er det et relevant forhold at det i dette masterprosjektet ble lagt opp til nokså omfattende datainnsamling, der data fra to ulike datakilder ble samlet inn og vurdert. Dette bidrar til å styrke studiens samlede validitet fordi man har hatt en større mengde og mer variert data å belyse problemstillingen med.

Selv om to ulike datasett bidrar til styrket validitet, er det antagelser og forutsetninger hos informanter og respondenter som kan bidra til å svekke validiteten. Det er viktig at informanter og respondenter har et noenlunde omforent syn på hva teknologien LLM er og hvilken type LLM det i realiteten er aktuelt å vurdere for saksbehandlingen. Det er eksempelvis mye forskning på grunnmodellers prestasjoner, og det er primært grunnmodeller som ChatGPT, CoPilot og DeepSeek som har vært hyppigst omtalt i media. Det må derfor legges til grunn at deltagerne i prosjektet kan tenke på ujusterte grunnmodeller som «fasit» på hva LLM er, og ikke domenespesifikke modeller, slik de er omtalt i teorien.

Det skal heller ikke legges skjul på at denne allmenne forståelsen av LLM har blitt utnyttet i masterprosjektet. I arbeidet rettet mot saksbehandlerne ble sågar begrepet «kunstig intelligens» benyttet som forenkling i spørreundersøkelsen, selv om det eksplisitt ble beskrevet at det med KI menes store språkmodeller (jamfør vedlegg B). Balansen har stått mellom å gi omfattende bakgrunnsinformasjon og teori om LLM, og for eksempel beskrive forskjellen på domenespesifikke modeller og grunnmodeller, mot viktigheten av at spørreundersøkelsen var kort og enkel. I datainnsamlingens del 1 ble det tydelig formulert fra informantene at om det var et ønske om høy svarprosent, måtte undersøkelsen være kort. Dette ble derfor lagt til grunn i utarbeidelsen av spørreskjema. Man må av nevnte grunner anta at en del saksbehandlere har tenkt på, og gitt svar basert på, en overbevisning om at LLM er avgrenset til de kjente grunnmodellene. Dette bidrar til å svekke studiens validitet og er en feilkilde.

3.6 Store språkmodeller som hjelpemiddel

Ettersom denne masteroppgaven dreier seg om anvendelser av store språkmodeller, er det naturlig at LLM-verktøy også er testet og benyttet i planleggingen og utføringen av masterarbeidet. Det kjente og kommersielle LLM-verktøyet *ChatGPT*, fra selskapet OpenAI, er benyttet som hjelpemiddel. Verktøyet er brukt til idégenerering, forslag til strukturering av kapitler og tekst, teorisøk, generering av LaTeX-kode for referanser og tabeller og generell sparring om tematikken. Versjonen som er benyttet som støtte er betalingsvarianten ChatGPT 4o. Fra Europa-lanseringen i slutten av februar 2025 er ChatGPTs funksjon «Deep Research» benyttet. Her resonnerer modellen mer enn normalt, som skal bidra til mer reflekterte og presise svar (OpenAI, 2025).

I tillegg er Googles *Notebook LM* benyttet i mindre grad. Her har bruken begrenset seg til å oppsummere relevante kilder og teori. Disse oppsummeringene har blitt presentert i både skriftlig form og i podcast-form, kalt «Dyp samtale». Hensikten har vært å tilegne seg et overblikk over store mengder teori på en auditiv måte (Google, 2024).

Det har vært et bevisst valg å bruke, utfordre og utforske LLM-verktøy i masterprosessen. Dette har imidlertid blitt gjort med stor varsomhet og bevissthet om teknologiens begrensninger og feilkilder. Faktaopplysninger er dobbeltsjekket manuelt, og tekst er ikke avskrevet direkte. NMBUs reglement for bruk av KI legger føringer for hvilken bruk universitetet aksepterer fra studenter (NMBU, 2025). Dette reglementet, sist oppdatert 28. januar 2025, er fulgt i denne masteroppgaven.

4 Resultater

4.1 Kvalitativ del

I denne delen av resultatkapittelet blir funnene fra ekspertintervjuene, del 1 av datainnsamlingen, presentert. I den tematiske analysen er funnene knyttet tett til informantenes faktiske utsagn, og sitater er brukt gjennomgående for å dokumentere og begrunne funn. Hovedfunn er presentert gjennom syv «temaer». Tabell 8 viser innholdet i analysen og koblingen mellom temaer og to overordnede temaer; kommunens potensial og teknologiens potensial. I det videre blir temaene presentert i den oppgitte rekkefølgen.

Tabell 8: Tematisk inndeling av funn fra ekspertintervjuer.

Overordnet tema	Tema
Kommunens potensial	1. Det er begrenset informasjonsdeling og kommunen er i liten grad datastyrt
	2. Mye av saksbehandlingen går til arbeid med tekst
	3. Skjønnsvurderinger og menneskelig dialog i saksbehandlingen er nødvendig
	4. Ansatte i kommunen har varierende kompetanse og ledelsen må ha ansvaret
Teknologiens potensial	5. Fagsystemet begrenser teknologiske innovasjoner
	6. Algoritmer kan egne seg bedre enn eksisterende, ujusterte LLMer
	7. Det er bekymringer om rettferdighet og bærekraft i LLM

4.1.1 Kommunens potensial

Dette overordnede temaet omfatter organisatoriske forutsetninger og forhold i kommunen som påvirker hvilket potensial som finnes for LLM i saksbehandlingen. I dette ligger informantenes beskrivelser av kommunens digitale modenhet, dagens saksbehandlingsprosess, forhold knyttet til ansatte og konkrete barrierer og potensial for LLM i Oslo kommune. Fellestrekket ved temaene som er identifisert i denne delen av den tematiske analysen er at de tar for seg kommunen som organisasjon og potensialet som ligger der.

1. Det er begrenset informasjonsdeling og kommunen er i liten grad datastyrt

Et gjennomgående tema fra informantene er at det er relativt begrenset datadeling mellom kommunen og eksterne instanser, internt i kommunen og mellom fagsystem og ulike datakilder. Dette kommer til uttrykk ved at informanter forteller om at data ofte må flyttes manuelt fra et sted til et annet, uten at det er noe automasjon involvert. Informant BVA, som har sin erfaring fra bydelssektoren, forteller om dette i sin beskrivelse av en typisk saksbehandlingsprosess:

Da skriver man jo på nytt inn opplysningene fra de du har hentet inn opplysninger fra. Altså, du gjenforteller hva doktoren har skrevet, kanskje, deler i det, og andre skriftlige ting du har fått. Og muntlige, selvfølgelig. Så gjør du en vurdering i en journal, og så skal du ha dette over etter hvert i et vedtak.

Informant OOO, som jobber med digitalisering i kommunen, fremhever at det ligger et effektiviseringspotensial i å automatisere informasjonsflyten i dagens saksbehandling, men at det ikke nødvendigvis er LLM som løser dette:

Så tror jeg det er noe annet som også vil spare tid, som jeg ikke tror LLM er svaret på. Men det er å hente inn dokumentasjon innbyggeren må levere fra seg. Som vi i liten grad har automatisert i dag. Du trenger bostedsattest for noen tjenester. Du trenger en legeattest for noen ting. Disse tingene burde maskinene bare begynne å hente ut når noen ber om et eller annet.

Denne begrensede informasjonsdelingen som informantene problematiserer medvirker også til at flere vurderer kommunen som generelt lite datastyrt. At kommunen i for liten grad utnytter dataen til teknologisk innovasjon og analyse for å bedre tilpasse tjenester og tilbud. Informant OOO beskriver situasjonen:

Jeg tenker at kommunen fremdeles har en ganske lang vei å gå i å hente de gevinstene det er mulig å hente med teknologiutviklingen. Vi har ikke kommet langt nok med data, data governance, til egentlig å få nytte av ting som store språkmodeller.

Og informanten legger til:

Hver bydel har rundt 1000 forskjellige tjenester, hvis vi regner med sosiale og tjenester til barn osv. Det er vanskelig for meg å se for meg at ikke noen av de er mer verdifulle og gir bedre helseutfall enn andre. Men vi vet ikke.

Liten grad av datastyring i kommunen kan også ha gjort at enkelte digitaliseringsinitiativ faktisk har forverret kommunens prosesser, heller enn å forbedre dem. Informant UKE, som blant annet jobber med systemarkitektur og infrastruktur, beskriver digitaliseringsprosjekter som bidrar til å forsterke eksisterende, dysfunksjonelle prosesser i kommunen:

Umodenhet knyttet til digitalisering og manglende helhetlig styring har også bidratt til å sementere dysfunksjonelle prosesser i negativ forstand. Så det har ikke bare vært positive forhold knyttet til digitalisering.

Et sentralt poeng i forbindelse med kommunens evne til å styre etter data er at virksomhetene ikke nødvendigvis rapporterer og registrerer data likt. Dette medfører at potensialet for analyse og beslutningsstøtte blir mer krevende, og at datastyringen må foregå med flere forbehold om dataens sannhetsgehalt. Informant UKE forklarer dette som at man stadig sammenligner «epler og pærer» når kommunen vurderer data fra bydeler og andre virksomheter:

Men det er likevel forskjellige virksomheter som har egne områder eller forskjellige måter å bruke disse funksjonene på. Det betyr at de også lagrer data [...] som tilsynelatende er samme informasjonsverdi, men som brukes på forskjellig måte. Det betyr at når man sammenligner to forskjellige virksomheter, så er det vanskelig å se på dataene som ligger i ett og samme system fordi de brukes på forskjellige måter. Man sammenligner epler og pærer.

Samtidig er et fellestrekk i informantenes uttalelser at de ser at kommunen er i en endringsprosess for å imøtegå kritikken om dårlig informasjonsflyt og liten grad av datastyring. Informant HEL2, som jobber med fagsystemene i helsesektoren, beskriver det som et paradigmeskifte at man nå beveger seg fra systemforvaltning til informasjonsforvaltning.

Den viktigste endringen vår, som nesten er et paradigmeskifte, er at vi har endret fokus fra systemforvaltning til informasjonsforvaltning. At vi ser på det som den viktigste delen av jobben vår å forvalte informasjon. Det å sørge for at data er tilgjengelig der det trengs, når du trenger det. Og at vi selvfølgelig da styrer den i forhold til tjenestlige behov. At vi har kontroll på hvilke data, hvilke verdier, hvem har tjenestebehov osv. Som er løsrevet fra enkeltapplikasjoner til å se på det mer som datasett.

2. Mye av saksbehandlingen går til arbeid med tekst

Et annet sentralt funn i materialet er at mye av dagens saksbehandlingsprosess innebærer både lesing og produksjon av store mengder tekst. Lesingen bunner i saksbehandlerens behov for å samle inn og vurdere store mengder skriftlig dokumentasjon, som journalinformasjon og attester, i tillegg til at de må ha forståelse for og anvende lovverk, skriftlige rutiner og standarder på en korrekt måte. Produksjonen av tekst inkluderer skriftlig pasientkommunikasjon og formulering av vedtaksbrev - dokumentet som beskriver hvilke helse- eller omsorgstjenester innbyggeren har rett på.

Flere informanter viser til at den store graden av tekstarbeid indikerer et potensial for effektivisering og forbedring. Informant BVA slår fast at skrivedelen er der det største potensialet for LLM ligger, men at det ligger et potensial i lese- og dokumentasjonsdelen av saksbehandlingen også:

Helt klart skrivedelen. Den er innlysende. Men den kan òg være til å forme, samle informasjon. Altså vurderingsgrunnlag som vil kunne [bidra til] det du spurte om, effektivisering, likebehandling og kvalitet. De tre vil være løst, kunne løses altså, i hvert fall der det finnes opplysninger i systemet fra før av.

Også informantene med digitaliseringsbakgrunn var enige i at betydelige deler av saksbehandlingen går til teksthåndtering. Og flere peker på et stort gevinstpotensial ved å bruke LLM til å plukke ut relevant informasjon fra store mengder skriftlig data. «Mitt inntrykk er at saksbehandlere må bruke mye tid på å lese gjennom dokumentasjon for å finne ut av situasjonen til innbyggeren, oppfyller den de konkrete kriteriene osv. Og det må pløyes gjennom mye tekst for å finne svaret på det» sier informant OOO om saksbehandlingsprosessene i kommunen idag. Informanten forklarer potensialet for LLM i en teksttung saksbehandling slik:

Det vi har litt erfaring med, er jo å tenke på hva trenger saksbehandlerne der ute. Og de trenger beslutningsstøtte. [...] Og da mener jeg ikke at en LLM skal ta beslutningene. Men de trenger støtte til å løfte fram de kritiske momentene i en sak. De trenger antakeligvis støtte til å trekke ut essens av en mengde dokumenter. Så i stedet for å lese, la oss si, 40 sider med et eller annet om et individ som trenger hjelp til noe. Å kunne stille spørsmålet: Har personen A, B og C? Og få svaret på det kjapt. Det trenger saksbehandlerne. Men vi har ingen praktisk erfaring med å gjøre noe sånt.

Kommunen er også igang med LLM-prosjekter innenfor informasjonsinnhenting, som ved intervju tidspunktet var på vei til å lanseres. Informant UKE beskriver at fundamentet, «byggeklosser», for tekstanalyse og smartsøk ved hjelp av LLM er etablert

i kommunens felles IKT-plattform, og at det allerede er utviklet en sluttbrukertjeneste for å håndtere store mengder tekstdata i plan- og byggesaker. Informanten beskriver prosjektet slik:

Så vi har nå [...] en sluttbrukertjeneste som heter Inspektørinnspill, som analyserer alle høringsinnspill. Den er utviklet nå for plan- og bygnings-etaten, men dette er en modell som er reell utenfor Oslo kommune og på andre områder enn høringsinnspill for plan og bygg. Her har vi allerede sett at man kan spare betydelig tid. Opptil tolv årsverk er Plan- og bygningsetatens vurdering. Og da er det ikke for å si opp disse ansatte. Men for at de ansatte skal bruke tiden sin på kvalitativt arbeid, og ikke lese 4000 sider og flytte tekst, som språkmodellen faktisk gjør bedre.

Informanten poengterer også at infrastrukturen er bygget opp slik at den er LLM-nøytral, og kan bytte til en norsk språkmodell med «gjennomgående transparens til data den trenes på» når dette i fremtiden blir tilgjengelig.

Når det er snakk om å generere forslag til tekst, som vedtaksbrevene, er det på den annen side mer divergerende syn blant informantene. Flere uttrykker bekymringer knyttet til bias og transparens, som beskrives ytterligere i tema 7 *Det er bekymringer om rettferdighet og bærekraft i LLM* i denne analysen. Men at det er mulig å anvende LLM til å lage forslag til for eksempel vedtaksbrev er på det rene, selv om gode og anvendelige dokumentmaler kan være vel så hjelpsomme, ifølge informant OOO.

Og så er det mange som snakker om å skrive vedtaksbrev. At du kan bruke det til det. Det er klart du kan det. Det kan du også ha maler for uten at du trenger LLM til det. Og det har de jo i dag. Mitt inntrykk er at de er litt sånn vanskelig tilgjengelig å bruke. Fordi gammelt fagsystem og sånn. Men så jeg vet ikke hvor mye bedre det blir med LLM hvis de malene hadde vært skikkelig gode i utgangspunktet, da.

Mange informanter trekker også frem et potensielt bruksområde for LLM i tekstøymed som er oversettelse til andre språk og til enklere språk. Informant BVA trekker frem det vedkommende beskriver som en «jussifisering» av vedtakstekstene, der lovhenvisninger og komplisert språk blir benyttet i større grad enn før - som gjør det vanskelig for utsatte innbyggere å håndtere teksten. Informant OOO trakk samme problem frem i sitt intervju:

Men der jeg tror dette har et potensial som jeg ikke syns folk snakker så mye om, er det å skrive god og lesbar tekst for utsatte brukergrupper som kanskje ikke har så høy systemkompetanse om det offentlige systemet.

Og ikke minst for å kunne både lese og skrive tekst på andre språk. Vi har veldig mange språk i Oslo kommune, men saksbehandlerne våre bruker norsk. Nettsidene våre og søknadsskjemaene våre er på norsk. Å få klarspråklige brev tilbake på det språket du har valgt, kunne dette vært bra.

Informant HEL2 konkluderer på samme måte: «De er jo blitt bedre og bedre på å oversette disse modellene. Så den type oversettelse er selvfølgelig et kjempepotensial».

3. Skjønnsvurderinger og menneskelig dialog i saksbehandlingen er nødvendig

Det tredje temaet og hovedfunnet er at det menneskelige elementet av saksbehandlingen er vesentlig, og at en del prosesssteg ikke nødvendigvis kan automatiseres. Mer konkret handler dette om at informantene beskriver at skjønnsvurderinger i saksbehandlingen og menneskelig kommunikasjon med innbyggere er nødvendig. Særlig det førstnevnte kan ha sine negative implikasjoner, da etterprøvbarhet og transparens blir svekket i saksbehandlingen. Dette slår ifølge informantene også ut i at det er til dels store variasjoner i hvordan regelverket praktiseres og hvilket vedtaksnivå kommunen ender med å gi til innbyggere. Informant HEL1 i Helseetaten forklarer at kommunen skal foreta en skjønsmessig helhetsvurdering og at dette er lovpålagt, og dette kan gi utslag i til dels svært ulikt vedtaksnivå fra innbygger til innbygger. Som eksempel trekkes det frem at samme dokumentasjon potensielt kan føre til betydelig høyere vedtaksnivå fra en bydel til en annen:

Brukerstyrt personlig assistanse er et godt eksempel på at i én bydel så får du noen få timer, og så får du flere ganger så mange timer i neste bydel basert på samme dokumentasjon. For det er en sånn vurdering, det er et skjønn. Men så stor ulikhet som vi ser, det er jo ikke riktig, da.

Årsakene til den voldsomme potensielle forskjellen på vedtak i en sak med samme dokumentasjon kan ifølge informant BVA komme av at man i saksbehandlingen vurderer mer enn kun diagnose og funksjonsevne hos innbyggeren. Rent personlige preferanser og personlighet kan også ha en innvirkning på hvilket vedtaksnivå som er riktig i en sak.

Hvis du går mye på funksjon, går du mye på diagnose, og tar den biten, så kan samme diagnose ha et ulikt funksjonsnivå. Eller folk må behandles litt ulikt fordi vi er forskjellige. Selv om vi har de samme sykdommene eller utfordringene. Som personlighetstypen meg, så er jeg en som helst

vil kare meg til toalettet, krype ned i stedet for å ta imot hjelp. Og så er det noen andre som ikke foretrekker det. Og da er det en likebehandling faktisk å gi ulik hjelp til de ulike personlighetstypene, mener jeg. Derfor får du ikke likebehandling i at kanskje det du trodde var samme problematikk, fordi innbyggeren skal kunne påvirke dette selv.

Denne skjønnsmessige delen av saksbehandlingen har som nevnt sine negative implikasjoner, og flere informanter trekker frem at det er krevende å redegjøre for hvilke konkrete avveininger som er gjort i en skønnsvurdering. Dette påvirker også potensialet for hvilke teknologiske verktøy man kan bruke, ettersom stor grad av skønnsvurdering ikke er forenelig med en stor grad av regelbaserte løsninger, som eksempel. Informant OOO i digitaliseringsetaten Origo utbroderer dette:

Det brukes utrolig mye skjønn i det å fatte vedtak. Mye er ikke regelbasert. Ting som er regelbaserte, det hadde vi ikke trengt LLM til heller. Du kunne hatt en god, gammel algoritme og en regelmotor og tatt de beslutningene. Men i dag så, det er i hvert fall det jeg blir fortalt, så bruker de så mye skjønn at de klarer nesten ikke selv å redegjøre for hva [...] reglene er som gjør at de lander på den beslutningen de gjør. Jeg tror at hadde vi hatt en god regelmotor, et tydelig regelverk for hva som er forsvarlig i hvilket tilfelle, og LLM til å forfatte og lese gjennom tekst, det er bra. Det kunne spart mye tid.

I dette hovedfunnet ligger det at menneskelig involvering er viktig også utenom de skjønnsmessige vurderingene. Informanter trekker frem at dialogen med innbyggere, mange av dem utsatte og med dårlig helse, er en viktig del av saksbehandlingen. Informant BVA, som er leder for bydelens saksbehandlere, poengterer at hjemmebesøk og fysisk dialog med innbyggere er viktig for å identifisere reelle pleiebehov.

Ideelt sett så skal vi jo være på et hjemmebesøk eller et vurderingsbesøk hos alle nye. Som er min kjepphest. For å få tatt de riktige vurderingene så er det lurt å ha sett pasienten. Tatt noen funksjonstester kanskje selv. Derfor man tenker at ellers har jeg ikke trengt å ha helsepersonell her.[...] Det er jo nettopp den undersøkelsen og å få avdekket symptomene som krever det menneskelige, så å si. Den vurderingen de gjør, er jo å få avdekket hva som egentlig er problemet ditt.

Men informanten erkjenner også: «Men hvis noen svarer helt ærlig og redelig hva som er problemet, og du skanner kroppen og alt det, så er det klart maskinene vil gjøre det bedre».

4. Ansatte i kommunen har varierende kompetanse og ledelsen må ha ansvaret

Det fjerde hovedfunnet peker på store variasjoner i kunnskapsnivå om LLM blant kommunens ansatte og at ansvaret for implementering av ny teknologi må være tydelig plassert hos ledelsen. Flere trekker frem at mange er svært optimistiske til hva LLM og KI kan bidra med i kommunen, uten å nødvendigvis ha tilstrekkelig med kunnskap om teknologien og potensialet i Oslo kommune. Informant 000 beskriver lederes og ansattes kompetanse på LLM slik:

Jeg tror mange er veldig optimistiske til hva vi kan få til. Spesielt på ledernivået rundt omkring så er man veldig optimistisk. Og det syns jeg er bra, at vi er teknologioptimistiske. Og så, selv om optimisme er bra, så tror jeg kanskje at mange tror at dagens LLM-modeller er mer enn de er. At de reelt sett analyserer faktum i dataene, ikke bare predikerer tekst. Så der mangler vi litt kanskje på forståelsen av hva du kan anvende dette til. Per i dag.

Informanten understreker imidlertid at den teknologiske utviklingen går fort, og at det vedkommende sier raskt kan endre seg. Informanten er tydelig på at de ansattes kunnskap om LLM ikke er det eneste som betyr noe, men at det også er begrenset modenhet når det kommer til ansattes tillit til LLM idag. Informanten fra den andre digitaliseringsetaten, UKE, deler vurderingen, men legger til at måten kommunen jobber med ansatte og KI på er god:

Modenheten når det kommer til data og informasjonsverdier, den er lav. Men modenheten i forhold til ansatte og hvordan vi jobber med kunstig intelligens sentralt, den mener jeg er høy. Så er det klart at det skal veldig mye mer til når du snakker om ansatte i Oslo kommune, så er det 57 000. Så det er også veldig mange som ikke har peiling på hva dette dreier seg om. Og det er veldig forskjellig hvordan ledere tilnærmer seg dette, for dette er egentlig først og fremst et lederansvar.

Viktigheten av ansatte med god kunnskap om kunstig intelligens og tett involvering i implementering av eksempelvis språkmodeller begrunnes av informant UKE blant annet med at teknologien påvirker alle ansatte, og at det er nødvendig å «mobilisere kreftene i kommunen som helhet» for å lykkes. Informanten mener at Oslo kommune i dette henseendet ligger foran sammenlignbare kommuner.

[...] Det viktig at dette ikke blir et IKT- og digitaliseringsanliggende [...].
For det dreier seg om alle ansatte. Så den tilnærmingen har vi hatt helt

fra slutten av 2023, da vi startet arbeidet med kunstig intelligens i stort. At vi må mobilisere kreftene. Vi må invitere de ansatte til å forstå hva kunstig intelligens dreier seg om.

Flere informanter er altså av den oppfatning av at det er en risiko knyttet til ansattes og kommunens forståelse for hva LLM egentlig er. Informant UKE beskriver situasjonen slik: «Og så tror jeg nok også at det er en risiko knyttet til modenhet generelt. At man ikke kan ta lett på den ansattes forståelse for hva en språkmodell er. LLM-en skjønner ingenting i utgangspunktet». Informanten poengterer videre viktigheten av å ha sikkerhetsmekanismer for å betrygge ansatte om at en potensiell LLM-assistert saksbehandling vil gå riktig for seg:

Så de dataene som genereres, de er dine som person. Så du er kvalitets-sikrer. Og derfor så må det ligge veldig mye sikkerhetsmekanismer i den utviklingen vi gjør på toppen, men det må også ligge prosessuelle sikkerhetsmekanismer. At de ansatte faktisk har kunnskap om hvordan er det dette egentlig fungerer, sånn at man ikke stoler blindt på den kunstige intelligensen.

Flere informanter trekker også frem at det er nødvendig å plassere ansvaret for teknologibruken hos ledelsen, slik at ansatte vil være trygge på at verktøyene de bruker er trygge og grundig vurdert av kommunens ledelse. Informant HEL1 argumenterer for viktigheten av dette i forhold til de ansatte: «så er det jo veldig viktig at det er god ledelse, tett på ledelse, sånn at man opplever at det man gjør, er riktig». Informant HEL2 legger vekt på at de ansatte blir betrygget med grundige analyser og risikovurderinger før LLM eventuelt tas i bruk:

Og så tenker jeg nok at det er veldig viktig at det er godt dokumentert alle vurderingene som er gjort [...]. Sånn at de som skal bruke dette, vet at her har det vært en grundig jobb med å sikre at det ikke skjer noe feil. Og at et ansvar er tydelig plassert i ledelse.

Informantene understreker gjennomgående at tidlig og reell ansattinvolvering er en forutsetning for å lykkes med implementering av teknologi som LLM. Informant UKE formulerte det slik: «Det å involvere virksomheter og ansatte i de fasene i utviklingen av kunstig intelligens, det tror jeg er kjempeviktig for å skape tillit. At man skjønner hvordan det hele henger sammen». Samtidig er det en bekymring om hvordan denne involveringen i praksis blir utført, og informant OOO poengterer at ledelsen eksplisitt må engasjere ansatte på et tidlig stadium for å sikre god nok involvering. Ellers blir de ansattes syn i liten grad hensyntatt, ifølge vedkommende.

Jeg tror ansattes syn i veldig liten grad, med mindre ledelsen målbærer det inn i tidlig fase av at Oslo kommune ønsker å teste ut noe, prøve ut noe, se om noe funker, så tror jeg ansattes syn i veldig liten grad er med i den beslutningen om at dette skal vi sjekke om er noe. Men når det kommer til å innføre ting, så skjer det jo reelt partssamarbeid. Hvor de ansatte skal medvirke i den type beslutninger.

4.1.2 Teknologiens potensial

Dette overordnede temaet omfatter teknologiske forutsetninger og forhold ved LLM og kommunens tekniske fagsystem som påvirker hvilket potensial som finnes for LLM i saksbehandlingen. I dette ligger informantenes beskrivelser av teknologiens styrker og svakheter, beskrivelser av dagens fagsystem «Geric»¹, alternativet med mer regelstyrte algoritmer over LLM og konkrete bekymringer knyttet til blant annet bias, transparens og klimapåvirkning med LLM. Fellestrekket ved temaene som er identifisert i denne delen av analysen er at de tar for seg tekniske forhold og teknologien LLM som sådan.

5. Fagsystemet begrenser teknologiske innovasjoner

Hovedfagsystemet i helse- og omsorgssektoren er et gjentakende tema hos samtlige informanter. Fagsystemet, kalt *Geric*, er der majoriteten av all informasjon og data behandles i sektoren. Det er også her saksbehandlingen gjøres, og systemet benyttes ifølge informant HEL2, som er ansvarlig for fagsystemene i Helseetaten, av 24 000 ansatte i Oslo kommune. Samme informant beskriver at «helse- og omsorgsfeltet i Oslo utvikler seg kontinuerlig, spesielt innenfor lavterskeltilbudene», og at «på grunn av fleksibiliteten til *Geric*, ender det ofte med at nye tjenester legges inn i *Geric*, dersom man ikke har noen andre klare alternativer». Omfanget til *Geric* beskrives av informantene som stort og komplisert. Informant HEL2 skildrer en kompleksitet som både har positive og negative sider ved seg:

Det er en fleksibilitet og en svøpe. Det takler mye, og det krever mye av de som skal bruke det. Og de fleste bruker ett hjørne av *Geric*. Hvert sitt hjørne. Og ser *Geric* som det lille hjørnet de er i.

Mer spesifikt er *Geric* et elektronisk pasientjournalssystem (EPJ), som over tid har fått koblet på en rekke administrative rutiner og APIer for å håndtere diverse oppgaver, inkludert saksbehandling. Og som følge av dette er systemet ifølge informantene relativt dårlig på mange av oppgavene det er satt til å gjøre. Informant HEL2 utdyper:

Det er jo egentlig en EPJ pluss, pluss, pluss. Og er vel egentlig ikke særlig

god på noen av delene. Naturlig nok, når du skal dekke så mange områder. Det dekker alt fra sykehjem til lavterskeltjenester lokalt i bydeler. Saksbehandling, økonomi, refusjon, vederlagsberegning for egenandeler. Både for sykehjem, andre institusjoner og tjenester i hjemmet. Så det er et enormt spekter av tjenester som benytter Gericca.

Kompleksiteten til fagsystemet er altså i seg selv en utfordring. I tillegg er systemet ifølge flere informanter gammelt og utdatert. Men på grunn av kompleksiteten og hvor avhengig ansatte, andre systemer og prosesser i kommunen er av Gericca, er det vanskelig å gjøre tiltak. Det medfører at fagsystemet binder opp kapasitet og hemmer teknologisk innovasjon på feltet. Informant UKE utdyper:

Og det er jo en utfordring som går langt utover helseområdet og Helseetaten nå, at systemene er ikke modernisert. De binder opp teknisk kapasitet på den felles IKT-plattformen, som gjør at vi må ha gammel teknologi f.eks., som vi egentlig har planlagt med å avvikle lenge. Men pga. avhengigheten som disse gamle systemene har, så er vi i en litt sånn krevende situasjon med teknisk gjeld, da.

Det er heller ingenting i dagens saksbehandlingsmoduler i Gericca som gir forslag eller direkte støtte til saksbehandlerne i sin saksbehandling. Det finnes riktignok rutiner som skal bidra til likebehandling, men ikke maskinell hjelp i systemet som sådan. Dette beskrives av informant HEL2 slik:

Man har jo disse byomfattende rutinene som er forankret i «Standard for tjenester i hjemmet». Så i den grad saksbehandlerne kjenner til og benytter den, så er jo det et hjelpemiddel for å behandle like saker likt. Men du har ikke noe maskinell hjelp til å gjøre det. Og det er ikke noen API-er i Gericca som gjør at du kan hente ut informasjon, kna det på en eller annen måte med KI på utsiden, og plassere det tilbake igjen heller.

I forlengelsen av dette er det tydelig at informantene med teknisk ansvarsområde er enige om at en LLM direkte fra markedet ikke uten videre kan integreres i dagens fagsystem. Til det er fagsystemet for utdatert og LLMene på markedet for lite utprøvd og spesialisert. Flere informanter viser imidlertid til at det er mulig å ekstrahere data fra fagsystemene som kan brukes med LLM. Informant OOO forklarer dette slik:

Men min hunch, uten at jeg er teknolog, er at det ikke er mulig å slippe inn LLM i en produksjonsversjon av mange av de fagsystemene. De er

20 pluss år gamle. De tåler en begrenset mengde trafikk. Så jeg tror man måtte ekstrahert dataene.

Utfordringen med fagsystemene er samtidig, ifølge informant UKE, at de skjærer «dataene eller informasjonverdiene litt isolert» og at man dermed ikke ser prosesser i sammenheng. Fagsystemets innretning gjør altså at analyse av data på tvers er krevende. Dette beskrives av flere som en forutsetning for å ta i bruk teknologi som LLM, at man kan analysere, måle og teste med data fra fagsystemene. Informant OOO beskriver situasjonen i tjenestene med fagsystem som Gericas slik:

Tjenestene er preget av gamle fagsystemer hvor vi ikke helt klarer å benytte dataene, som da bor i de fagsystemene, til å faktisk gjøre analyse. Og å få til det, å ha orden på det, tenker jeg er en forutsetning for å både kunne utnytte teknologi som LLM og annen maskinlæring. Til å kunne drive med stordata og faktisk bli en datadrevet virksomhet. Og å kunne drive datadrevet kunnskapsbasert tjenesteutvikling, da. For hvis ikke vi vet om de tjenestene vi har faktisk funker, som jeg ville påstå i mange tilfeller, vet ikke vi om det funker å gi folk avlastning på seniorsenter eller opphold på helsehus. Eller hvordan skal vi da utvikle tjenestene våre på en bærekraftig måte. Det mener jeg personlig er den største utfordringen vi har.

Samtidig slår informantene fra Helseetaten fast at de jobber med å tilgjengeliggjøre data fra fagsystemet, slik at det lettere kan gjøres teknologisk innovasjon og analyse: «Og vi jobber da med å tilgjengeliggjøre dataene i disse store systemene, både for tilpasning av tjenester til den enkelte, til tjenesteutvikling og til styring». Imidlertid, selv med tilgjengelig data, poengteres det at det er behov for å bearbeide dataen før den kan brukes for eksempel til å trene opp en språkmodell. Dette arbeidet beskrives som omfattende, men ikke umulig, av informant OOO:

Det er mulig å hente ut data derfra. Men jeg tipper det trengs en jobb for å gjøre de dataene analyserbare. Det er ikke gitt at de har metadata som gir mening. Det er kanskje usikkerhet om datakvalitet. Så de er ikke selvbeskrevne og klare til å tolke. For det er et ganske gammelt system. Det er ikke sånn man bygde opp databasen og datamodellen. Er det mulig å gjøre den jobben? Ja, det er mulig, med min påstand. Jeg sier ikke det er lett, men det er mulig.

6. Algoritmer kan egne seg bedre enn eksisterende, ujusterte LLMer

Dette temaet fanger blant annet opp informantenes vurdering av at språkmodeller slik de foreligger i dag, ikke uten videre kan brukes i kommunal saksbehandling.

Bakgrunnen for at informantene konkluderer på denne måten er sammensatt, og det vises blant annet til at markedet for spesialiserte (justerte) LLM foreløpig ikke er modent og at det for eksempel ikke er på plass sertifiseringsordninger som sikrer god kvalitet på modellene. Flere informanter trekker frem at det derfor vil være en betydelig jobb å justere og spesialisere en LLM for at den skal ha en bruksverdi i saksbehandlingen, i tillegg til at selve arbeidsprosessene nødvendigvis også må endres for å hente gevinster ved bruk av LLM. Informant OOO beskriver dette slik:

Jeg kjenner ikke til at noen i markedet gjør dette med tanke på saksbehandling i kommunene. Hvis kommunene selv skal i gang med det, så tenker jeg det vil ta litt tid. Fordi organisasjonen finnes ikke. Denne modellen må justeres, trenes. Kanskje vi må få ut noe data først for å gjøre en ordentlig god jobb med det. [...] Vi må bygge litt videre på den modellen selv for å kunne bruke den til dette. Men ideelt sett så designer man jo en helt ny arbeidsprosess som utnytter alle de fordelene du får av teknologien, da. Så det krever jo masse av både ansatte og ledere. Og at teknologien er god nok og klarer å beskrive hva som er mulig.

Flere informanter peker også på at optimismen rundt LLM til dels er overvurdert, og at man ikke foreløpig er der at man kan hente de store gevinstene av bruk av LLM i saksbehandlingen. Til det peker flere på at det på kort sikt er et større potensial i å hente ut gevinster ved å lage klassiske, regelstyrte algoritmer med den strukturerte dataen kommunen har tilgjengelig og automatisere prosesssteg ved hjelp av dette. Informant HEL2 er ganske tydelig på hvor vedkommende mener fokuset bør ligge den nærmeste tiden:

Skal du ta LLM i bruk, så må du ha veldig god kontroll på hva du bruker den til. Og vite veldig godt hva som ligger til grunn for de rådene du får. For å kalle det det. Og hovedbudskapet mitt da er vel egentlig det at jeg synes at LLM-ene er litt vel hypet. Og at vi sitter egentlig på enorme mengder strukturerte data som jeg heller vil at vi skal bruke til å forbedre tjenestene.

Informanten forklarer videre:

Idag vil jeg heller slå slag for de gode, gamle algoritmene. La oss bruke de strukturerte dataene mens man får på plass sertifiseringsordninger og at LLM-markedet bygger seg opp. For det kommer til å komme et marked, fordi det er så hypet, og har så store forventninger knyttet til seg. Men la oss bruke mønstergjenkjenning og de andre tingene som vi har større grad av kontroll på.

Informant UKE viser også til at potensialet ikke kun ligger i LLM alene, men i stor grad også i LLM i kombinasjon med algoritmer:

Og så er det viktig å si at det ikke bare innenfor språkmodeller potensialet ligger. Potensialet ligger kanskje vel så mye i å kombinere språkmodeller med prediksjonsalgoritmer på egne data. For eksempel innenfor bemanningsplanlegging, effektiv bemanningsplanlegging.

Et fellestrekk i materialet er at informantene mener at de delene av saksbehandlingen som ikke innebærer menneskelige skjønnsvurderinger i større grad bør automatiseres. Det trekkes frem at dette også er viktig for at saksbehandlerne får viet tilstrekkelig oppmerksomhet til skjønnsutøvelsen og ikke bruker tid på kriteriestyrte vurderinger som en regelstyrt algoritme like gjerne kan utføre. Informant HEL2 forklarer:

I en del saksbehandling er det jo så å si ikke noe skjønn. Det å få sortert ut det som bare er regelstyrt til noe du automatiserer, på et eller annet vis, hadde beredet grunnen for den skjønnsutøvelsen som skal til. Sørget for at den får tilstrekkelig oppmerksomhet og rom i saksbehandlingen.

Informanten trekker også frem at LLM kan brukes for å tilrettelegge for automatisjon ved algoritmer, ved at en språkmodell kan hente ut og generere strukturert data basert på ustrukturert data som ligger i fagsystemene. Mengden av data i de gamle fagsystemene vurderes som «ekstremt stor», og å kunne konvertere dette til anvendbar strukturert data vurderes som et potensial ved LLM-teknologien:

Jeg har jo også en forventning om at KI kan hjelpe oss til å omdanne brødtekst til strukturert informasjon. Det ligger jo ekstremt store mengder informasjon i de gamle systemene. Hva om jeg hadde klart å konvertere det til SNOMED-CT kategorier, f.eks.? Og på sikt kan forhåpentligvis KI hjelpe oss til å oversette en muntlig konsultasjon til strukturert informasjon i type CT-kategorier.

Flere informanter understreker samtidig at automatisering alene ikke løser alle utfordringer. Der potensialet er størst, er ifølge informantene å kombinere automatisering av kriterierstyrte vurderinger med god beslutningsstøtte til saksbehandlerne i saker som krever større grad av skjønnsvurderinger. HEL2 beskriver dette som det «ultimate»:

Den mest lavthengede frukten er nok automatisering. Mens det ultimate er jo da å kombinere det med beslutningsstøtte. Og vi kommer dit, men vi er ikke der i dag. Men ja, det ligger helt klart et potensial der òg.

7. Det er bekymringer om rettferdighet og bærekraft i LLM

Det siste temaet i analysen dreier seg om informantenes bekymringer for LLMs rettferdighet og bærekraft. Temaet belyser risikoer som må håndteres for å kunne realisere teknologiens potensial i saksbehandling. I rettferdighet ligger det bekymringer om språkmodellens risiko for bias og manglende transparens til treningsdata. Dette medvirker til at flere informanter mener man ikke umiddelbart kan si ja til å bruke LLM til å generere forslag til vedtaksbrev, for eksempel, per nå. På direkte spørsmål om man kunne stolt på et utkast til vedtak fra LLM idag, dersom modellen fikk samme data som saksbehandlerne har, svarte informant HEL2 følgende:

Vi vet jo ikke helt hvilket grunnlag disse LLMene har. Altså hva er treningsgrunnlaget deres. Så jeg vil ikke si umiddelbart ja til det. Jeg trenger at noen har kvalitetssikret de opplæringsdataene som LLMen bruker. Rett og slett også med tanke på faren for type fordommer, at den utvikler fordommer. Men hvis du setter på plass en hel bråte med forutsetninger rundt, knyttet til krav til kvalitet på det meste, og i tillegg har menneskelig kvalitetssikring av sluttproduktet, så ja. Men jeg kan ikke svare ja på et så generelt spørsmål som det du kom med. Det er veldig mange forutsetninger.

Tillits- og maktforholdet mellom en innbygger og kommunen er også vurdert som viktig å balansere i møte med ny teknologi som LLM. Dette gjelder spesielt i forbindelse med bias, som flere informanter er bekymret for. Informant OOO beskriver at de som har kunnskap om bias i LLM har grunn til å være bekymret:

Det er et skjevt maktforhold mellom kommunen og innbyggerne. De som vet noe om bias i modeller og teknologiutvikling, vil nok da være bekymret for at de stiller svakere. Så jeg tenker at både de som kjenner til hvor mye bias det er i treningen av ulike teknologimodeller, og de som føler at de er en svak gruppe, har jo grunn til å være litt redde, da.

I materialet trekkes det også frem at mengden av skjønnsvurderinger, slik det er redegjort for mer inngående under tema 3 *Skjønnsvurderinger og menneskelig dialog i saksbehandlingen er nødvendig*, gjør problematikken med bias viktigere å adressere og LLM vanskeligere å benytte. Informant OOO er bekymret for at saksbehandlere enten vil stole for mye på, eller ikke ha tillit til, en generert tekst fra LLM. Og uansett utfall er bias en bekymring som må svares ut:

Per i dag kan vi ikke bare ta en modell og være sikre på at den vurderer ting uten bias, at den faktisk fanger opp de viktige momentene osv. Så vi

kunne ikke bare erstattet et menneske, en saksbehandler, med en LLM i dag. Jeg tror heller ikke den teknologien ut av boksen er moden nok til at vi kan la den gi beslutningsstøtte, for det tror jeg folk ville lent seg mye på. Eller jeg tror det fins to varianter her. Noen ville ikke stolt på det i det hele tatt og uansett gjort akkurat den jobben de gjør i dag. Mens noen ville tatt alt den sa for god fisk og fattet vedtaket basert på det.

I forlengelsen av dette er informantene opptatt av viktigheten av transparens i språkmodellene. Mer spesifikt hvilke data som LLMene er trent opp på. Dette beskrives av flere som én av de klart største bekymringene knyttet til LLM. Informant UKE er blant disse og beskriver viktigheten av transparens og kvalitet på treningsdata som kjerneutfordringer:

De største utfordringene, det er jo antagelig transparens til data. At vi har ikke transparens til dataene, treningsdataene, grunnmodellene er trent på. Og så har vi jo noe som heter videreutviklet trening og trening som er mer relatert til tjenesteområdene. Det kan vi ha transparens på og vi kan ha kontroll på selv [...]. Men det med datakvalitet, det er nok et sånn typisk utfordringsområde. Datakvalitet og ikke minst sensitive data.

Et fellessyn blant informantene i forbindelse med dette er at denne transparensen er nødvendig for å dokumentere for og betrygge beslutningstagere og saksbehandlerne om at LLM er trygt og rettferdig å benytte i saksbehandlingen. Informant OOO beskriver viktigheten av godt dokumenterte og transparente modeller overfor pliktoppfyllende saksbehandlere:

Det blir ofte vanskelig å få folk til å endre måten de arbeider på, hvis de ikke tror på denne nye tingen. Og mitt inntrykk er at det er veldig mange veldig pliktoppfyllende folk som jobber med saksbehandling, som er opptatt av at det blir riktig. Så jeg tror at det er viktig at de opplever at de vet at den modellen er rettferdig og riktig. At det ikke blir feil og tilfeldig.

Det utvikles for tiden nye grunnmodeller som sannsynligvis vil bedre situasjonen knyttet til transparens. Informant UKE forklarer at det ikke er transparens til flere av de internasjonale, kommersielle grunnmodellene, men at det utvikles norske modeller som bedre kan svare ut bekymringene knyttet til transparens i tiden som kommer:

Vi har ingen transparens, verken til OpenAI eller Meta f.eks. [...] Men når du kommer til de norske grunnmodellene og for eksempel også den som nå er i arbeid hos Nasjonalbiblioteket, så er selve merverdien at vi har eierskap. Vi vet hvilke data grunnmodellene er trent på.

Dette temaet tar også for seg LLMenes egnethet i et bærekraftperspektiv. Språkmodeller krever stor datakraft og er energikrevende, som gjør at teknologien ikke bør benyttes uvettig. Informant UKE er opptatt av denne problematikken:

Det er jo viktig når man snakker om store språkmodeller, at man har forståelse for at det er nevralt nettverk i bunnen. Det er mange lag og dimensjoner. Det har stort avtrykk, stort karbonavtrykk, hvis man bruker store språkmodeller helt blindt.

Og informanten erkjenner at kommunen i testingen og prøvingen av LLM «skyter mye spurv med kanon». Dette kan bedres med mer domenespesifikke språkmodeller, som informanten sier kommunen har planer om å bruke mer av fremover, slik at kommunen ikke bruker unødig med ressurser.

4.2 Kvantitativ del

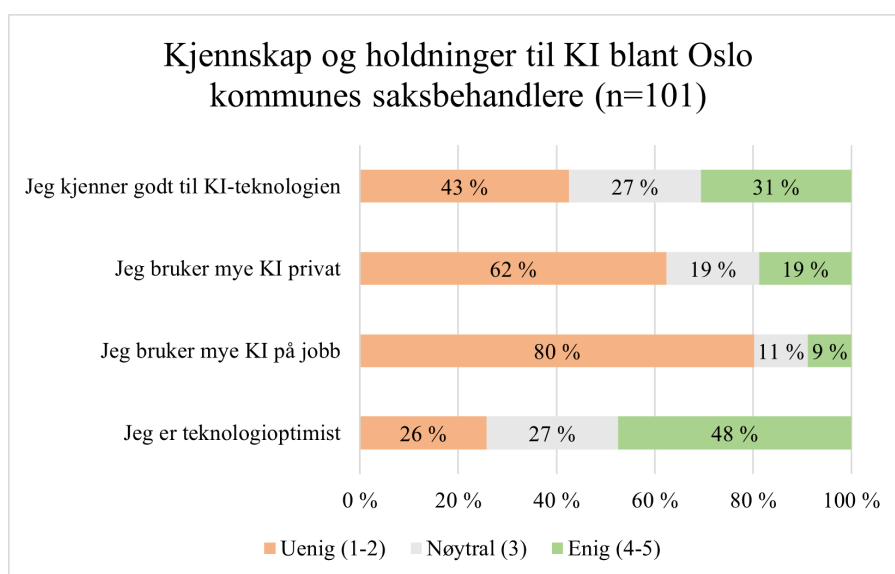
I denne delen av resultatkapittelet blir funnene fra spørreundersøkelsen til kommunens saksbehandlere, del 2 av datainnsamlingen, presentert. Det er funn fra data-materialet som er av særlig interesse og relevans for forskningsspørsmål 2, *Hvilke utfordringer finnes i dagens saksbehandling og hva kan LLM bidra med ifølge saksbehandlerne selv?*, som blir fremhevet. Hovedfunnene er fordelt på fire kategorier som vises under. Videre blir funnene presentert i den oppgitte rekkefølgen.

1. Bruk av og kjennskap til LLM
2. Utfordringer i eksisterende saksbehandling
3. Forventninger og bekymringer om LLM i saksbehandling
4. Implementerings- og gevinstvurderinger

I spørreundersøkelsen er begrepet kunstig intelligens brukt bevisst og konsekvent, da dette er et mer kjent begrep enn språkmodeller blant folk flest. Det fremkom samtidig eksplisitt og i uthevet skrift at det er store språkmodeller det refereres til når KI og kunstig intelligens ble benyttet i spørreundersøkelsen. Henvisningene til KI i dette delkapittelet skal dermed leses som LLM, slik LLM er definert i teorikapittelet.

4.2.1 Bruk av og kjennskap til LLM

Figur 8 viser saksbehandlerne bruk, kjennskap og holdning til KI (LLM) vurdert på en Likert-skala fra 1 til 5. Det er et mindretall på 31 prosent av saksbehandlerne som svarer at de kjenner godt til KI-teknologien. Resten er nøytrale eller uenige i påstanden, med den største andelen av disse som uenige. Det er ellers tydelig i data-materialet at saksbehandlerne i begrenset grad bruker KI. Dette gjelder både privat og på jobb, med henholdsvis 62 og 80 prosent som svarer at de er uenige i påstanden om at de bruker KI mye privat og på jobb. Imidlertid svarer nesten halvparten (48 prosent) av saksbehandlerne at de vurderer seg selv som teknologioptimister.



Figur 8: Respondentenes prosentvise bruk, kjennskap og holdninger til KI. Likert-skala 1-5 gruppert til «uenig», «nøytral» og «enig» per påstand.

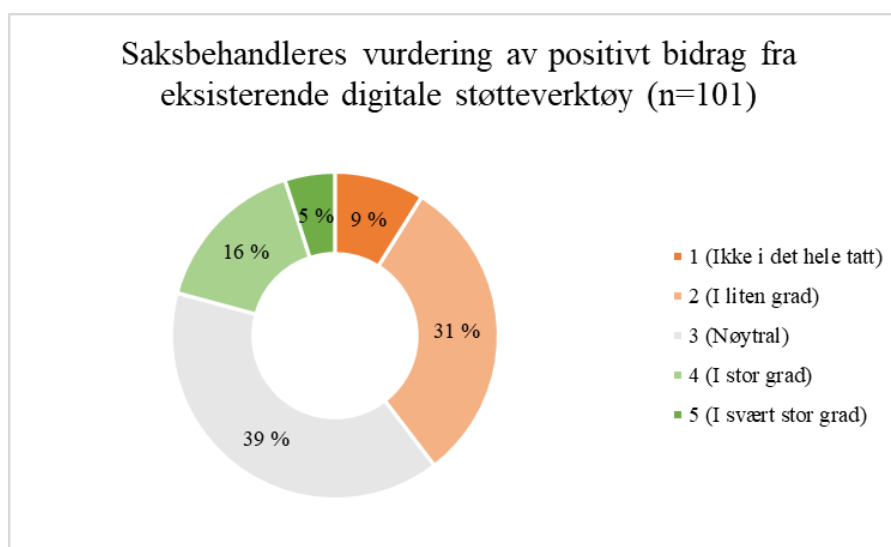
Funnene gjør seg gjeldende også i gjennomsnittlig score for kjennskap og bruk av KI som ligger under 3 (nøytralt) for de tre første påstandene (tabell 9). Klart lavest er gjennomsnittlig score for påstanden om bruk av KI på jobb. Gjennomsnittsscore for påstanden om teknologioptimisme er på sin side 3,36 med medianverdi 3. Det er relativt lik spredning i dataen, med det største standardavviket (SD) på påstanden om bruk av KI privat (1,27) og den minste spredningen på påstanden om bruk av KI på jobb (1,11).

Tabell 9: Respondentenes bruk, kjennskap og holdninger til KI. Gjennomsnitt, median og SD av Likert-skala 1-5 per påstand (n=101).

	Jeg kjenner godt til KI-teknologien	Jeg bruker mye KI privat	Jeg bruker mye KI på jobb	Jeg er teknologioptimist
Gjennomsnittsscore	2,88	2,26	1,76	3,36
Medianscore	3	2	1	3
Standardavvik	1,25	1,27	1,11	1,18

4.2.2 Utfordringer i eksisterende saksbehandling

Det er om lag like stor andel av saksbehandlerne som vurderer at de digitale støtteverktøyene de har tilgjengelig ikke i det hele tatt eller i liten grad bidrar til forbedret saksbehandling (40 prosent) som er nøytrale til samme påstand (39 prosent). Kun om lag en femtedel av saksbehandlerne mener at dagens systemer bidrar til forbedret saksbehandling i stor eller svært stor grad. Figur 9 viser saksbehandlernes vurdering av bidraget fra dagens digitale systemer.



Figur 9: Respondentens vurdering av positivt bidrag (forbedring) fra eksisterende digitale støtteverktøy/-systemer. Prosentfordeling av Likert-skala (1-5).

Saksbehandlerne oppgir i gjennomsnitt en Likert-score på 2,77, under det nøytrale 3, med medianverdi på 3 og standardavvik 0,99. Variasjonen på dette spørsmålet er således noe mindre enn i de foregående resultatene.

Sentralt i spørsmålet om hvorvidt LLM kan egne seg i saksbehandlingen er hva teknologien i så fall skal løse. Saksbehandlerne ble spurt om hvilke forhold ved dagens saksbehandling som utgjorde de største utfordringene. Saksbehandlerne kunne velge

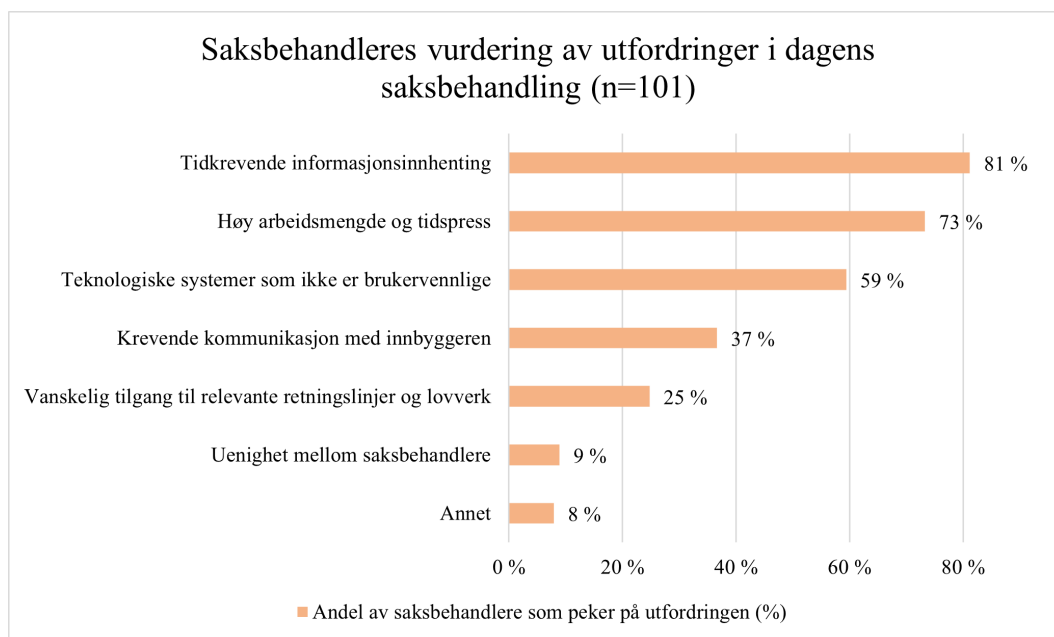
mellom seks på forhånd oppgitte forhold, i tillegg til å ha muligheten til å legge til egne. Det var anledning til å velge flere alternativer, og det ble avgitt totalt 295 svar. Tabell 10 viser antall avgitte svar og andel av saksbehandlerne som peker på hver utfordring, inkludert «Annet» som representerer saksbehandlerne egne svar utover de forhåndsoppgitte.

Blant de åtte fritekstsvarene som utgjør «Annet» problematiseres fagsystemet Gerica i stor grad (5/8 fritekstsvaer). Det pekes på at systemet er utdatert og ustabilt, og at enkelte deler oppleves som «tidstyver». For øvrig nevnes blant annet økende krav til saksbehandlingen, krevende bydelsøkonomi og uklare retningslinjer og lovverk som utfordringer i dagens saksbehandling.

Tabell 10: Respondentenes vurdering av utfordringer i dagens saksbehandling (n=101).

Utfordring	Antall avgitte svar	Andel av saksbehandlere som peker på utfordringen
Tidkrevende informasjonsinnhenting	82	81 %
Uenighet mellom saksbehandlere	9	9 %
Høy arbeidsmengde og tidspress	74	73 %
Vanskelig tilgang til relevante retningslinjer og lovverk	25	25 %
Teknologiske systemer som ikke er brukervennlige	60	59 %
Krevende kommunikasjon med innbyggeren	37	37 %
Annet	8	8 %
Sum avgitte svar	295	

Figur 10 viser resultatene i rekkefølge basert på andel saksbehandlere som peker på den aktuelle utfordringen. Utfordringer knyttet til tid er det saksbehandlerne i størst grad trekker frem. En stor majoritet av respondentene peker på tidkrevende informasjonsinnhenting (81 prosent) og høy arbeidsmengde og tidspress (73 prosent) som de to mest sentrale utfordringene. Også teknikk og informasjonsflyt problematiseres, der 59 prosent mener at teknologiske systemer som ikke er brukervennlige er en utfordring og 37 prosent mener at krevende kommunikasjon med innbyggeren er en utfordring. Ellers er det tydelig at få saksbehandlere peker på uenighet med kolleger som en utfordring i saksbehandlingen idag.



Figur 10: Respondentenes vurdering av utfordringer i dagens saksbehandling, målt etter andel av respondenter som peker på hver utfordring.

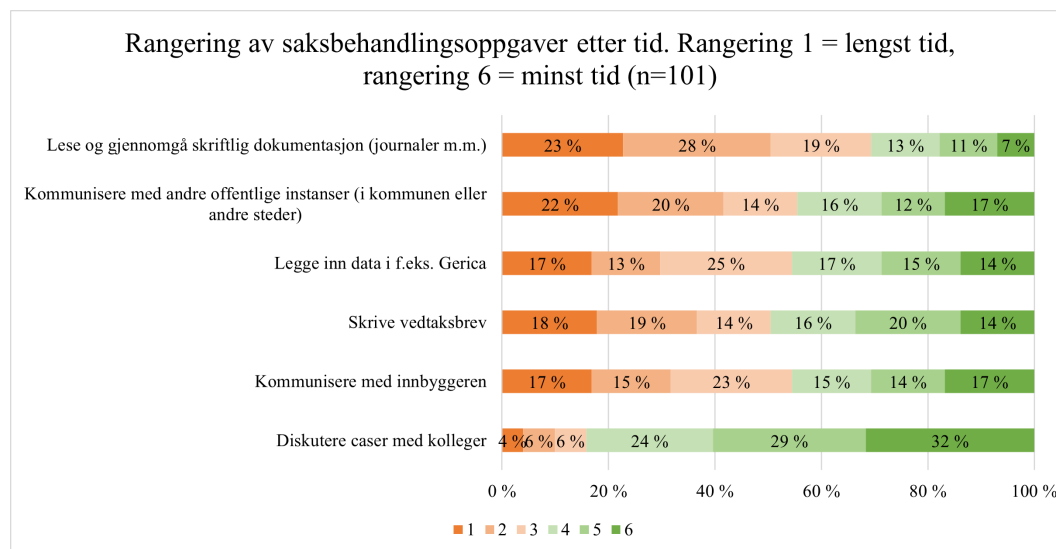
Også sentralt for å beskrive utfordringsbildet i dagens saksbehandling er å identifisere hvilke «tidstyver» som finnes i arbeidet. I denne sammenheng ble saksbehandlerne spurt om å rangere seks på forhånd utvalgte prosesssteg etter tidsbruk. Tabell 11 viser de seks oppgavene/prosessene med tilhørende rangeringer fra saksbehandlerne. Lave rangeringer er ekvivalent med høy tidsbruk, og vice versa. Det er tydelig at å lese og gjennomgå skriftlig dokumentasjon er vurdert som den mest tidkrevende oppgaven. Gjennomsnittsrangeringen for denne oppgaven er 2,82, som er betraktelig lavere enn de andre. Ser man på spredningen i datasettet er denne også relativt lav for denne oppgaven, med et standardavvik på 1,54 rangeringsnumre.

I motsatt ende av resultatene finner vi oppgaven «diskutere caser med kolleger», med en gjennomsnittsrangering på 4,62 og median på 5. Også her er spredningen relativt liten, med 1,36 rangeringsnumre i standardavvik.

Tabell 11: Respondentenes rangering av saksbehandlingsoppgaver etter hvor tidkrevende de er. Rangering 1 = lengst tid, rangering 6 = minst tid (n=101).

Oppgave	Antall rangeringer						Deskriptiv statistikk		
	1	2	3	4	5	6	Gj.snitts-rangering	Median-rangering	SD rangering
Lese og gjennomgå skriftlig dokumentasjon (journaler m.m.)	23	28	19	13	11	7	2,82	2	1,54
Kommunisere med andre offentlige instanser (i kommunen eller andre steder)	22	20	14	16	12	17	3,27	3	1,78
Legge inn data i f.eks. GERICA	17	13	25	17	15	14	3,42	3	1,63
Skrive vedtaksbrev	18	19	14	16	20	14	3,43	3	1,72
Kommunisere med innbyggeren	17	15	23	15	14	17	3,45	3	1,69
Diskutere caser med kolleger	4	6	6	24	29	32	4,62	5	1,36

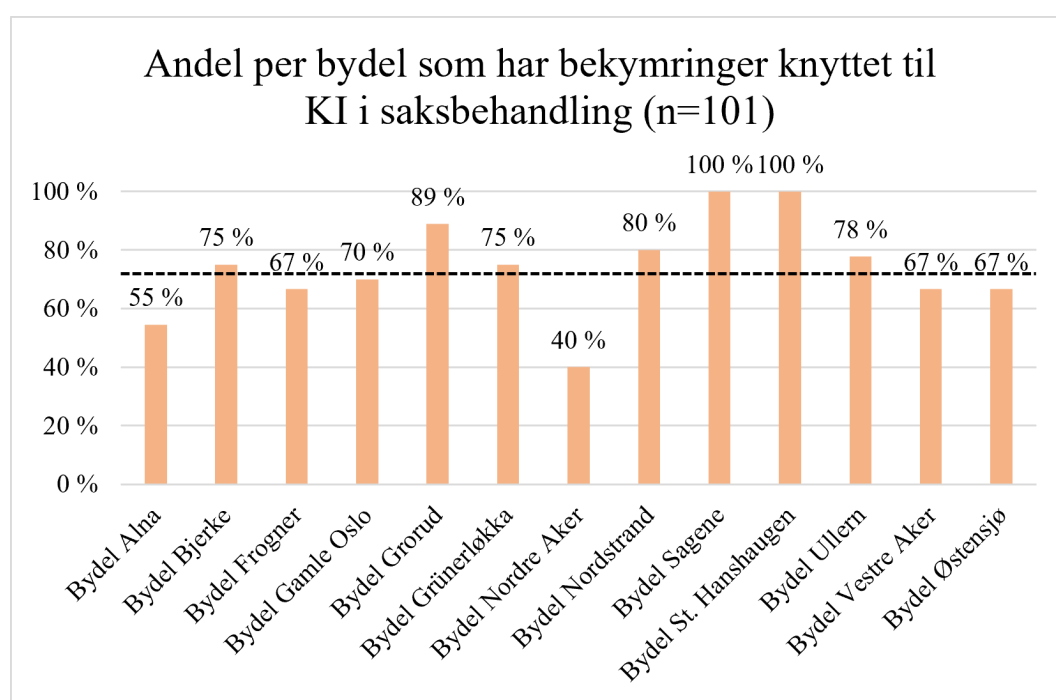
Figur 11 viser den prosentvise fordelingen av ulike rangeringer per oppgave. Den relativt like gjennomsnitts- og medianrangeringen kommer godt til syne på de utvalgte oppgavene, unntatt «lese og gjennomgå skriftlig dokumentasjon» og «diskutere caser med kolleger» som skiller seg ut i hver sin ende. Det er tydelig at saksbehandlerne er relativt samstemte om rangeringen av disse etter tidsbruk.



Figur 11: Respondentenes prosentvise rangering av saksbehandlingsoppgaver etter hvor tidkrevende de er. Rangering 1 = lengst tid, rangering 6 = minst tid.

4.2.3 Forventninger og bekymringer om LLM i saksbehandling

Det tredje sentrale fokuset i datainnsamlingen var hvilke positive forventninger og bekymringer saksbehandlerne hadde til bruk av språkmodeller i saksbehandlingen. Dette var viktig for å identifisere både barrierer mot og fortrinn for teknologien. I alt 73 respondenter svarte at de hadde bekymringer knyttet til bruk av LLM i saksbehandling. 39 oppgav at de ikke hadde bekymringer. Andelen av saksbehandlere som hadde bekymringer varierer stort per bydel. Figur 12 viser at andelen med bekymringer er klart lavest i bydel Nordre Aker (40 prosent) og høyest i bydelene Sagene og St. Hanshaugen der samtlige respondenter hadde bekymringer knyttet til LLM i saksbehandling.



Figur 12: Andel respondenter med bekymringer om KI i saksbehandling per bydel. Stiplet linje er gjennomsnittsandelen (72 prosent).

Respondentene ble spurt om hvilke, av et forhåndsgitt utvalg på fire, konkrete forhold de var bekymret for. Tabell 12 viser resultatene av dette spørsmålet. En stor majoritet har bekymringer om LLMs nøyaktighet og rettferdighet (82 prosent) og juridiske forhold/personvern knyttet til teknologien (70 prosent). Et mindretall på om lag en tredjedel av respondentene peker på hver av bekymringene «Egen eller bydelens teknologiforståelse» og «Egen jobbhverdag».

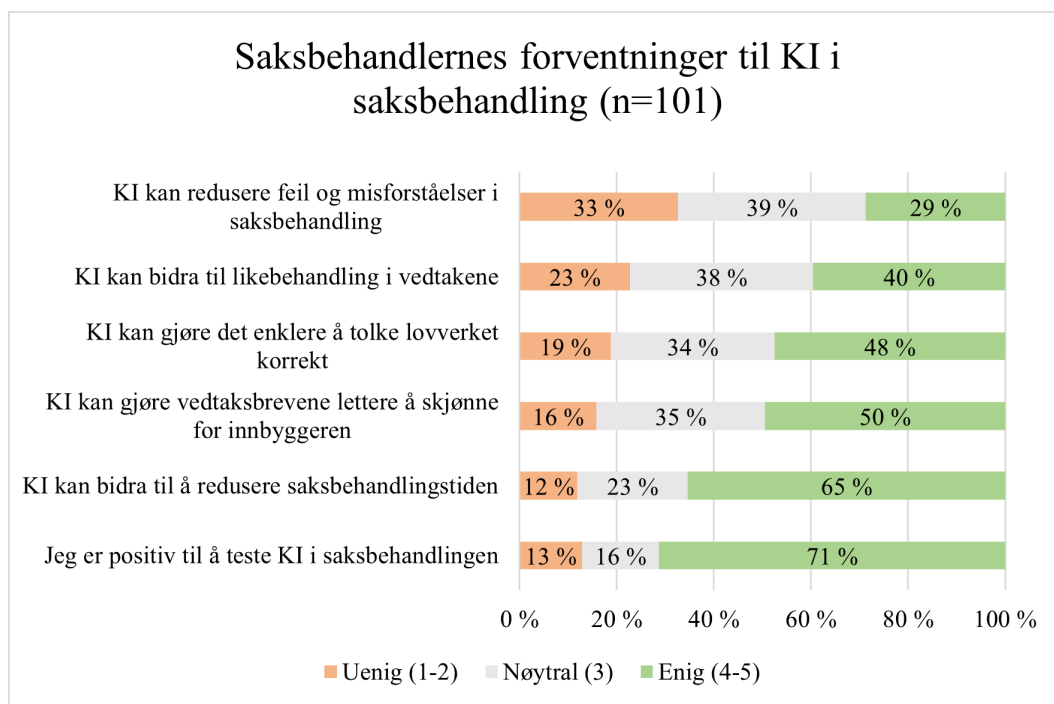
I tillegg fikk respondentene anledning til å skrive inn egne bekymringer i fritekstsvar. Åtte respondenter gjorde dette, og svarene dreier seg i stor grad om bekymringer

knyttet til LLMs evne til å håndtere og gjøre skjønnsvurderinger (6/8 fritekstsvaer). I tillegg pekes det på bekymringer om at LLM vil gå på bekostning av personlig kontakt/kommunikasjon og at det trengs en mer inngående beskrivelse av teknologien for at respondentene kan gi opplyste svar på spørreundersøkelsen.

Tabell 12: Respondentenes bekymringer om bruk av KI i saksbehandlingen, flervalg (n=73).

Bekymring	Antall avgitte svar	Andel av som har bekymringen
Bekymringer om teknologiens nøyaktighet/rettferdighet	60	82 %
Bekymringer om juridiske forhold/personvern	51	70 %
Bekymringer om egen eller bydelens teknologiforståelse	25	34 %
Bekymringer om egen jobbhverdag	23	32 %
Annet	8	11 %
Sum avgitte svar	167	

For å vurdere saksbehandlernes (positive) forventninger til LLM i saksbehandling ble de bredt om å rangere seks påstander etter enighetsgrad (Likert-skala) fra 1 (svært uenig) til 5 (svært enig). Figur 13 viser den prosentvise fordelingen av resultatene gruppert til enig, nøytral og uenig. Resultatene varierer en del per påstand, der det er størst uenighet blant saksbehandlerne om hvorvidt KI kan redusere feil og misforståelser i saksbehandling. Her er det om lag en tredeling av respondentene mellom uenig, nøytral og enig. I motsatt ende av skalaen er påstanden om at KI kan bidra til å redusere saksbehandlingstiden, der 65 prosent av saksbehandlerne er enige i påstanden. Likeledes er saksbehandlerne i overveldende grad positive (71 prosent) til å teste ut teknologien, og kun 13 prosent av de spurte ønsker ikke dette.



Figur 13: Respondentenes forventninger til bruk av KI i saksbehandlingen, målt etter andel av respondenter som peker på hvert forhold. Likert-skala 1-5 gruppert til «uenig», «nøytral» og «enig» per påstand.

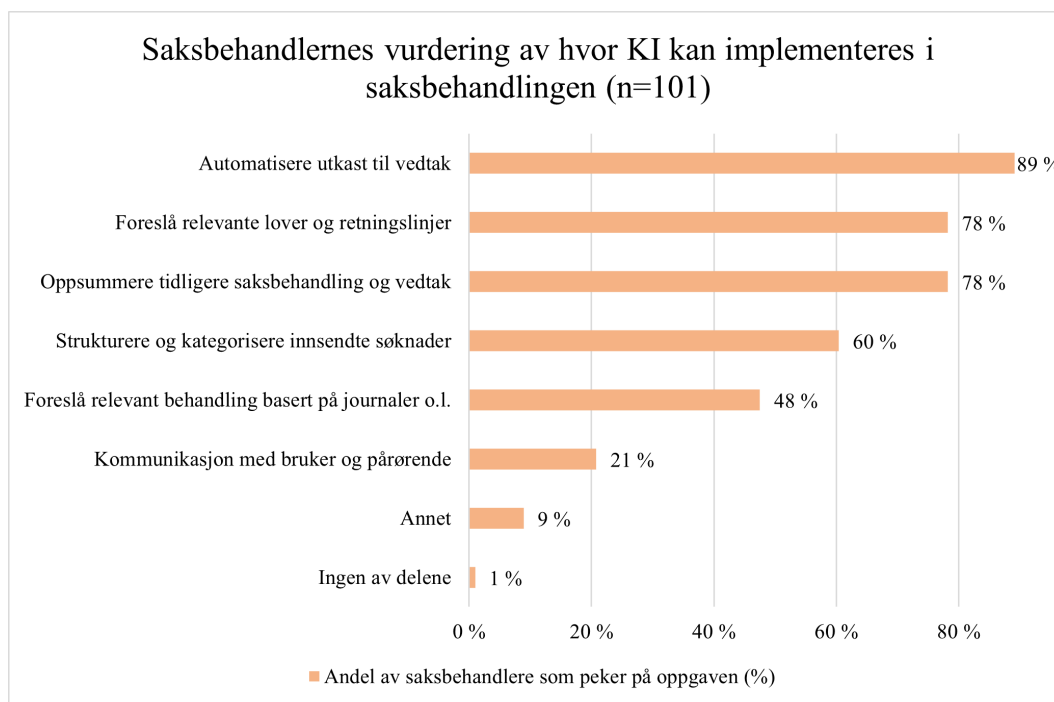
Tabell 13 viser deskriptiv statistikk på saksbehandlerne forventninger til LLM i saksbehandling. Her kommer det også tydelig frem at saksbehandlerne i snitt er nøytrale til påstanden om at KI kan redusere feil og misforståelser (3,01 i gjennomsnittlig Likert-score mot 3 som er nøytralt). Påstanden om at KI kan bidra til å redusere saksbehandlingstid har på sin side en snittverdi på 3,70 og median på 4, som underbygger at respondentene i større grad er enige i påstanden. Det er også på denne påstanden det er minst spredning i dataen, med et standardavvik på 0,99. Spørsmålet om hvorvidt saksbehandleren selv er positiv til å teste ut KI i saksbehandlingen har den høyeste gjennomsnittsscoren (3,96), men samtidig høyest variasjon (1,14), som kan tyder på at det i større grad er ulike meninger på dette spørsmålet.

Tabell 13: Respondentenes forventninger til bruk av KI i saksbehandling. Gjennomsnitt, median og SD av Likert-skala 1-5 per påstand (n=101).

Forventning	Gj.snittsscore	Medianscore	Standardavvik
KI kan redusere feil og misforståelser i saksbehandling	3,01	3	1,12
KI kan bidra til likebehandling i vedtakene	3,30	3	1,11
KI kan gjøre det enklere å tolke lovverket korrekt	3,39	3	1,06
KI kan gjøre vedtaksbrevne lettere å skjønne for innbyggeren	3,50	4	1,04
KI kan bidra til å redusere saksbehandlingstiden	3,70	4	0,99
Jeg er positiv til å teste KI i saksbehandlingen	3,96	4	1,14

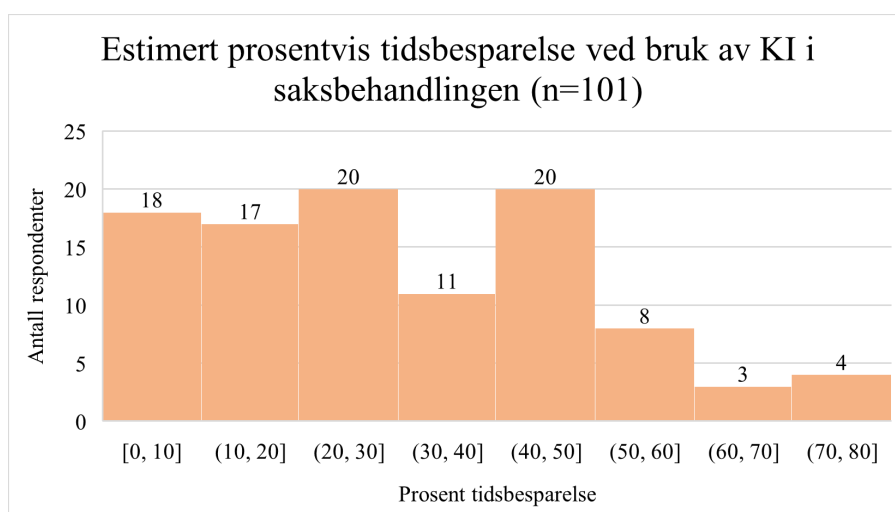
4.2.4 Implementerings- og gevinstvurderinger

For å undersøke mer konkret hva saksbehandlere i Oslo kommune mener at LLM kan bidra med, ble det samlet inn og analysert data på konkrete implementerings- og gevinstvurderinger. Saksbehandlerne ble blant annet spurt om hvor og til hva KI kan implementeres for i saksbehandlingen. Figur 14 viser resultatene av dette flervalgsspørsmålet som relativ andel av antall respondenter. Nærmere 9/10 saksbehandlere (89 prosent) peker på at LLM kan implementeres for å automatisere utkast til vedtak. 78 prosent vurderer på samme vis at teknologien kan implementeres til å oppsummere tidligere saksbehandling og vedtak og til å foreslå relevante lover og retningslinjer. Kun omtrent en femtedel av saksbehandlerne mener at KI kan implementeres i kommunikasjonen til bruker og pårørende. Én saksbehandler vurderer at teknologien er uegnet til samtlige bruksområder.



Figur 14: Respondentenes vurdering av hvor i saksbehandlingen KI kan implementeres, flervalg.

Ni saksbehandlere la inn egenformulerte svar under «Annet». Her pekes det blant annet på at LLM kan brukes til å oppsummere omfattende tekstdata og generere data som transkriberinger, referater, ledsagerbevis og behovbeskrivelser. I tillegg tas det opp at LLM kan brukes til å finne lignende saker for å sikre likebehandling og å forenkle vedtakstekster. Det pekes også på at et nytt fagsystem som er enklere enn Gerica hadde vært positivt, og at eksisterende dokumentasjon i journalsystemet ikke er tilstrekkelig som grunnlag for å behandle saker.



Figur 15: Histogram over estimert prosentvis tidsbesparelse ved bruk av KI til det respondentene mener teknologien kan brukes til i saksbehandlingen.

Saksbehandlerne ble også utfordret til å kvantifisere hvilken tidsbesparelse de mente at LLM ville gi dersom teknologien ble brukt slik de selv mente den kunne, i en gjennomsnittlig saksbehandlingsprosess. Resultatet av denne kvantifiseringen kan sees i figur 15 der saksbehandlerne estimat på prosentvis tidsbesparelse er presentert i et histogram med ti prosentpoeng i intervall. Det er tydelig at det er relativt stor spredning i svarene, og det kan ikke sies å være en tydelig normalfordeling, for eksempel. Men det er to toppestimer mellom 20 og 30 og mellom 40 og 50 prosent tidsbesparelse.

Dataen ble videre analysert deskriptivt for å finne at det gjennomsnittlige estimatet er 34,6 prosent tidsbesparelse og medianestimatet er 30 prosent tidsbesparelse. I tillegg ble gjennomsnittet i populasjonen beregnet med 95 % konfidensintervall. Konfidensintervallet (CI) for gjennomsnittet er [31.5 % , 37.6 %]. Tabell 14 presenterer de statistiske målene. Estimaten lå mellom 0 og 80 prosent, og spredningen er nokså stor; standardavviket er 20 prosentpoeng, mens interkvartilbredden (differansen mellom Q3 og Q1) er 30 prosentpoeng.

Tabell 14: Statistiske mål for estimert prosentvis tidsbesparelse og ekvivalent kostnadsbesparelse ved bruk av KI til det respondentene mener teknologien kan brukes til i saksbehandlingen (n=101).

Statistisk mål	Prosent tidsbesparelse estimert av saksbehandlere	Ekvivalent kostnadsbesparelse for utvalget (gitt arbeidskraftkost. på 903 902 kr/årsverk)	Ekvivalent kostnadsbesparelse for kommunen (gitt arbeidskraftkost. på 903 902 kr/årsverk og 260 årsverk)
Minimum	0,0 %	–	–
1. kvartil (Q1)	20,0 %	18 258 820 kr	47 002 904 kr
Median	30,0 %	27 388 231 kr	70 504 356 kr
Nedre CI-grense gj.snitt	31,5 %	28 747 233 kr	74 002 777 kr
Gjennomsnitt	34,5 %	31 537 141 kr	81 184 719 kr
Øvre CI-grense gj.snitt	37,6 %	34 327 049 kr	88 366 661 kr
3. kvartil (Q3)	50,0 %	45 647 051 kr	117 507 260 kr
Maksimum	80,0 %	73 035 282 kr	188 011 616 kr
Standardavvik (SD)	20,0 %	18 257 690 kr	46 999 995 kr
Interkvartilbredde (IQR)	30,0 %	27 388 231 kr	70 504 356 kr

Ved videre arbeid med estimatene for tidsbesparelse var det mulig å regne ut et grovt estimat på økonomisk konsekvens av bruken av LLM i den utstrekning som saksbehandlerne har estimert er mulig. Beregningen ble gjort med SSBs statistikk for totale arbeidskraftkostnader¹. For utvalget isolert (n=101) er den ekvivalente kostnadsbesparelsen ca. 32 millioner kroner ved gjennomsnittlig besparelse og ca. 27 millioner kroner ved medianbesparelse. For populasjonen, det vil si samtlige 260 saksbehandlere² i Oslo kommune, er den ekvivalente kostnadsbesparelsen mellom ca. 74 millioner kroner og 88 millioner kroner ved gjennomsnittlig besparelse og et konfidensintervall på 95 %. Imidlertid er det fortsatt stor spredning i estimatene, og standardavviket utgjør ca. 47 millioner kroner for kommunen samlet, og interkvartilbredden utgjør ca. 71 millioner kroner.

For å kartlegge LLMs potensial til helautomatisering ble saksbehandlerne spurt om det finnes typer av saksbehandling som kunne egnet seg til dette. Majoriteten (63 prosent) svarte at det ikke fantes saksbehandling der dette var mulig. 37 prosent av saksbehandlerne mente imidlertid at det fantes slike typer av saksbehandling, og disse fikk oppfølgingsspørsmålet om hvilke konkrete tjenestetyper de mente egnet seg til helautomatisering. Tabell 15 viser hvilke typer saksbehandling som 36 saksbehandlere mente egnet seg til helautomatisering. Det er tydelig at spesielt saksbehandling i forbindelse med TT-kort og trygghetsalarm er noe mange av respondentene (42

¹SSB anslo i 2023 en gjennomsnittlig total arbeidskraftkostnad per årsverk i helse- og sosialtjenester på 903 902 kroner (SSB, u.å.).

²Det er anslått rundt 260 saksbehandlere i Oslo kommunes 15 bydeler, se kap. 3.2.2 Kvantitativ del.

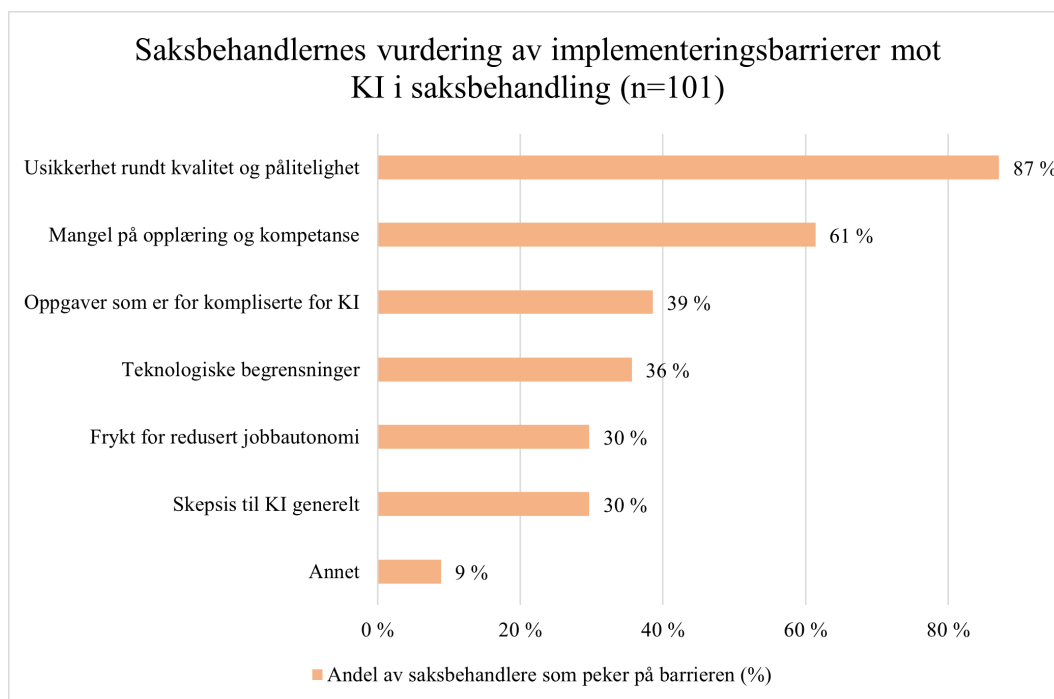
prosent) mener kan helautomatiseres.

Tabell 15: Saksbehandling som respondenter mener kan helautomatiseres (n=36).

Saksbehandlingstype som kan helautomatiseres	Antall	Andel
TT-kort	15	42 %
Trygghetsalarm	15	42 %
Ledsagerbevis	8	22 %
Praktisk bistand	6	17 %
Omsorgsbolig/Omsorg+	4	11 %
Støttekontakt	3	8 %
Koordinator/individuell plan	3	8 %
Vedtak om dagsenter	2	6 %
Annet (tjenestetyper nevnt én gang)	4	11 %
Ikke konkrete tjenestetyper	4	11 %

Utover rene bekymringer og forventninger for teknologien, er det relevant å vite noe om konkrete implementeringsbarrierer. Saksbehandlerne ble spurt om dette ved hjelp av et flervalgsspørsmål med seks forhåndsoppgitte barrierer og muligheten for å legge inn fritekstsvar («Annet»). Figur 16 viser prosentvis fordeling på hver barriere. Det ble totalt avgitt 294 svar.

Fritekstsvarene varierer stort. Blant annet trekkes personvern/juridiske forhold, skjønnsvurderinger, komplisert fagsystem, sikkerhet og den fagetiske profesjonstekningen fra sykepleierutdanningen frem som implementeringsbarrierer.



Figur 16: Respondentenes vurdering av implementeringsbarrierer mot KI i saksbehandling, flervalg.

Til sist i spørreundersøkelsen ble saksbehandlerne spurt om hvor viktig det var at saksbehandlere også involveres i utvikling og testing av KI. Her er utvalget nokså unisone om at dette er vesentlig. Gjennomsnittlig Likert-score på spørsmålet er 4,51 (mot 5 som er maksimalt). 89 prosent av saksbehandlerne var enige (score 4 og 5) i påstanden.

Øvrige generelle fritekstinnspill og refleksjoner fra saksbehandlerne om tematikken kan leses i vedlegg [A](#).

5 Diskusjon

I dette kapitlet blir de sentrale funnene og studiens implikasjoner drøftet opp mot masteroppgavens problemstilling; *Finnes det anvendelser av store språkmodeller (LLM) som kan forbedre saksbehandlingen av søknader om helse- og omsorgstjenester i Oslo kommune?* Masteroppgaven har operert med to innganger til denne problemstillingen, som konkretiserte seg i to forskningsspørsmål:

1. Hvilket potensial er det kommunens ledende fagpersoner identifiserer for LLM i saksbehandlingen?
2. Hvilke utfordringer finnes i dagens saksbehandling og hva kan LLM bidra med ifølge saksbehandlerne selv?

Disse angriper problemstillingen med en kvalitativ tilnærming og en kvantitativ tilnærming, og resultatene er presentert for hver av dem i foregående kapittel. I denne diskusjonen blir resultatene sammenlignet, knyttet til teori og vurdert samlet opp mot problemstillingen ved hjelp av triangulering. Triangulering av de to resultatsettene brukes for å styrke diskusjonen og belyse konsistens og divergens mellom del 1 og 2.

5.1 Innledende betraktninger

Innledningsvis bør det knyttes en refleksjon til at denne masteroppgaven tar for seg en teknologi i voldsom utvikling, slik det er redegjort for i kapittel 2.2.2 Teknologiens utvikling. Det medfører at eksisterende forskning i forrykkende tempo blir utdatert, og at selv den relativt nye forskningen det er referert til i denne oppgaven vil henge bak den teknologiske utviklingen. Begrensninger og forbehold som er identifisert i litteraturen kan dermed allerede være løst eller endret vesentlig, om man antar at teknologien stadig håndterer sine egne problematiske sider. Derfor vil det i den videre diskusjonen være noe varsomhet knyttet til hvor stor vekt det legges på tidligere forskning på feltet.

Tilsvarende problematikk kan også gjøre seg gjeldene for informantene (heretter eksperter) og respondentene (heretter saksbehandlere). Det vil si at egen forståelse og kunnskap om LLM kan være utdatert og at erfaringer man eksempelvis har gjort seg med å teste tidlige utgaver av kommersielle LLM-verktøy blir tungtveiende for synet på teknologien idag, selv om den altså utvikler seg hurtig. I samme refleksjon ligger det også en betraktning om at det i allmennheten i stor grad er de generelle, store grunnmodellene som ChatGPT og CoPilot man tenker på når man blir spurt om LLM. Dette på tross av at det antagelig er domenespesifikke, spesialiserte modeller

som er aktuelle til bruk i saksbehandlingen i helse- og omsorgssektoren og ikke ujusterte grunnmodeller.

Mer overordnet i denne studien kan det observeres en spenning mellom det store teoretiske potensialet som er identifisert i litteraturen og de betydelige praktiske barrierene mot implementering som er funnet gjeldende i datamaterialet. Det er i denne sammenhengen også en divergens mellom resultatene som baserer seg på ekspertintervjuene og de som baseres på spørreundersøkelsen til saksbehandlere. Ekspertene er i større grad avholdene og varsomme i en del spørsmål enn saksbehandlerne synes å være, selv om det fremkommer helt klare bekymringer om teknologien fra sistnevnte også. Dette medvirker til et spenn mellom resultatene på de to forskningsspørsmålene, fordi data fra to ulike perspektiver blir sammenlignet. I det videre vil sentrale funn bli drøftet, både de som har divergerende syn og der resultatene fra del 1 og del 2 av datainnsamlingen er konsistente.

Problemstilling og forskningsspørsmål har i denne studien dreiet seg om Oslo kommune som en enhet, selv om kommunen har en komplisert og fragmentert organisering, slik kapittel 2.7 Oslo kommunes organisering beskriver. Saksbehandlingen er fordelt geografisk mellom femten separate bydeler, og ansvar for den teknologiske biten er fordelt mellom tre kommunale etater - som samtlige var representert i ekspertintervjuene. Når resultatene er presentert og blir diskutert i dette kapittelet, er det dermed verdt å ha i mente at de organisatoriske forskjellene i resultater er til dels betydelig. Forskjellen illustreres i spørsmålet om hvorvidt saksbehandlerne har bekymringer for bruk av LLM i saksbehandlingen, der resultatene varierer fra 40 prosent av saksbehandlerne i én bydel, og 100 prosent i en annen. Fragmentert organisatorisk ansvar kan derfor være en faktor som påvirker de samlede resultatene og muligheten kommunen har til å lykkes med LLM.

Som en siste innledende betraktning er det relevant med en refleksjon om hvordan teorikapittelet blir anvendt i denne diskusjonen. Som nevnt er det utvist generell varshomhet knyttet til vektleggelsen av (potensielt utdatert) litteratur. Delkapitlene om overordnet potensial i helsesektoren og metaanalyser anvendes hyppig i diskusjonen relativt til andre deler av teorien. Disse delkapitlene tar for seg de generelle funnene om LLM i forskningen. Delkapitlene om begreps- og teknologiforklaring, status og utvikling samt politiske krav og målsetninger inneholder viktig kontekstualiserende litteratur. Delkapitlene om Oslo kommunes organisering og KIAS-rapporten er viktige for *forretningsforståelsen*, som er CRISP-DMs første hovedfase. Det er imidlertid verdt å merke seg at de tre empiriske studiene i liten grad trekkes inn i diskusjonen, da de representerer relativt smale konklusjonspunkter. De har samtidig en verdi som illustrasjon på mulige anvendelser og effekter av bruk av LLM i ulike kontekster.

5.2 Hovedfunn

Resultatene to deler har ikke presentert funn som er direkte sammenlignbare kategori for kategori, da de angriper problemstillingen på ulike måter. Derfor er det i diskusjonen av studiens hovedfunn etablert seks diskusjonspunkter som kobler resultater fra spørreundersøkelsen til saksbehandlere med temaene fra den tematiske analysen av ekspertintervjuer. Diskusjonen deles følgelig inn i disse diskusjonspunktene:

1. Variert kompetanse og bruk
2. Digitalisering med forbedringspotensial
3. Teksthåndteringspotensial i saksbehandlingen
4. Gevinstvurderinger
5. Skjønn, bias og transparens som barrierer
6. Er LLM riktig teknologi?

5.2.1 Variert kompetanse og bruk

Først og fremst er det i resultatene tydelig at det er stor forskjell på de nasjonalpolitiske målsetningene om bruk av KI i offentlig sektor, og den faktiske bruken blant saksbehandlere i Oslo kommune. Der digitaliseringsministeren hadde et mål om bruk av KI i 80 prosent av offentlig sektor i løpet av 2025 (NRK, 2024), er det 80 prosent av saksbehandlerne i denne studien som sier de er *uenige* i påstanden om at de bruker KI mye på jobb. Det er også begrenset kunnskap om KI-teknologien i Oslo kommune, skal vi tro både ekspertene og saksbehandlerne. Det er heller ikke identifisert konkret anvendelse av domenespesifikke språkmodeller i kommunen mer enn i ett tilfelle, som ved denne masteroppgavens utarbeidelse var i en lanseringsfase i kommunen.

Når det gjelder den store variasjonen i kunnskap om språkmodeller er dette et funn det er relevant å ha med seg tidlig i diskusjonen. Saksbehandlerne oppgir selv at de har en ganske begrenset kunnskap til teknologien, der kun 31 prosent av dem mener de kjenner teknologien godt. Dette er konsistent med ekspertenes vurdering av ansattes kunnskapsnivå, som de vurderer som høyst variert. Også på ledernivå i kommunen, ifølge enkelte av dem, er kunnskapsnivået variert. Dette kan medvirke til at LLM-teknologien i noen sammenhenger er beskrevet med i overkant positive ordelag, eller er «hypet», for å bruke en informants begrep.

Ledelsens varierende kunnskapsnivå representerer også en faktor som kan påvirke LLMs potensial i saksbehandlingen, da flere av ekspertene peker på at ansvaret for eventuell implementering må plasseres tydelig i ledelsen. Dette fordi flere av ekspertene understreker at digitalisering til syvende og sist er et ledelsesansvar. Konsekvensen av den begrensede kunnskapen om LLM-teknologi blant ledere og ansatte er imidlertid mer diskutabel, ifølge Denecke mfl. (2024). I Delphi-utvalget var det dissens på spørsmålet om «Lack of understanding of the underlying technology» kvalifiserte som en svakhet eller risiko ved LLM, jamfør tabell 3, selv om flertallet av ekspertutvalget riktignok mente dette.

5.2.2 Digitalisering med forbedringspotensial

Det synes derimot å være stor grad av konsistens i resultatene når det kommer til kommunens og fagsystemets evne til å gi ansatte digital støtte, dele og styre etter data på en hensiktsmessig måte. Her er det en lang vei å gå. Det fremkommer tydelig av resultatene at saksbehandlerne opplever begrenset grad av støtte fra de digitale systemene, og at mye tid går til å flytte data manuelt. Ekspertene er av samme oppfatning, og trekker frem at det er begrenset informasjonsdeling i kommunen og at fagsystemet Gerica i helse- og omsorgssektoren har betydelige svakheter som følge av systemets kompleksitet.

Det er stor grad av konsistens i resultatene når det gjelder synet på Gerica. Ekspertene vurderer det elektroniske pasientjournalsystemet som svært komplekst og av en størrelse som gjør at brukere av fagsystemet kun forholder seg til en liten del av systemet, avhengig av hvilken rolle de har. Det er interessant at selv med en nokså avgrenset bruk av systemet, svarer likevel 59 prosent av saksbehandlerne at de jobber med teknologiske system som ikke er brukervennlige. Dette kan tyde på at både størrelse og design er utfordringer med dagens fagsystem. Gerica vurderes på samme tid som utdatert i ekspertintervjuene, på en måte som gjør at teknisk kapasitet bindes opp til et gammelt system i kommunen. Det begrenser andre fremtidige satsninger.

I denne sammenhengen synes det å være et paradoks. Kommunen har altså over tid lagt inn så mange funksjoner i fagsystemet, for å bedre og effektivisere prosesser, at systemet nå er blitt så stort, utdatert og har så mange avhengigheter at det begrenser teknologiske innovasjoner som kan bidra til nettopp bedre og mer effektive prosesser. Denne tekniske gjelden, som en informant beskriver det som, kan medføre at kommunen bruker ressurser og kapasitet på et system som hemmer teknologisk innovasjon, som LLM potensielt kan gjøre utgjøre.

Ekspertene trekker samtidig frem at utfordringen med begrenset digital støtte og informasjonsflyt ikke er enestående for bydelene eller helse- og omsorgssektoren. Det

er et tydelig diskursfellestrekk blant ekspertene om at kommunen i liten grad er datastyrt. Det siste er i direkte strid med Oslo kommunes vedtatte strategi som slår fast at kommunen nettopp skal være datastyrt (Oslo kommune, 2023b). Denne divergensen mellom politikk og praksis bekrefter Spilker (2024) sin påstand om at det er en betydelig kontrast mellom politiske forestillinger om digitalisering og den praktiske implementeringen i virksomhetene.

Den tydelige mangelen på smidig informasjonsdeling og datastyring kan tolkes to-delt; på den ene side at potensialet for teknologisk innovasjon generelt, og LLM spesielt, blir redusert fordi kommunen og systemene i liten grad bruker dataen sin godt idag. At man rett og slett ikke er moden for teknologien. På den annen side kan man tolke nettopp denne mangelen på datastyring, støtte til saksbehandlere og informasjonsflyt som et stort potensial for forbedring ved for eksempel bruk av språk-modeller. I denne tolkningen er det også relevant å trekke inn Denecke mfl. (2024) fra teorien, som slo fast at det var konsensus blant ekspertene i Delphi-studien om at «More efficient data handling and extraction» var en generell effekt av å bruke LLM i helsesektoren (tabell 2). Legger man dette til grunn impliseres det at bruk av LLM i seg selv kan bidra til å bedre informasjonshåndtering og datastyring.

5.2.3 Teksthåndteringspotensial i saksbehandlingen

En viktig konsensus er at både saksbehandlere og eksperter vurderer at det å lese og gjennomgå skriftlig dokumentasjon, rett og slett bearbeiding og arbeid med store tekstmengder, er en hovedutfordring i dagens saksbehandlingsprosess. Saksbehandlerne vurderer sågar denne oppgaven som den mest tidkrevende i saksbehandlingen, når de rangerer ulike oppgaver. 78 prosent av saksbehandlerne vurderer også at det å oppsummere tidligere saksbehandling og vedtak er et område der LLM kan implementeres. Dette er konsistent med ekspertintervjuene og Denecke mfl. (2024) sine funn som viste at det var konsensus blant eksperter om at oppsummering var en av oppgavene som det var sannsynlig at LLM vil kunne bidra med i helsesektoren (tabell 1).

I diskusjonen om dokumentasjonsinnhenting og håndtering av tekst er det relevant at kommunen allerede har utviklet byggeklosser for tekstanalyse og smartsøk ved hjelp av LLM, slik det fremkommer i ekspertintervjuene. Infrastrukturen er plassert i kommunens felles IKT-plattform, og bruker teknologien i en sluttbrukertjeneste innenfor plan- og byggesaker i skrivende stund. Praktisk erfaring fra implementering av denne typen teknologi er mangelvare, som det er beskrevet i metaanalysen til Bedi mfl. (2024), og lærdommen derfra bør sees i sammenheng med behovet for enklere og raskere tekstgjennomgang i helse- og omsorgssaker også. Behovet og potensialet for tidsbesparelse er gjennomgående beskrevet som stort i datamaterialet til denne masteroppgaven.

Når det gjelder tekstproduksjon, eksempelvis til vedtaksbrev, er det større grad av divergens mellom resultatene fra del 1 og del 2 av datainnsamlingen. Mens saksbehandlerne i svært stor grad (89 prosent) mener at LLM kan implementeres for å automatisere utkast til vedtak, er det kommet begrunnede tilbakemeldinger fra ekspertintervjuene på at det per nå er for mye usikkerhet knyttet til teknologien til at man helt og holdent kan stole på et utkast til vedtak generert av LLM. Dessuten trekkes det frem i ekspertintervjuene at det kan være at klassiske maler og tydeligere retningslinjer kan ha vel så stor effekt. Dette er også relevant med tanke på å ikke «skyte spurv med kanon», slik en informant beskrev utfordringene knyttet til bærekraft med LLM, når enklere og mindre energikrevende løsninger kan bidra til ønsket forbedring.

En kontrast til ekspertenes skepsis til tekstproduksjon i denne studien er funnene til Meng mfl. (2024). I metaanalysen ble medisinsk skriving identifisert som én av fire hovedområder der LLMer kan «transformere» helsetjenesten. Forfatterne begrunnet dette med at studiene viser at språkmodellene kan utarbeide utkast til medisinske dokumenter. Divergensen mellom ekspertintervjuene og funnene i den nevnte metaanalysen kan tyde på at tekstproduksjon ved bruk av LLM ikke nødvendigvis er kategorisk eller for all fremtid uegnet i saksbehandlingen, men at det kan være andre bruksområder som innebærer mindre risiko og er mer nærliggende på kortere sikt.

5.2.4 Gevinstvurderinger

I vurderingen av potensiell gevinst ved bruk av LLM i saksbehandling er det interessant å merke seg at saksbehandlerne anslår en tidsbesparelse (gjennomsnitt 34,5 prosent, median 30 prosent) som er nokså lik som det KIAS-rapporten skisserte i 2024 (30 prosent). Undeland mfl. (2024) anslo også at tidsbesparelsen på 30 prosent representerte en årlig økonomisk gevinst på minst 55 millioner kroner for Oslo kommunes helse- og omsorgstjenester samlet. Dette er betydelig lavere enn den beregnede ekvivalente kostnadsbesparelsen for kommunen som denne masteroppgaven presenterer (mellom 74 millioner og 88 millioner kroner ved bruk av gjennomsnittsestimat for besparelse og et 95 % konfidensintervall).

Forskjellen i skissert økonomisk gevinst understreker at beregningen av ekvivalent kostnadsbesparelse er forbundet med stor usikkerhet, og at det er en teoretisk beregning heller enn et reelt estimat. Dessuten vil det naturligvis være kostnader forbundet med LLM-implementering som ikke er beregnet i denne studien. Av de generelle effektene av å benytte LLM i helsesektoren, er «Reduced health care costs» den lavest rangerte i utdraget fra Denecke mfl. (2024) i tabell 2 i teorien. Det er fortsatt konsensus blant ekspertene i Delphi-utvalget om at LLM vil påvirke kostnadsbildet positivt, men andelen som mener dette er lavere enn i de andre effektkategoriene i

utdraget. Dette underbygger at økonomiske gevinstvurderinger er heftet med stor grad av usikkerhet, og at behovet for mer helsepersonell antagelig gjør at de samlede kostnadene for helsetjenester uansett neppe blir redusert i årene som kommer.

5.2.5 Skjønn, bias og transparens som barrierer

Problematikken rundt graden av skjønnsvurderinger og risiko for manglende likebehandling er gjennomgående i resultatene. Som det fremkom i ekspertintervjuene er det eksempler på at skjønnsvurderingene idag kan gi store utslag i hvilket vedtaksnivå en innbygger får fra saksbehandler til saksbehandler og fra bydel til bydel. Undeland mfl. (2024) identifiserte økt likebehandling og likere vedtaksnivå som én av tre hovedgevinster ved bruk av LLM i saksbehandlingen. Samtidig trekkes det i ekspertintervjuene frem at det også finnes «god» forskjellsbehandling og situasjoner der samme diagnose og samme funksjonsnivå likevel bør gi forskjellig vedtak. Balansen mellom positiv skjønnsutøvelse og for store, negative forskjeller internt i kommunen er et dilemma som må adresseres videre og som påvirker hvor egnet LLM, algoritmer og annen teknologi er i saksbehandlingen.

Det er gjennomgående i resultatene at ekspertene har bekymringer knyttet til LLMs transparens og risiko for bias, og dette er spesifisert som ett av de syv temaene i den tematiske analysen. Selv om saksbehandlerne ikke ble bedt om å vurdere transparens og bias eksplisitt, fanges problematikken likevel opp i at hele 82 prosent av saksbehandlerne er bekymret for «teknologiens nøyaktighet/rettferdighet». I litteraturen er det flere studier som er konsistente med saksbehandlerne og ekspertenes syn på LLMs svakheter knyttet til disse to forholdene. Sallam (2023) sin metaanalyse, som riktignok fokuserte på ChatGPT spesifikt, trekker frem at manglende transparens og bias er problematisert i en rekke studier på tematikken. Denecke mfl. (2024) identifiserte konsensus blant ekspertene i Delphi-utvalget knyttet til risikoen for bias, men ikke knyttet til transparens (formulert som «Lack of explainability of system decision-making processes»), jamfør tabell 3. Selv om det var dissens i Delphi-utvalget på dette spørsmålet, var likevel majoriteten i utvalget enige i at manglende transparens var en svakhet ved LLM-baserte systemer i helsesektoren.

Bekymringen for bias og manglende transparens har samtidig et potensial for å bedres i nær fremtid. Slik det er beskrevet i ekspertintervjuene jobbes det med å utvikle helnorske grunnmodeller av blant annet Nasjonalbiblioteket. Når de norskutviklede modellene er ferdigstilt vil kommunen kunne ha full transparens til hvilken treningsdata som modellen er trent opp på. Dette vil derfor kunne redusere utfordringene knyttet til manglende transparens og tilhørende bekymring for bias i treningsdata betydelig. I denne forbindelse synes det også fornuftig at «bygggeklossene», det vil si infrastrukturen, kommunen har etablert i felles IKT-plattform er helt LLM-nøytral, slik at man enkelt kan skifte til en ny, norsk modell når denne er klar. Det kan antas

at resultatene i denne studien hadde stilt seg ganske annerledes om den norskutviklede og heltransparente modellen allerede var lansert og utprøvd.

5.2.6 Er LLM riktig teknologi?

Selv om denne studien tar for seg språkmodeller spesielt, er det relevant å diskutere at flere eksperter vurderer klassiske algoritmer som en mer lavhengende frukt på kort sikt i kommunen. Automatisering av kriteriestyrte oppgaver og bruk av strukturerte data i saksbehandlingen kan gi effektivitetsgevinster uten å ha store bekymringer om transparens, bias og bærekraft knyttet til seg. Samtidig trekkes det frem av ekspertene med helsefaglig ansvarsområde at en del skjønnsvurdering faktisk er nødvendig for å gi likeverdige tilbud til innbyggere, i tillegg til at majoriteten av saksbehandlerne (63 prosent) mener det uansett ikke finnes saksbehandlingstyper som kan helautomatiseres. Her ligger det en påfallende divergens mellom saksbehandlerne og et flertall av offentlige virksomheter i rapporten til Rambøll mfl., 2024 som på sin side mener at KI kan brukes til automatisert saksbehandling (jamfør figur 4). Riktignok er andelen offentlige virksomheter som mener dette redusert fra 2021 til 2024 (Rambøll mfl., 2024).

På grunn av manglende konsensus mellom eksperter, saksbehandlere og litteratur når det kommer til helautomatisering med LLM og/eller algoritmer kan det fremstå fornuftig å fokusere på algoritmer og LLM til beslutningsstøtte, heller en helautomatisering, i første omgang. Dette vil i så fall være i tråd med Baxter et al. (2024) som konkluderte med at LLM egnet seg bedre som beslutningsstøtte enn automatiseringsverktøy når det gjaldt pasientkommunikasjon som eksempel. På den annen side vil bruk av LLM og algoritmer kun til beslutningsstøtte sannsynligvis svekke gevinstpotensialet fordi menneskelig arbeidskraft blir nødvendig. Og som nevnt i teorien viste Vaccaro mfl. (2024) i sin metaanalyse av mer enn hundre studier at beslutninger basert på både menneskelig og kunstig intelligens har en tendens til å underprestere sammenlignet med de to hver for seg. Spørsmålet om beslutningsstøtte kontra automatisering kan således regnes som inkonklusivt i denne studien.

5.3 Implementering

Reddy (2024) konkluderte blant annet med at det på generelt grunnlag foreløpig ikke eksisterer en klar vei for implementering og integrering av generativ KI i dagens systemer og strukturer i helsevesenet. Dette er gjenkjennelig i datamaterialet i denne studien. Samtidig er det flere resultater som kan bidra til å styrke implementeringspotensialet for LLM i kommunal saksbehandling. I dette delkapittelet vil studiens implementeringsimplikasjoner diskuteres mer inngående enn i den generelle

diskusjonen. Som nevnt i introduksjonen er det denne masteroppgavens intensjon at funn fra prosjektet skal kunne utgjøre et bidrag til et utvidet kunnskapsgrunnlag for norske kommuner.

Forskning på LLM, som denne studien, er ferskvare og blir raskt utdatert fordi teknologien forbedres så hurtig. Der det er mindre brå endringer er de organisatoriske rammene, forutsetningene og prosessene som påvirker implementeringspotensialet til LLM. De identifiserte utfordringene og fortrinnene Oslo kommune har når det kommer til digital modenhet, organisatorisk struktur, ansattes kunnskapsnivå, teknisk infrastruktur og så videre bør gi relevant kunnskap i et implementeringsøyemed for både Oslo og andre kommuner. Det er som nevnt få studier som har hatt implementeringsvurderinger som evalueringskriterium i analysene sine (Bedi mfl., 2024). Denne masteroppgaven et bidrag til å bedre denne statistikken.

Et fellestrekk i resultatene er at det er behov for mer kunnskap, utprøving i trygge omgivelser og involvering av ansatte. Saksbehandlerne selv er svært opptatt av å bli involvert i utviklingen; 71 prosent er positive til å teste LLM i saksbehandlingen, og hele 89 prosent vurderer det som viktig at saksbehandlerne involveres i testing og utvikling av LLM. Det synes å være en kontrast mellom de ansattes ønske om involvering, og den faktiske involveringen, slik den er beskrevet i ekspertintervjuene. Samtidig pekes det i enkelte intervjuer på at kommunen i større grad involverer ansatte i KI-arbeidet nå enn med tidligere digitaliseringsinitiativ. I det videre arbeidet med språkmodeller og KI synes det viktig å videreføre denne involveringen, både for å øke den generelle forståelsen for teknologien blant Oslo kommunes ansatte, og ikke minst å klare å identifisere kloke bruksområder ved hjelp av de ansattes involvering og kunnskap.

Kommunens utvikling av infrastruktur for smartsøk og tekstanalyse i den felles IKT-plattformen bør som nevnt følges tett og utredes videre, med henblikk på mulig utnyttelse også i helse- og omsorgssektoren, der behovet for mer effektiv informasjonsinnhenting er stort. Erfaringer og data fra denne typen domenespesifikke språkmodeller i reell bruk i kommunal sektor er det få, om noen, eksempler på. Videre forskning bør derfor undersøke effektene av bruken av den nye infrastrukturen og overføringsverdien den måtte ha til helse- og omsorgsfeltet.

Teoretiske analyser og mer forskning tar oss likevel bare et visst steg lenger når man diskuterer praktisk implementering av LLM. Undeland mfl. (2024) konkluderte blant annet med at hurtig utprøving av kommunal KIAS (kunstig intelligens-assistert saksbehandling) ved bruk av MVP (Minimum Viable Product) bør være en del av veien videre. Også Reddy (2024) argumenter for behovet for utstrakt utprøving i reelle omgivelser og kontinuerlig iterasjon, og mener dette må drive utviklingen. Basert på saksbehandlerne oppgitte villighet til å teste LLM i saksbehandlingen og de lavthengende implementeringsmulighetene som ekspertene har vurdert som noen-

lunde adekvate, synes det naturlig at man i arbeidet med denne problemstillingen videre faktisk tar i bruk og tester ut enkle applikasjoner med LLM i saksbehandlingen. Ikke for å ta beslutninger eller helautomatisere, men for å hente erfaringer fra LLM i relativt ufarlige sammenhenger. I ekspertintervjuene er det trukket frem blant annet følgende anvendelser av LLM som er vurdert til å ha relativt lav risiko knyttet til implementering:

- Oversettelse av vedtaksbrev til andre språk enn norsk
- Forenkling og klarspråkliggjøring av tekst
- Smartsøk i retningslinjer, lovverk og dokumentasjon
- Oppsummering og analyse av eksisterende dokumenter

En fornuftig begynnelse på det videre FoU- og implementeringsarbeidet basert på funnene i denne masteroppgaven synes dermed å være å teste ut domenespesifikke språkmodeller i trygge og kontrollerte omgivelser, sammen med saksbehandlere, til anvendelsene som er nevnt ovenfor. Sammen med nøye oppfølging av effektene av de nye byggeklossene for smartsøk og tekstanalyse i felles IKT-plattform og videre forskning kan dette medvirke til at Oslo kommune får hentet ut gevinster av bruk av store språkmodeller i saksbehandling i helse- og omsorgstjenesten.

5.4 Konklusjon

Samlet sett viser funnene at det eksisterer anvendelser av store språkmodeller som kan forbedre saksbehandlingen i helse- og omsorgstjenesten i Oslo kommune. Basert på studiens dybdeintervjuer med ledende fagfolk, spørreskjemadata fra en høy andel av kommunens saksbehandlere og relevant litteratur er det identifisert et mulighetsrom for bruk av språkmodeller i saksbehandlingen. Det er identifisert et klart behov for bedre verktøy for teksthåndtering i saksbehandlingen, primært knyttet til informasjonsinnhenting og bistand til tekstproduksjon. Potensialet for tidsbesparelse ved bruk av LLM vurderes av saksbehandlerne til å ligge på mellom 31,5 % og 37,6 %, gitt et 95 % konfidensintervall på gjennomsnitt.

Bekymringer knyttet til graden av skønnsvurderinger, transparens og bias er samtidig høyst tilstedeværende. I denne sammenheng er det forhåpninger om at utviklingen av norske grunnmodeller vil bedre situasjonen. Kommunens evne til å dele og styre etter data vurderes samtidig som relativt dårlig, som ikke er i tråd med politiske ambisjoner. Fagsystemet for helse- og omsorgssektoren er også så komplekst og utdatert at det binder opp teknisk kapasitet i kommunen som kunne blitt brukt til andre ting. Kunnskapsnivået og erfaringen med LLM blant saksbehandlere er på

sin side dokumentert som svært variert, som både påvirker teknologiens potensial og dataen som denne masteroppgaven bygger sine funn på.

På kort sikt er det derfor vurdert at en fornuftig begynnelse på videre FoU- og implementeringsarbeid vil være å teste ut domenespesifikke språkmodeller i trygge og kontrollerte omgivelser, sammen med saksbehandlere. Anvendelser til oversettelse, klarspråkliggjøring, smartsøk og oppsummering av tekst er identifisert som anvendelser med relativt lav risiko knyttet til seg. Sammen med nøye oppfølging av effektene av de nye byggeklossene for smartsøk og tekstanalyse i felles IKT-plattform og videre forskning kan dette medvirke til at Oslo kommune får hentet ut gevinster av bruk av store språkmodeller i saksbehandling i helse- og omsorgstjenesten.

Referanser

- Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology* (tekn. rapp. Nr. 31422) (Lest 1. april 2025). National Bureau of Economic Research. <https://doi.org/10.3386/w31422>
- Angelova, V., Dobbie, W. S., & Yang, C. S. (2023). *Algorithmic Recommendations and Human Discretion* (tekn. rapp. Nr. 31747) (Lest 1. april 2025). National Bureau of Economic Research. <https://doi.org/10.3386/w31747>
- Banerjee, D., Singh, P., Avadhanam, A., & Srivastava, S. (2023). Benchmarking LLM powered Chatbots: Methods and Metrics [Lest 30. mars 2025]. *arXiv preprint arXiv:2308.04624*. <https://arxiv.org/pdf/2308.04624>
- Baxter, S. L., Longhurst, C. A., Millen, M., Sitapati, A. M., & Tai-Seale, M. (2024). Generative artificial intelligence responses to patient messages in the electronic health record: early lessons learned [Lest 25. mars 2025]. *JAMIA Open*, 7(2), ooae028. <https://doi.org/10.1093/jamiaopen/ooae028>
- Bedi, S., Liu, Y., Orr-Ewing, L., Dash, D., Koyejo, S., Callahan, A., & Nigam, H. S. (2024). A Systematic Review of Testing and Evaluation of Healthcare Applications of Large Language Models (LLMs). *medRxiv*. <https://www.medrxiv.org/content/10.1101/2024.04.15.24305869v4>
- Bommasani, R., & Liang, P. (2021 oktober). *Reflections on Foundation Models* [Lest 3. mars 2025]. <https://hai.stanford.edu/news/reflections-foundation-models>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at Work [Lest 1. april 2025]. *The Quarterly Journal of Economics*. <https://doi.org/10.1093/qje/qjae044>
- Chapman, P., Kerber, R., Clinton, J., Khabaza, T., Reinartz, T., & Wirth, R. (1999 mars). *CRISP-DM Discussion Paper March 1999* (tekn. rapp.). <https://mineracaodados.wordpress.com/wp-content/uploads/2012/12/crisp-dm-no-brand.pdf>
- Datatilsynet. (2023 Desember). *Tid for språkmodeller i sandkassa* [Lest 2. mars 2025]. <https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2023/tid-for-sprakmodeller-i-sandkassa/>
- Den nasjonale forskningsetiske komité for medisin og helsefag (NEM). (2009). *Veiledning for forskningsetisk og vitenskapelig vurdering av kvalitative forskningsprosjekt innen medisin og helsefag* (tekn. rapp.). <https://www.forskningsetikk.no/retningslinjer/med-helse/vurdering-av-kvalitative-forskningsprosjekt-innen-medisin-og-helsefag>
- Denecke, K., May, R., LLMHealthGroup & Romero, O. R. (2024). Potential of Large Language Models in Health Care: Delphi Study [Lest 8. mars 2025]. *Journal of Medical Internet Research*, 26, e52399. <https://doi.org/10.2196/52399>
- Digdir. (2024). *Veiledning for ansvarlig bruk og utvikling av kunstig intelligens* (tekn. rapp.) (Lest 2. mars 2025). Digitaliseringsdirektoratet. <https://www.digdir.no>

- [no/kunstig-intelligens/veiledning-ansvarlig-bruk-og-utvikling-av-kunstig-intelligens/4601](#)
- Elster, A. C., & Tidemann, A. (2025). Maskinlæring [Lest 28. mars 2025]. <https://snl.no/maskinl%C3%A6ring>
- Executive Office of the President. (2023 november). *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (tekn. rapp.) (Lest 3. mars 2025). Federal Register. <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>
- Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., & Chen, J. H. (2024). Large Language Model Influence on Diagnostic Reasoning. *JAMA Network Open*, 1–12. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2825395>
- Google. (2024). NotebookLM now lets you listen to a conversation about your sources [Lest 11. mai 2025]. <https://blog.google/technology/ai/notebooklm-audio-overviews/>
- Helsedirektoratet. (2024 mai). *Store språkmodellar i helse- og omsorgstenesta – eit kunnskapsgrunnlag* (tekn. rapp.). Helsedirektoratet. <https://www.helsedirektoratet.no/rapporter/store-sprakmodellar-i-helse-og-omsorgstenesta>
- Hem, E. (2025). IMRAD [Lest 11. mai 2025]. <https://sml.snl.no/IMRAD>
- Ho, A., Besiroglu, T., Erdil, E., Owen, D., Rahman, R., Guo, Z. C., Atkinson, D., Thompson, N., & Sevilla, J. (2024). Algorithmic Progress in Language Models [Lest 30. mars 2025]. *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS)*. https://proceedings.neurips.cc/paper_files/paper/2024/file/6b066da6a23bc55f9b887e7298102884-Paper-Conference.pdf
- Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in Hiring [Lest 1. april 2025]. *The Quarterly Journal of Economics*, 133(2), 765–800. <https://doi.org/10.1093/qje/qjx041>
- Hovland, K. M. (2024 mars). *Tek-nedgang: – Tror du kan trekke dette tilbake til DeepSeek* [Lest 3. mars 2025]. <https://e24.no/boers-og-finans/i/Vz25P6/tek-nedgang-tror-du-kan-trekke-dette-tilbake-til-deepseek>
- Jakobsen, K. M. (2025 januar). *Holdt tre allmøter om sammenslåing: Oslo kommunes store omveltning er i gang* [Lest 1. april 2025]. <https://www.ao.no/holdt-tre-allmoter-om-sammenslaing-oslo-kommunes-store-omveltning-er-i-gang/s/5-128-960610>
- Khan, S. D., Ullah, R., Rahim, M. A., Rashid, M., Ali, Z., Ullah, M., & Ullah, H. (2022). An efficient deep learning framework for face mask detection in complex scenes. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 159–169.

- KS. (2024). Nye tall fra KS: Svært krevende økonomisk situasjon også i 2025 [Lest 11. mai 2025]. <https://www.ks.no/fagomrader/okonomi/kno/nye-tall-svart-krevende-okonomisk-situasjon-i-2025/>
- Kumar, H., Rothschild, D. M., Goldstein, D. G., & Hofman, J. M. (2023). Math Education with Large Language Models: Peril or Promise? [Lest 1. april 2025]. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4641653>
- Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhang, X., Zhao, T., Panalkar, A., Mehta, D., Pasquali, S., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., ... Zhao, L. (2024). Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey [Lest 13. mai 2025]. *Association for Computing Machinery*. <https://arxiv.org/pdf/2305.18703v7>
- Machot, F. A., Horsch, M. T., & Ullah, H. (2024). Building Trustworthy AI: Transparent AI Systems via Large Language Models, Ontologies, and Logical Reasoning (TranspNet). *arXiv preprint arXiv:2411.08469*.
- Malt, U., & Grønmo, S. (2025). Likert-skala [Lest 4. mai 2025]. <https://snl.no/Likert-skala>
- Meng, X., Yan, X., Zhang, K., Liu, D., Cui, X., Yang, Y., Zhang, M., Cao, C., Wang, J., Wang, X., Gao, J., Wang, Y. S., & Ji, J. (2024). The application of large language models in medicine: A scoping review [Lest 3. mars 2025]. *iScience*, 27(5), 109713. <https://doi.org/10.1016/j.isci.2024.109713>
- Microsoft. (u.å.). *What Are Large Language Models (LLMs)?* [Lest 28. mars 2025]. <https://azure.microsoft.com/nb-no/resources/cloud-computing-dictionary/what-are-large-language-models-llms>
- Moreno, E., Alvarez-Lozada, L. A., Arrambide-Garza, F. J., Quiroga-Garza, A., & Elizondo-Omaña, R. E. (2023). Comment on Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 2023, 11, 887 [Lest 7. mars 2025]. *Healthcare*, 11(21), 2819. <https://doi.org/10.3390/healthcare11212819>
- NMBU. (2025 januar). *Retningslinjer for bruk av kunstig intelligens (KI) ved NMBU*. <https://www.nmbu.no/retningslinjer-bruk-av-kunstig-intelligens-ki-ved-nmbu>
- NRK. (2024 februar). *Regjeringen vil at 80 prosent av offentlig sektor bruker KI – «Urealistisk», mener KI-forsker* [Lest 2. mars 2025]. https://www.nrk.no/norge/regjeringen-vil-at-80-prosent-av-offentlig-sektor-bruker-ki_-_urealistisk_-mener-ki-forsker-1.16843972
- OpenAI. (2025). Introducing Deep Research [Lest 11. mai 2025]. <https://openai.com/index/introducing-deep-research/>
- Oslo kommune. (u.å.-a). *Byråd for finans* [Lest 1. april 2025]. <https://www.oslo.kommune.no/politikk/byradet/byrad-for-finans/>

- Oslo kommune. (u. å.-b). *Byråd for helse* [Lest 1. april 2025]. <https://www.oslo.kommune.no/politikk/byradet/byrad-for-helse/>
- Oslo kommune. (u. å.-c). *Helseetaten* [Lest 12. april 2025]. <https://www.oslo.kommune.no/etater-foretak-og-ombud/helseetaten/>
- Oslo kommune. (u. å.-d). Organisasjonskart Oslo kommune. <https://www.oslo.kommune.no/getfile.php/13519743-1725352281/Tjenester%20og%20tilbud/Politikk%20og%20administrasjon/Politikk/Slik%20styres%20Oslo/Organisasjonskart-TBF-2024%20september%281%29.pdf>
- Oslo kommune. (u. å.-e). *Oslo Origo* [Lest 1. april 2025]. <https://www.oslo.kommune.no/etater-foretak-og-ombud/oslo-origo/>
- Oslo kommune. (u. å.-f). Slik styres Oslo. <https://www.oslo.kommune.no/politikk/slik-styres-oslo/>
- Oslo kommune. (u. å.-g). *Utviklings- og kompetanseetaten* [Lest 1. april 2025]. <https://www.oslo.kommune.no/etater-foretak-og-ombud/utviklings-og-kompetanseetaten/>
- Oslo kommune. (2023a). Møtebok – Bystyret 22. mars 2023 [Lest 12. april 2025]. https://einnsyn-fillager-api.api.oslo.kommune.no/fil?virkomhet=976819853&filnavn=b9f9f61e223b414caf68b8086bbf3653_82a754d3f5d911d4272eaafc58d7c3eb.pdf
- Oslo kommune. (2023b). *Strategi for kunnskapsbasert utvikling, innovasjon og digital teknologi for Oslo kommune* (tekn. rapp.) (Lest 30. mars 2025). Oslo kommune. https://einnsyn-fillager-api.api.oslo.kommune.no/fil?virkomhet=976819837&filnavn=7ee7780ed67a4e3f8f635e5f27b7a130_bf9842b6d6bf77ed1cb3b1d6f3ef883f.pdf
- Rambøll, Digitaliseringsdirektoratet, IKT-Norge, KS - Kommunesektorens organisasjon & NTNU. (2024). *IT i praksis 2024* (tekn. rapp.) (Lest 28. mars 2025). Digitaliseringsdirektoratet. https://www.digdir.no/sites/sogn/files/2024-09/IT_i_praksis_2024.pdf
- Reddy, S. (2024). Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*. <https://implementationscience.biomedcentral.com/articles/10.1186/s13012-024-01357-9>
- Regjeringen. (2020). *Nasjonal strategi for kunstig intelligens* (tekn. rapp.). Regjeringen. <https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf>
- Regjeringen. (2023 februar). *Tid for handling (NOU 2023:4)* (tekn. rapp.). Helse- og omsorgsdepartementet. <https://www.regjeringen.no/contentassets/337fef958f2148bebd326f0749a1213d/no/pdfs/nou202320230004000dddpdfs.pdf>
- Regjeringen. (2024). Forslag til forordning om kunstig intelligens (KI-forordningen) [Lest 11. mai 2025]. <https://www.regjeringen.no/no/sub/eos-notatbasen/>

- notatene / 2021 / juni / forslag - til - forordning - om - kunstig - intelligens - ki - forordningen/id2884935/
- Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns [Lest 7. mars 2025]. *Healthcare*, 11(6), 887. <https://doi.org/10.3390/healthcare11060887>
- Spilker, H. S. (2024). *KI i offentlig sektor: Muligheter og utfordringer* [Masteroppgave]. Norges teknisk-naturvitenskapelige universitet (NTNU). <https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/3141694/no.ntnu%3ainspera%3a187582943%3a68750634.pdf?sequence=1&isAllowed=y>
- SSB. (u. å.). 07685: Sysselsatte, etter næring (SN2007) og kjønn (personer) [Lest 3. mai 2025]. <https://www.ssb.no/statbank/table/07685/tableViewLayout1/>
- SSB. (2020 juni). 11190: Sykelighet. Sykdom, skade eller funksjonshemming (prosent), etter type sykelighet, statistikkvariabel og år. <https://www.ssb.no/statbank/table/11190/tableViewLayout1/>
- SSB. (2024a). Vi blir flere, men også mye eldre [Lest 11. mai 2025]. <https://www.ssb.no/befolkning/befolkningsframskrivinger/statistikk/nasjonale-befolkningsframskrivinger/artikler/vi-blir-flere-men-ogsaa-mye-eldre>
- SSB. (2024b september). *Dobling i bruk av KI* [Lest 28. mars 2025]. <https://www.ssb.no/teknologi-og-innovasjon/informasjons-og-kommunikasjonsteknologi-ikt/statistikk/bruk-av-ikt-i-naeringslivet/artikler/dobling-i-bruk-av-ki>
- Stoltenberg, C. (2024). Deskriptiv [Lest 4. mai 2025]. <https://snl.no/deskriptiv>
- Svardtal, F. (2024). Metaanalyse [Lest 10. mars 2025]. <https://snl.no/metaanalyse>
- Teknologirådet. (2024 mai). *Generativ kunstig intelligens i Norge* (tekn. rapp.) (Lest 2. mars 2025). Teknologirådet. <https://media.wpd.digital/teknologiradet/uploads/2024/05/Generativ-kunstig-intelligens-i-Norge-DIGITAL.pdf>
- Thorsnæs, G., Tvedt, K. A., & Bjorvand, H. (2024). Oslo. <https://snl.no/Oslo>
- Tidemann, A., & Arnesen, L. (2025 februar). Kunstig intelligens. https://snl.no/kunstig_intelligens
- Undeland, T., Wiklund, O., Agasøster, E., Sunde, H. K., Askeland, F. J., Waage, T. L., & Nilsen, A. (2024). *Kunstig Intelligens-Assistert Saksbehandling* (tekn. rapp.). Oslo kommune. <https://1drv.ms/b/s!AlviGXEscLVegfZHgEZF202IkCzIsw?e=cLSR8f>
- Universitetet i Oslo. (u. å.). *Autotekst* [Lest 27. april 2025]. <https://www.uio.no/tjenester/it/lyd-video/autotekst/>
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis [Lest 3. april 2025]. *Nature Human Behaviour*, 8, 2293–2303. <https://doi.org/10.1038/s41562-024-02024-1>

A Øvrige innspill om tematikken fra respondenter

Tabell A.1: Øvrige fritekstsvar fra saksbehandlere om LLM i saksbehandling.

Svar
På høy tid at det innføres, mange prosesser og tekniske oppgaver i datasystemer kunne vært utført av KI
Få prøvd KI ut, stor tro på at det vil være nyttig, spørsmålet er i hvilken grad
Tenker at mange delprosesser i saksbehandlingen bør starte først.
Ikke undervurdere den menneskelige kontakten. Bare se hva som skjedde da «kaffe-koppen» i hjemmetjenesten ble tatt bort. Den lille praten kunne demme opp for ensomhet og isolasjon.
Hvordan få frem nyanseringer ifh hva folk søker og hva som trolig blir det rett etter faglig og skjønnsmessig vurderinger...
Må ikke bli for automatisert da saksbehandling dreier seg om brukernes(søkernes) utfordringer/plager
KI er per i dag ikke integrert i fagsystem (Geric) og man får dermed ikke brukte det i saksbehandlingen. Uklart hvor trygt det er å overføre informasjon om bruker på eksternt KI system. Positiv til at KI kan automatisere en del prosesser, slik at saksbehandlingen går raskere.
Erfaringsmessig implementerer ofte Oslo kommune nye systemer uten at de er tilstrekkelig utprøvd, og uten at nødvendig opplæring er gitt. Dette medfører økt skepsis til nye systemer og rutiner.
Jeg tror kunstig intelligens kan være med på å lage standardtekster og fraser slik at vedtakene ikke ser vesentlig forskjellige ut, basert på hvem som har skrevet det og i hvilken kommune/bydel den er skrevet i. I tillegg kan kunstig intelligens i en større grad gå gjennom alt av lovverk og finne aktuelle paragrafer i forskrifter som vil kunne begrunne en søknad med aktuelt funksjonsnivå. Mye tid går på å innhente eldre opplysninger som gemmer seg i et stort mangfold av journaler og dokumenter, kunstig intelligens kan sile ut og oppsummere på en mer effektiv måte slik at saksbehandler alltid sitter med riktig og nødvendig informasjon for vedtaks grunnlag
Det hadde vært effektivt om vi kunne ha hatt kartleggingssamtaler som tok opp lyd, og som skrev referat fra samtalen.

Svar (forts.)
KI kan sikkert være et nyttig verktøy for å sende ut forvaltningsmeldinger f.els. KI kan ikke erstatte våres vurderinger i enkelte saker. KI kan gi forslag/løsninger men saksbeh.bør ta den endelige avgjørelsen. Gradvis implementering av KI kan være lurt.
prøv det ut asap
Positiv til sortering og prioritering av saker, få oversikt over hvor langt hver sak har kommet, økt rettferdighet, sikre pasientrettigheter. Sikre at det ikke gjøres feil i saksbehandlingen, KI vil gi flere rett til tjenester enn saksbehandler og leder når økonomi er presset. Pasienten har flere rettigheter enn det er økonomi til å oppfylle. Lovverket er veldig romslig.
Tenker at KI i saksbehandling vil kunne være et nyttig og effektivt hjelpemiddel, dersom det blir brukt på riktig måte.
Viktig med opplæring i forhold til hvordan den kan implementeres i saksbehandling
spennende utvikling
Det vil kreve ekstra av oss for å legge inn riktige opplysninger, om vedtak skal skrives automatisk.
Jeg har tro på at «enkle» tjenestesøknader kan behandles av KI ved at brukeren svarer på en rekke spørsmål. Dersom det ikke er «avvik» i svarene, får brukeren automatisk innvilget, men dersom det er «avvik» går saken til en fysisk saksbehandler. Med info om stikkontroller. Heia KI!
KI vil kunne redusere arbeidsmengden ved å legge inn standardopplysninger i Geric og kunne komme med innspill på vurderinger. Det vil fortsatt være behov for at en ansatt går igjennom saksopplysningene for å se om relevant informasjon er med. KI kan brukes til en del standard svar og referere til lovverk i lignende saker. Selv om mange saker er like, vil det allikevel være individuelle og ulike behov som må tas hensyn til.
punkt 8 gikk ikke å besvare med piler, det var lite informativt hvordan man skal bruke scoret, fikk det ikke til elektronisk. Svar på punkt 8. er: punkt 1,3,5 og 6 tar lengst id.

Svar (forts.)

Det er svært viktig å ha med i vurderingen at alle vedtak er basert på innbyggerens individuelle behov og funksjonsnivå, samt hjemmesituasjon. Litt usikker om hvordan KI kan fange opp all relevant informasjon. Ved revurderinger av vedtak må saksbehandlere selv gjennomgå saken på nytt, se om det er endringer i behovet o.l, skal avsluttes o.l. Ikke alle saker skal automatisk forlenges. Det er mye en saksbehandler kan fange opp ved revurderinger, noe jeg ikke vet hvordan KI skal fange opp.

Spennende!

Gleder meg til å ta det i daglig bruk - tror det vil gjøre endel av sorteringen i jobben enklere- Ikke minst når det skal skrives avslag.

Vi bruker store ressurser på å holde tjenestenivået nede. Det er policy nå. I utgangspunktet kan nok alle søknader som kommer til oss vris til at de har rett til tjenester. Jobben består i å holde omfanget nede. Det er krevende å på en måte lete etter argumenter for at søkere skal få et så lite omfang som loven tillater. Dette blir en helhetlig vurdering tatt i et tverrfagligt rom.

B Elektronisk spørreskjema (utskrift)

Bruk av kunstig intelligens i saksbehandling av helse- og omsorgstjenester i Oslo kommune

Denne undersøkelsen er en del av en masteroppgave ved NMBU, der det undersøkes hvordan store språkmodeller (kunstig intelligens, KI) kan forbedre saksbehandlingen av søknader om helse- og omsorgstjenester i Oslo kommune. Ditt svar vil bidra til å kartlegge muligheter og utfordringer ved implementering av denne teknologien.

Store språkmodeller er verktøy som bruker kunstig intelligens for å produsere og forstå tekst. Det mest kjente eksempelet er GPT-modellen som ligger til grunn for blant annet ChatGPT og Copilot, men det finnes andre varianter også. **Det er denne teknologien det refereres til når det står kunstig intelligens eller KI i denne spørreundersøkelsen.**

Mer konkret ønskes det svar på følgende forskningsspørsmål:

- *Hvilke organisatoriske og teknologiske faktorer påvirker saksbehandlers tillit til og aksept for store språkmodeller?*

Tid: Om lag 5 minutter.

Svarfrist: Fredag 4. april / **utsatt svarfrist onsdag 9. april.**

Anonymitet: Svarene er anonyme, og ingen personlige opplysninger vil bli publisert.

Utfyllende informasjon: Se informasjonsskrivet.

* Obligatorisk

Samtykke til deltagelse

1. Jeg har mottatt informasjon om studien, og ønsker å delta. Jeg samtykker til at mine svar behandles frem til prosjektet er avsluttet. *

☐ Ja

☐ Nei

Bakgrunnsspørsmål

2. Hvilken bydel jobber du i? *

- ☐ Bydel Alna
- ☐ Bydel Bjerke
- ☐ Bydel Frogner
- ☐ Bydel Gamle Oslo
- ☐ Bydel Grorud
- ☐ Bydel Grünerløkka
- ☐ Bydel Nordre Aker
- ☐ Bydel Nordstrand
- ☐ Bydel Sagene
- ☐ Bydel St. Hanshaugen
- ☐ Bydel Stovner
- ☐ Bydel Søndre Nordstrand
- ☐ Bydel Ullern
- ☐ Bydel Vestre Aker
- ☐ Bydel Østensjø

3. Hvor lenge har du jobbet med saksbehandling innen helse- og omsorgstjenester? *

- ☐ Mindre enn 1 år
- ☐ 1-3 år
- ☐ 4-7 år
- ☐ 8-12 år
- ☐ Mer enn 12 år
- ☐ Jobber ikke med saksbehandling

4. Hvor ofte jobber du med saksbehandling av søknader om helse- og omsorgstjenester? *

- ☐ Daglig
- ☐ Ukentlig
- ☐ Månedlig
- ☐ Sjeldnere

5. Teknologipåstander (Skala 1-5: 1 = Helt uenig, 5 = Helt enig) *

	1	2	3	4	5
Jeg kjenner godt til KI-teknologien	☐	☐	☐	☐	☐
Jeg bruker mye KI privat	☐	☐	☐	☐	☐
Jeg bruker mye KI på jobb	☐	☐	☐	☐	☐
Jeg er teknologioptimist	☐	☐	☐	☐	☐

Erfaringer med dagens saksbehandling

6. I hvilken grad opplever du at dagens digitale systemer forbedrer saksbehandlingen? (Skala 1-5: 1 = Ikke i det hele tatt, 5 = I svært stor grad) *

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

7. Hvilke utfordringer møter du i dagens saksbehandlingsprosess? (Flervalg) *

- ☐ Tidkrevende informasjonsinnhenting
- ☐ Uenighet mellom saksbehandlere
- ☐ Høy arbeidsmengde og tidspress
- ☐ Vanskelig tilgang til relevante retningslinjer og lovverk
- ☐ Teknologiske systemer som ikke er brukervennlige
- ☐ Krevende kommunikasjon med innbyggeren
- ☐ Annet

8. Hvilke oppgaver tar lengst tid i en gjennomsnittlig saksbehandlingsprosess? (Rangering: Høyest = tar lengst tid, Lavest = kortest tid) *

Kommunisere med andre offentlige instanser (i kommunen eller andre steder)

Kommunisere med innbyggeren

Legge inn data i f.eks. Gerica

Lese og gjennomgå skriftlig dokumentasjon (journaler m.m.)

Diskutere caser med kolleger

Skrive vedtaksbrev

Holdninger til bruk av kunstig intelligens

9. Tror du at KI kan bidra til å effektivisere saksbehandlingen? (Skala 1-5: 1 = Ikke i det hele tatt, 5 = I svært stor grad) *

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

10. Har du noen bekymringer knyttet til bruk av KI i saksbehandling? (Ja/Nei) *

- ☐ Ja
- ☐ Nei

11. Hvis ja på forrige spørsmål, hvilke bekymringer har du? (Flervalg) *

- ☐ Bekymringer om juridiske forhold/personvern
- ☐ Bekymringer om teknologiens nøyaktighet/rettferdighet
- ☐ Bekymringer om egen jobbhverdag
- ☐ Bekymringer om egen eller bydelens teknologiforståelse
- ☐ Annet

12. Hvor enig er du i følgende utsagn? (1 = Helt uenig, 5 = Helt enig) *

	1	2	3	4	5
KI kan bidra til likebehandling i vedtakene	☐	☐	☐	☐	☐
KI kan bidra til å redusere saksbehandling stiden	☐	☐	☐	☐	☐
KI kan redusere feil og misforståelser i saksbehandling	☐	☐	☐	☐	☐
KI kan gjøre det enklere å tolke lovverket korrekt	☐	☐	☐	☐	☐
KI kan gjøre vedtaksbrevene lettere å skjønne for innbyggeren	☐	☐	☐	☐	☐
Jeg er positiv til å teste KI i saksbehandling en	☐	☐	☐	☐	☐

Implementering og bruk i praksis

13. Hvilke deler av saksbehandlingen tror du KI kan brukes til? *(Flervalg)* *

- ☐ Automatisere utkast til vedtak
- ☐ Foreslå relevante lover og retningslinjer
- ☐ Strukturere og kategorisere innsendte søknader
- ☐ Oppsummere tidligere saksbehandling og vedtak
- ☐ Kommunikasjon med bruker og pårørende
- ☐ Foreslå relevant behandling basert på journaler o.l.
- ☐ Ingen av delene
- ☐ Annet

14. I en gjennomsnittlig søknadsprosess, hvor mye tid tror du at du kunne spart dersom KI ble brukt til det du valgte i forrige spørsmål? *(Prosent tidsbesparing fra 0 til 100)* *

Tallet må være mellom 0 og 100

15. Finnes det tjenestetyper der saksbehandlingen kunne blitt gjort fullt og helt av KI? *(Ja/Nei)* *

- ☐ Ja
- ☐ Nei

16. Hvis ja på forrige spørsmål, hvilke tjenestetyper er det snakk om?

17. Hva tror du skal til for at saksbehandlere skal ha tillit til KI?

18. Hva er etter din mening de viktigste barrierene for å ta i bruk KI? (Flervalg) *

- ☐ Mangel på opplæring og kompetanse
- ☐ Usikkerhet rundt kvalitet og pålitelighet
- ☐ Frykt for redusert jobbautonomi
- ☐ Teknologiske begrensninger
- ☐ Skepsis til KI generelt
- ☐ Oppgaver som er for kompliserte for KI
- ☐ Annet

19. Hvor viktig mener du det er at saksbehandlere involveres i utvikling og testing av KI? (Skala 1-5: 1 = Ikke i det hele tatt, 5 = Svært viktig) *

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

Avslutning

Takk for at du tok deg tid til å delta i undersøkelsen! Dine svar vil bidra til forskning om hvordan Oslo kommune kan bruke teknologi for å forbedre saksbehandlingen av søknader om helse- og omsorgstjenester.

20. Har du andre innspill eller tanker om bruk av KI i saksbehandling?

Dette innholdet er verken opprettet eller godkjent av Microsoft. Dataene du sender, sendes til skjemaieieren.

 Microsoft Forms



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway