Bacheloroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

Ekern, Vetle
Våge, Sebastian Sakseid

Hvordan håndtere utfordringer I utviklingsprosjekter tilknyttet kunstig intelligente systemer

En casestudie av Kantega AS

Bacheloroppgave i Digital Forretningsutvikling
Veileder: Casandra Grundstrom
Medveileder: Ursula Sokolaj
Mai 2024

NTNU
Kunnskap for en bedre verden

Ekern, Vetle
Våge, Sebastian Sakseid

Hvordan håndtere utfordringer I utviklingsprosjekter tilknyttet kunstig intelligente systemer

En casestudie av Kantega AS

Bacheloroppgave i Digital Forretningsutvikling
Veileder: Casandra Grundstrom
Medveileder: Ursula Sokolaj
Mai 2024

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

NTNU
Kunnskap for en bedre verden

Sammendrag

Kapabilitetene til kunstig intelligens (KI) har de siste årene opplevd en drastisk utvikling på bakgrunn av teknologiske fremskritt, og har raskt blitt en del av verdiproposisjonene til mange organisasjoner. Spesielt har store språkmodeller fått mye oppmerksomhet de siste årene. De unike egenskapene til slik teknologi byr på nye muligheter, men også nye utfordringer. Blant annet introduserer karakteristikkene til KI et behov for endringer i arbeidsprosessene tilknyttet utviklingen av slike systemer. Denne studien har derfor undersøkt hvordan arbeidsprosessene i utviklingen av KI formes av teknologien, samt hvilke sentrale utfordringer som oppstår på bakgrunn av dette samspillet. For å utforske en slik problematikk benytter denne oppgaven en kvalitativ casestudie av et prosjekt som omhandler utviklingen av en programvareløsning basert på OpenAI sin ChatGPT-4. Studien har benyttet teori om generativ KI, utviklingsprosessen i KI-prosjekter og etablerte utfordringer for KI-prosjekter.

Resultatene gir innsikt i utfordringer våre fire informanter fra caseprosjektet opplever i utviklingsarbeidet. Våre funn indikerer at noen sentrale utfordringer er: (1) teknologiens virkemåte, (2) etablering og kommunikasjon av problem, mål og krav, (3) data og hallusinering og (4) forklarbarhet og lovverk. Oppgaven tydeliggjør påvirkningen disse utfordringene har på utviklingsarbeidet og legger frem mulige håndteringsstrategier. Håndteringsstrategiene fokuserer på forståelsen for en ny arbeidsprosess, sentrale aktiviteter i forkant av utviklingsarbeidets begynnelse, samt aktiv deltakelse fra forskjellige aktører gjennom hele arbeidsprosessen med forståelse for forretningskonteksten og for teknologien. Til slutt identifiseres områder i faglitteraturen som burde addreseres av fremtidig forskning.

Abstract

The capabilities of artificial intelligence (AI) have experienced drastic improvements in recent years due to technological advancements and have quickly become part of the value propositions of many organizations. In particular, large language models have received a lot of attention in recent years. The unique properties of such technology offer new opportunities but also new challenges. The characteristics of AI introduce a need for changes in the work processes associated with the development of such systems. This study has therefore examined how the processes of AI development are shaped by the technology, as well as what key challenges arise from this interplay. To explore such issues, this thesis reports on a qualitative case study focused on a software solution based on OpenAI's ChatGPT-4, currently under development. The study is based on theory about generative AI, the development process in AI projects, and established challenges for AI projects.

The results provide insight into the challenges our four informants from the case project experience in their development work. Our findings indicate that some key challenges are: (1) the non-deterministic nature of the technology, (2) the establishment and communication of problems, goals, and requirements, (3) data and hallucination, and (4) explainability and legislation. The thesis clarifies the impact these challenges have on the development work and presents possible strategies for handling these challenges. The strategies focus on the understanding of a new work process, key activities in the preliminary stages of the development work, and active participation from various stakeholders with an understanding of the business context and the technology. Finally, the thesis develops a research agenda that identifies areas in the literature that should be addressed by future research.

Forord

Denne bacheloroppgaven er skrevet våren 2023 som en avsluttende avhandling for studieprogrammet Digital Forretningsutvikling ved Instituttet for Datateknologi og Informatikk på Norges teknisk-naturvitenskapelige universitet (NTNU). Oppgaven er utformet fra egne interesser sammen med oppgavestiller – Kantega.

Gjennom bachelorgraden har vi fått innsyn i hvordan de tekniske og sosiale elementene i organisasjoner samspiller, og deres gjensidig avhengige natur. I lys av slike perspektiver og den eksplosive veksten innenfor fagfeltet for generativ kunstig intelligens under vår studietid, ble vi videre nysgjerrige for hvilke implikasjoner slik innovativ teknologi gir for etablerte organisatoriske prosesser.

Prosjektet har vært svært lærerikt, men også tidvis veldig krevende. Vi har tatt et dypdykk i et stort og fremoverlent fagfelt, som har utvidet vår faglige horisont og gitt inspirasjon for videre studier. Da det var stor variasjon i mengden etablert litteratur som fantes rundt de ulike tematikkene, fikk vi en prøvesmak på vanskelighetene tilknyttet å drive forskning ved et lite utforsket vitenskapelig landskap. Vi har også lært mye om gledene og utfordringene ved å samarbeide i et slikt intensivt arbeid, og vi ser på dette som verdifull lærdom vi kan ta med oss videre inn i fremtidige prosjekter.

Vi vil gjerne takke Kantega og deres Trondheimskontor, for et spennende samarbeid. Takk til kontaktpunktene våre Øyvind Lillerødvann, Jon Espen Ingvaldsen og Kristin Wulff, for rask kommunikasjon og behjelpelige svar. Vi vil også rette en takk til informantene som fant tid i travle arbeidsdager for å stille til intervju, noe som gav oss dyrebar informasjon til oppgavens datagrunnlag. Til slutt ønsker vi også å gi en stor takk til våre veiledere, Casandra Grundstrom og Ursula Sokolaj, for å ha vært fantastiske sparringspartnere og for å ha bidratt med verdifulle tilbakemeldinger.

Innhold

Figurer x
Tabeller x
Forkortelser/symboler x
1 Introduksjon 11
1.1 Oppgavens hensikt og forskningsspørsmål 12
1.2 Oppgavens struktur 12
2 Teori 13
2.1 Generativ kunstig intelligens 13
2.1.1 Hva er generativ kunstig intelligens 13
2.1.2 Teorier og teknikker for anvendelsen av generativ KI 14
2.2 Utviklingsprosessen for KI-prosjekter 16
2.2.1 Problemforståelse 17
2.2.2 Utviklingsarbeidet 18
2.2.3 Modellovervåking 18
2.3 Etablerte utfordringer for KI-prosjekter 19
2.3.1 Kravspesifisering 19
2.3.2 Konflikter mellom prosessen i KI-prosjekter og agil metodikk 20
2.3.3 Datahåndtering 22
2.3.4 Kvalitet 22
2.3.4.1 Forklarbarhet og lovverk 22
2.3.4.2 Feil i svar og hallusinering 24
3 Metode 26
3.1 Kvalitativ metode 26
3.2 Casestudie 26
3.2.1 Vårt casestudie 27
3.3 Datainnsamling 28
3.3.1 Intervjuer 28
3.3.2 Utvalgsstrategi 29
3.4 Dataanalyse 30
3.5 Forskningsetikk 31
4 Resultater 32
4.1 Teknologiens virkemåte 32
4.2 Etablering og kommunikasjon av problem, mål og krav 34
4.3 Data og hallusinering 35
4.4 Forklarbarhet og lovverk 36

5 Diskusjon 39
5.1 Teknologiens virkemåte 39
5.2 Etablering og kommunikasjon av problem, mål og krav 41
5.3 Data og hallusinering 43
5.4 Forklarbarhet og lovverk 44
5.5 Forbehold og svakheter ved vår forskning 45
6 Bærekraft 47
7 Konklusjon 48
7.1 Videre forskning 50
Referanser 52
Vedlegg 57

Figurer

Figur 1 - Utviklingsprosess for ML-systemer (Nascimento et al., 2019, s. 3).....17
Figur 2 - SDI rammeverket (Tjora, 2019, s. 4, vår oversettelse) 30

Tabeller

Tabell 1 - Etablerte utfordringer i KI-prosjekter 24
Tabell 2 - Deltakerdemografi.....29

Forkortelser/symboler

KI	Kunstig Intelligens
AI	Artificial Intelligence
LLM	Large Language Model
ML	Machine Learning / Maskinlæring
NLP	Natural Language Processing
XAI	Explainable Artificial Intelligence
DPI	Dots per inches
NTNU	Norges teknisk-naturvitenskapelige universitet
PDF	Portable Document Format

1 Introduksjon

Økningen av anvendte kunstig intelligente (KI) systemer har de siste årene vært eksplosiv (Giray, 2021). Det siste tiåret har kapabilitetene til KI-systemer forbedret seg på bakgrunn av fremskritt innen blant annet stordata, algoritmer, maskinlæring og maskinvare (C. Zhang & Lu, 2021). Slike forbedringer har igjen muliggjort fremveksten av disruptive teknologier som selvkjørende biler (Chen et al., 2015), avansert bilde-gjenkjenning (Zoph et al., 2018), og ikke minst naturlig tekstgenerering, som tok verden med storm gjennom utgivelsen av ChatGPT (Leiter et al., 2023). Selv om kunstig intelligens ikke er et nytt konsept, er dens nylige tilgjengelighet og integrasjon på tvers av verdikjedene til bedrifter i ulike sektorer et nytt fenomen. En studie gjort av MIT Sloan Management Review viser at 80% av organisasjonene som ble undersøkt betraktet KI som en strategisk mulighet, og enda flere bedrifter ser på slik teknologi som en mulighet til å oppnå konkurransefordeler (Enholm et al., 2022). I tillegg viser Berente et al. (2021) til estimater som oppgir at halvparten av alle bedrifter aktivt implementerer KI-systemer anno 2020, og mye tyder også på at slike systemer vil ha bred adopsjon de neste årene.

Introduksjonen av KI-systemer i organisasjoner og deres utviklingsprosesser er dog ikke problemfri, da slike systemer differensierer seg stort fra mer tradisjonell programvare. KI skiller seg blant annet fra tidligere teknologier ved at den har større autonomi og muligheter for dyp læring (Berente et al., 2021). En annen unik karakteristikk ved KI er avhengigheten av data, da både kvaliteten og volumet av disse. (Enholm et al., 2022). Samtidig som KI innehar unike tekniske kapabiliteter og utfordringer, skiller den seg også fra annen teknologi fordi det kan være svært vanskelige å tolke beslutningsprosessene til slike systemer (Berente et al., 2021). Dette gir blant annet helt nye etiske og juridiske dilemmaer. På bakgrunn av forskjeller som disse, har inntoget til KI-systemer skapt behov for nye retningslinjer i utviklingsprosjekter, da organisasjoner nå står ovenfor en rekke nye unike utfordringer (Wan et al., 2021; Giray, 2021; Berente et al., 2021).

Hvordan organisasjoner og deres ansatte arbeider med programvareutvikling har siden 1950-tallet vært et sentralt tema i litteraturen tilknyttet utviklingsprosjekter, men fagfeltet har siden den gang gjennomgått store endringer, i likhet med teknologiene slike prosesser underbygger (Boehm, 2006). Med frembruddet av kunstig intelligens står fagfeltet nok en gang ovenfor endringer. Teknologiske differensieringer ved KI, slik som de vi introduserte ovenfor, gir implikasjoner for hvordan man kan effektivt utvikle programvaresystemer. På bakgrunn av utfordringer ved å navigere utviklingsarbeidet tilknyttet KI, virker det å være stor etterspørsel for arbeidsmetodikker og strategier som bidrar til å løse slike (Nascimento et al., 2020). Dette kommer også tydelig frem i forskning, og en litteraturstudie av Giray (2021) viser at antallet primærstudier tilknyttet programvareutvikling av KI-systemer har doblet seg hvert år siden 2018. Det formidles også om at det fortsatt eksisterer et stort gap i litteraturen tilknyttet forståelsen og håndteringen av utfordringene som slikt arbeid medbringer (Nascimento et al., 2020; Giray, 2021; Enholm et al., 2022).

1.1 Oppgavens hensikt og forskningsspørsmål

Fagfeltet for kunstig intelligens er bredt og omfavner både tekniske og organisatoriske perspektiver. På tross av at det finnes store og voksende mengder forskning på feltet, virker de ulike perspektivene å lide av silo-tenkning (Berente et al., 2021). Det kan videre argumenteres for at gapet i litteraturen slik det er beskrevet innledningsvis har bakgrunn i slik silo-tenkning. Denne oppgaven har derfor som hensikt å bistå med innsikt til forskningsarbeidet for å fylle dette kunnskapsgapet, med hensyn til både tekniske og organisatoriske faktorer. Vi søker å gjøre dette ved å kartlegge hvordan arbeidsprosessen i utviklingsarbeidet formes av egenskaper ved teknologien, som igjen krever spesielle strategier for håndtering. For å sikre forskningens integritet og samtidig føye oss til kravene for en bacheloroppgave, må vi dog gjøre nødvendige avgrensninger basert på de empiriske dataene fra caseprosjektet. Vi gjør derfor avgrensninger ved å kun ta for oss utfordringer som var opplevd av informantene; Begrense oss til å kun se på utfordringer tilknyttet arbeidet med store språkmodeller; Kun se på disse utfordringene i lys av utviklingsarbeidet. I henhold til disse begrensningene kan vi konkretisere to forskningsspørsmål som videre underbygger oppgavens hensikt:

> **FS1:** Hvilke utfordringer oppleves i utviklingsarbeidet av et system bygd på store språkmodeller?
>
> **FS2:** Er det mulig å håndtere disse utfordringene, eventuelt hvordan?

Besvarelsen av forskningsspørsmålene vil bidra til oppgavens hensikt ved å belyse samspillet mellom kunstig intelligens som en disruptiv teknologi og arbeidsprosessene i et KI-prosjekt. Verdiene av slik forskning kan videre generaliseres ettersom slik innsikt kan fremme effektiv og vellykket utvikling av KI-systemer, som igjen muliggjør innhøstingen av forretningsverdier ved systemets implementasjon.

1.2 Oppgavens struktur

Oppgaven er delt inn i syv kapitler med et formål om å gjøre lesbarheten bedre og sørge for en tydelig rød tråd. Den første delen er introduksjonen som tar for seg bakgrunnen og hensikten for vår oppgave, forskningsspørsmålene oppgaven skal bidra mot og oppgavens oppbygning. I det andre kapittelet presenterer vi den teoretiske bakgrunnen som skal brukes for å undersøke forskningsspørsmålene. Kapittelet er delt inn i tre hovedtemaer – generativ kunstig intelligens, utviklingsprosess i KI-prosjekter og utfordringer i KI-prosjekter. I kapittel tre vil vi redegjøre for den metodiske tilnærmingen vi har anvendt i forskningsprosjektet. Videre vil resultatene fra den metodiske tilnærmingen presenteres i kapittel fire. Kapittelet sitt formål er å sette fokus på informantene sine opplevelser og meninger om utfordrende aspekter innenfor oppgavens tematikk. I kapittel fem diskuterer vi resultatene og teorien opp mot de presenterte forskningsspørsmålene for oppgaven. Vi vil i tillegg diskutere forbehold og svakheter ved vår forskning. Etter diskusjonen vil vi diskutere bærekraftig programvareutvikling i kapittel 6, før vi presenterer vår konklusjon og forsøker å besvare oppgavens forskningsspørsmål i kapittel syv. Avslutningsvis vil vi også diskutere muligheter for videre forskning innenfor oppgavens tematikk.

2 Teori

I dette kapitlet presenterer vi teorigrunnlaget fra tidligere forskning som er nødvendig for å skape en ramme for forskningsspørsmålene våre. Ettersom en sentral tematikk for oppgaven omhandler generativ kunstig intelligens, vil vi starte med å gi et overordnet innblikk i denne typen teknologi og hva som inngår i viktige tilknyttede begreper. Videre i dette kapittelet vil vi se på hva som kjennetegner KI-prosjekter og hvordan disse skiller seg fra annen programvareutvikling. Til slutt vil vi presentere utfordringer fra litteraturen som omhandler utviklingen av KI-systemer, samt eksperter sine meninger for hvordan disse kan håndteres. Da det kommer nye bidrag til litteraturen tilknyttet KI og dens utfordringer tilnærmet hver dag, og da denne bidragsraten er voksende (Nascimento et al., 2020), er det ikke nødvendigvis etablert konsensus for alt vi legger frem her i teorikapitlet. Utfordringene vi presenterer i dette kapitelet er kun en delmengde av alle utfordringer som kan knyttes opp til utviklingen av KI-systemer, men på bakgrunn av oppgavens omfang har vi valgt å fokusere på de med størst relevans for problemstillingen vår.

2.1 Generativ kunstig intelligens

2.1.1 Hva er generativ kunstig intelligens

Kunstig intelligens (KI) har de siste årene fått mye oppmerksomhet på bakgrunn av sin nye praktiske applikasjon til en rekke disipliner. Det har blitt gjort mange forsøk på å utarbeide en samlende definisjon på hva KI er for å skille teknologien fra annen informasjonsteknologi (Enholm et al., 2022). Blant annet har EU forsøkt å gjøre dette gjennom sitt arbeid med «AI act» (*EU AI Act*, 2023), men deres definisjon har blitt kritisert for å være for bred. Det understreker vanskeligheten av å sette et tydelig skille mellom hvilken teknologi som kan betegnes som kunstig intelligens og ikke. I denne oppgaven tar vi utgangspunkt i definisjonen til Wamba-Taguimdje et al. (2020, vår oversettelse) hvor kunstig intelligens blir definert som *en samling med teorier og teknikker brukt for å lage maskiner som kan simulere intelligens. KI er en generell term som innebærer bruken av datamaskiner for modellering av intelligent atferd, med minimal menneskelig innblanding (s. 4)*. Vi kan bygge videre på denne definisjonen og spesifisere generativ kunstig intelligens som slike modelleringer av intelligent atferd som har kapabilitet til å generere nytt innhold i form av tekst, bilder, kode, m.m.

Konseptet om maskiner som innehar kognitive egenskaper er ikke noe nytt, og den britiske matematikeren Alan Turing blir sett på som en av de første til å introdusere dette konseptet i artikkelen hans «Computing Machinery and Intelligence» (Turing, 1950). Her ble blant annet Turing-testen presentert, som lenge har vært menneskets førsteskanse for testing av de kognitive egenskapene til kunstig intelligens; Og med fremgangen vi har sett på fagfeltet de siste årene blir det diskutert om vi allerede har produsert modeller som består en slik test, da eksempelvis chatboten «Eugene Goostman» (Warwick & Shah, 2016). Dagens eksponentielle vekst for ulike instanser av KI skyldes som nevnt den teknologiske utviklingen innen blant annet maskinvare og prosesseringsalgoritmer (C. Zhang & Lu, 2021; Enholm et al., 2022). Det finnes mange teknikker for å realisere KI, men det er

særlig noen som går igjen: Maskinlæring (ML), Natural Language Processing (NLP), maskinsyn, ekspertsystemer, m.m. I henhold til vår oppgave vil de to førstnevnte være ekstra relevante da chatboter som er basert på store språkmodeller anvender både maskinlæring og NLP (Enholm et al., 2022), vi vil derfor se nærmere på disse i neste underkapittel.

Bruken av KI har et stadig større spenn og brer seg i dag over mange fagfelt. KI-systemer har bruksområder i hele næringskjeden til en bedrift, men hvordan disse brukes er primært fordelt på to kategorier: automatisering og augmentering (Enholm et al., 2022). Automatisering innebærer at maskiner erstatter menneskelig arbeid. Bruken av automatisering er ikke et nytt konsept, men tidligere automatiseringsprosesser har hovedsakelig omhandlet enkle oppgaver. I dag ser vi at maskiner kan erstatte mennesker på stadig vanskeligere kognitive oppgaver; Chatboter kan her være et godt eksempel, og vi kan spesifikt se på Klarna som implementerte en chatbot basert på en stor språkmodell som i dag utfører arbeidet til 700 mennesker (Marks, u.å.). Augmentering står i kontrast til automasjon ved å ikke ha som hensikt å erstatte mennesker, men heller forbedre informasjonsgrunnlaget til mennesker slik at de kan ta bedre beslutninger. Helsesektoren er et eksempel på hvor slike assisterende verktøy brukes; Noen leger har tilgang på KI-systemer som kan informere om indikatorer til kreft i en pasient, på basis av analyserte testresultater. Slik informasjon kan videre brukes av disse legene for å gjøre en informert beslutning (Enholm et al., 2022).

Fra sosioteknisk tradisjon vet vi at utviklingen av ny teknologi ikke er noe som skaper verdi i seg selv, men at verdiene og mulighetene kommer som en effekt av samspillet mellom teknologi og organisasjon (Bouwman et al., 2005). Det samme vil kunne sies om anvendelsen av KI-systemer, og det finnes store muligheter og fallgruver for utnyttelsen av slik teknologi. En bedrifts KI-kapabilitet defineres av Schmidt et al. (2020, vår oversettelse) som *deres evne til å bruke data, metoder, prosesser og folk på en måte som skaper nye muligheter for automasjon, beslutningstaking, samarbeid, o.l. Som ikke ville vært mulig på konvensjonelt vis (s. 3)*. Data står svært sentralt som en teknologisk faktor for å kunne muliggjøre en bedrifts KI-kapabilitet, og Enholm et al. (2020) forklarer at det her er viktig med nok volum og god kvalitet på disse. Svakt datagrunnlag kan også være en hindring for implementasjonen av KI-systemer og kan bidra til blant annet bias og hallusinering, som vi skal diskuterer senere. Forfatterne peker videre på organisatoriske faktorer for yteevnen til KI-systemer, her nevnes blant annet støtte fra ledelsen og kultur som essensielle for suksessfull adopsjon.

2.1.2 Teorier og teknikker for anvendelsen av generativ KI

For å legge grunnlaget til hvordan store språkmodeller fungerer er vi nødt til å først definere det teortiske fagfeltet slik teknologi bygger på: Natural Language Processing (NLP). NLP defineres som samlingen av ulike datateknikker for å automatisk analysere og representere menneskelig språk. Slike teknikker tar for seg oppgaver som blant annet maskinoversettelse, informasjonsgjenfinning, tekst oppsummering og sentiment analyse (Chowdhary, 2020). For at teknikker som disse skal kunne brukes krever det at maskiner har en dyp forståelse av naturlig språk, og derav ulike komponenter av lingvistikk som syntaks, semantikk og pragmatikk. Det er også en viktig forutsetning at maskiner lærer og tilpasser seg i samhandlingen med mennesker, her er maskinlæring en sentral bidragsyter.

Maskinlæring har ikke kun stor relevans for den praktiske anvendelse av generativ kunstig intelligens, men er en fundamental del av KI majoriteten av KI-systemer. Wang (2019, vår

oversettelse) forklarer at *maskinlæring muliggjør at maskinen kan lære uten eksplisitt programmering. Denne læringsprosessen gjøres av maskinen selv gjennom innsamling av data, datanalyse og ved å gjøre forutsigelser (s. 2).* Maskinlæring impliserer derfor at en maskin kan imitere menneskelig oppførsel ved å iterere på egen kunnskap og dermed lære, samt øke nøyaktighet for output. Innenfor maskinlæring finnes det igjen ulike metoder som kan anvendes, disse kan overordnet deles inn i: veiledet læring (eng: «supervised»), ikke-veiledet læring (eng: «unsupervised» og forsterket læring (eng: «reinforcement learning») (Enholm et al., 2022). I veiledet læring er treningsdataen markert med en målverdi slik at systemet kan finne mønstre som kommer til samme slutning som målverdien. Ikke-veiledet læring har derimot ikke målverdiene inkludert, og systemet må selv finne strukturer i treningsdataen. Det er også vanlig å kombinere veiledet og ikke-veiledet læring i treningen av modeller, slik metode kalles semi-veiledet læring (engelsk: «semi-supervised learning»). Forsterket læring står i kontrast til de forekommende metodene ved at det ikke baserer seg på forhåndsetablerte data, men heller innhenter informasjon gjennom interaksjon med et miljø. Modellen blir så vurdert og lærer basert på tilbakemeldingen den får (Enholm et al., 2022).

Tidlige anvendelser fra NLP-feltet var enkle maskiner som oversatte ord for ord uten noen kontekstuell sammenheng (Nadkarni et al., 2011). Feltet har siden den gang hatt stor vekst, og sammen med maskinlæring legger det i dag grunnlaget for teknologier som ChatGPT-4. Slik teknologi bygger videre på disse teoretiske rammeverkene og kombinerer dem med praktiske metoder som transformatorer for å effektivt prosessere og respondere til naturlig språk. En transformator kan defineres som en modellarkitektur som muliggjør effektiv behandling av data for å finne komplekse sammenhenger mellom disse (Vaswani et al., 2017). Dagens transformatorer skiller seg fra tidligere modeller, som «Recurrent Neural Networks» (RNN) og «Long Short-Term Memory networks» (LSTM), ved å prosessere og vekte relevansen av større sekvenser med data samtidig, i motsetning til å se på disse sekvensielt. Kjernemodulen i transformatorarkitekturen som muliggjør dette er en selv-oppmerksomhetsmekanisme (engelsk: «self-attention mechanism») som tillater modellen til å se på hver token (tekstbit) i inngangsverdiene parallell i en større kontekst. Dette tillater for vesentlig større effektivitet i prosesseringen av data, samt også mer kontekstuell presisjon ved utførelsen av en rekke NLP-oppgaver (Vaswani et al., 2017).

Transformatorarkitekturen og selv-oppmerksomhetsmekanismen ble først introdusert i «Attention is all you need» av Vaswani et al. (2017) og har siden da fått annerkjennelse som en nøkkelfaktor for den eksponentielle veksten av kraftige store språkmodeller de siste årene. Store språkmodeller, også kjent som Large Language Models (LLM), er tilpassede maskinlæringsmodeller som ofte baserer seg på transformatorarkitektur (Chang et al., 2024), kjente eksempler på LLMer er OpenAI sin ChatGPT og Google sin BERT. Disse modellene kan spesialiseres til å utføre ulike NLP-oppgaver, og mens styrken til førstnevnte er å generere relevant sammenhengende tekst basert på inngangsdata, er BERT spesialisert til å forstå og analysere tekst (Devlin et al., 2019). Selv om store språkmodeller som ChatGPT-4 i dag er svært gode til å effektivt utføre en rekke oppgaver, har disse også betydelige svakheter. En sentral egenskap med LLMer og andre former for KI-systemer er at de består av ikke-deterministiske algoritmer, dette er algoritmer som kan produsere ulike resultater selv med like inngangsverdier (Cormen & Leiserson, 2022). Disse algoritmene kalkulerer statistiske estimater om hvilken token som skal være den neste gitt konteksten. Modellene bedriver derfor i essens en kompleks gjettelek (Y. Zhang et al., 2023).

Et annet viktig aspekt med store språkmodeller er hvordan de trenes og datasettene som inngår i denne prosessen. Treningen forgår hovedsakelig i to faser: «pre-training» og «fine-tuning» (Devlin et al., 2019). Den første fasen skal lære modellen å tolke språklige sammenhenger ved å la den utføre oppgaver på ekstremt store datasett selv. I tilfellet med BERT ble «pre-training» gjennomført på en stor database av bøker (800 millioner ord), samt den engelske Wikipedia databasen (2500 millioner ord). BERT fikk så instruksjoner om å utføre to oppgaver med disse dataene; Første oppgave gikk ut på å gjette et maskert ord i en tekst ut ifra konteksten; Andre oppgave var å avgjøre om en gitt setning var en logisk fortsettelse basert på en forløpende setning (Devlin et al., 2019). Etter generell «pre-training» er ferdig brukes «fine-tuning» for å spesialisere modellen til ulike ønskede oppgaver. I denne prosessen er datasettene mindre og mer oppgavespesifikke, samt er læringsraten ikke like sterkt vektet da modellen nå bare skal justeres og ikke trenes fra bunnen. For denne fasen evalueres resultatene til modellen og det gjøres flere iterasjoner for å optimalisere ytelsen Devlin et al., 2019)

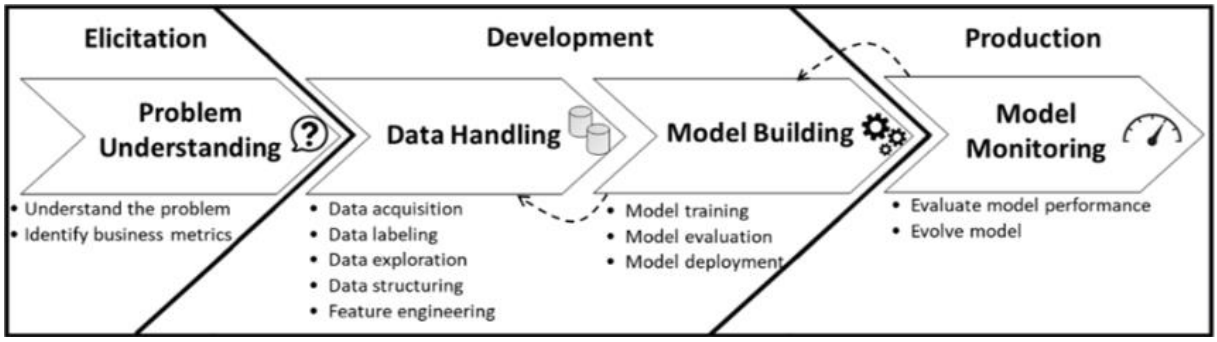
Selv om en språkmodell er trent på en enorm mengde data og spesialisert for å utføre en spesifikk oppgave, vil en svakhet i modellen være at den nå ikke vil få inn mer data og at den ikke har tilgang på ekstern data. Dermed vil den ikke kunne gi gode svar til brukere som eksempelvis lurer på nylige hendelser fra nyhetsbildet eller spesifikk informasjon fra den interne databasen til selskapet deres. Uttrykket «Retrival-augmented generation» (RAG) ble popularisert i en av artikkel av Lewis et al. (2020) og beskriver en av løsningene som har vært mest effektive for å løse dette problemet. RAG-metodikken gir språkmodellen muligheten til å dynamisk hente inn ekstra kontekst for å kunne generere bedre svar på spørsmål ved å hente ut datasegmenter fra eksterne datakilder (Gao et al., 2024). Denne metodikken har blitt svært populær å bruke for adopsjon av store språkmodeller i en forretningskontekst da den tillater for at slike systemer kan utnytte interne data, og senker sannsynligheten for hallusinering som vi skal se nærmere på senere i dette kapittelet.

2.2 Utviklingsprosessen for KI-prosjekter

For å forstå hvilke utfordringer som er tilknyttet prosjekter som utvikler kunstig intelligente systemer, er det viktig å forstå hvordan arbeidet i slike prosjekter foregår. De unike egenskapene ved teknologien i KI-systemer medbringer endringer i programvareutviklingsprosessen, i forhold til tidligere etablerte praksiser for tradisjonell programvareutvikling som agil metodikk og fossefallsmetoden (Nascimento et al., 2019; Vial et al., 2023). Data utgjør kjernen av KI- og ML-systemer (Enholm et al., 2022; Wan et al., 2021), og programvareløsninger som bygger på slike systemer er derfor svært datadrevne (Studer et al., 2021). Data sin sentrale rolle har sterk betydning for arbeidet med utviklingen og fører til en eksperimentell og sterkt iterativ prosess (Wan et al., 2021). I tillegg styrker det behovet for kontinuerlige oppdatering og forbedring av systemene gjennom programvareutviklingsprosessen (Wan et al., 2021).

I caseprosjektet vi studerer arbeides det med å utvikle en løsning som spesialiserer en stor språkmodell og trener den med kundedata ved anvendelse av teknikker som RAG. Selv om forespørselen etter slike verktøy har vokst eksplosivt, er fagfeltet for arbeid med metoder som RAG fortsatt ferskt, og det finnes svært lite litteratur som beskriver gode rammer for slike prosjekter. Ettersom prosjektet omhandler programvareutvikling basert på et KI-system skiller utviklingsprosessen seg betraktelig fra etablerte tradisjonelle metodikker som for eksempel agil metodikk (Vial et al., 2023). Vi har derfor valgt å ta utgangspunkt i en modell for utvikling av maskinlæringssystemer i denne delen av teorigrunnlaget. Store språkmodeller som chatGPT-4 baseres hovedsakelig på maskinlæringssystemer og NLP,

som nevnt i del 2.1.2, og i arbeidet med "fine-tuning" står data svært sentralt. Selv om vi ser ulikheter mellom utviklingen av LLMer og den forretningsmessige tilpasningen av disse, betrakter vi prosessering av data som kjernen for dem begge. Vi kommer derfor til å ta utgangspunkt i en prosessmodell Nascimento et al. (2019) har utviklet gjennom sin studie av arbeidsflyt i utviklingen av maskinlæringssystemer i små virksomheter. I artikkelen deler de arbeidet opp i tre deler; (1) informasjonsinnhenting, (2) utvikling, (3) produksjon, som omfatter fire faser; (1) problemforståelse, (2) datahåndtering, (3) modellbygging og (4) modellovervåking. Samtidig presiserer forfatterne at prosessen er sterkt iterativ og at aktivitetene i et ML-prosjekt ofte hopper frem og tilbake mellom steg underveis (Nascimento et al., 2019). Pilene som går fra fjerde til tredje fase og fra tredje til fjerde fase understreker dette i figur 1. Vi kommer også til å supplere prosessmodellen med betraktninger gjort i andre studier, som kan sees i sammenheng med arbeidet som beskrives i de forskjellige fasene.



Figur 1 - Utviklingsprosess for ML-systemer (Nascimento et al., 2019, s. 3)

2.2.1 Problemforståelse

Den første fasen i modellen presentert av Nascimento et al. (2019) omhandler aktiviteter som skjer i forkant av selve utviklingen og defineres som problemforståelse. I denne fasen jobber utviklerne med å etablere en forståelse for problemet som skal adresseres, samt å definere mål for prosjektet. Nascimento et al. (2019) skriver om at utviklingsteamet forsøker å identifisere hvilke kvantifiserbare beregninger virksomheten allerede bruker for å overvåke og vurdere ytelsen og effektiviteten til den gjeldende forretningsprosessen (eng: «business metrics»). Sokolaj et al. (2023) utdyper forklaringen av arbeidet i fasen, ved å understreke at utviklerne analyserer problemet i samarbeid med interessenter og bruker dette til å formulere krav, omfang og mål ved prosjektet. På denne måten kan man etablere en forståelse over akkurat hva kunden ønsker å oppnå med prosjektet og danne et solid grunnlag som bidrar til å sikre klarhet og retning for hele prosessen (Bandi et al., 2023). Namvar et al. (2023) argumenterer også for at kvaliteten på systemet har tilknytning til hvor godt utviklerne forstår problemet som skal adresseres.

Andre studier som omhandler prosessen for utviklingen av produkter og tjenester som maskinlæring er en del av, trekker også frem forståelse for forretningsproblem som et viktig første steg (Studer et al., 2021; Wan et al., 2021). Studer et al. (2021) har gjennom sin studie utviklet et forslag til et industri- og bruksnøytralt rammeverk for utvikling av ML-systemer med fokus på kvalitetssikring, kalt CRISP-ML(Q). Den første fasen i rammeverket omhandler blant annet forståelsen av forretningen og hva løsningen skal oppnå, i tillegg til forståelse for dataen som skal brukes. Wan et al. (2021) forklarer at den første fasen også brukes av utviklere til å kommunisere kapabiliteter og begrensninger ved teknologien til interessenter for å håndtere forventninger.

2.2.2 Utviklingsarbeidet

Etter å ha definert problemet som skal løses og målene ved prosjektet, er neste del av prosessen selve utviklingen av løsningen (Nascimento et al., 2019). Nascimento et al. (2019) deler dette steget inn i to faser; (1) datahåndtering og (2) modellbygging. Datahåndteringsfasen omhandler aktiviteter knyttet til dataen som skal anvendes i modellen, som datainnsamling, datautforskning, datastrukturering og preprosessering av dataen (Nascimento et al., 2019). Essensen i arbeidet er å finne ut hvilken data som er nødvendig for å møte kundens mål (Nascimento et al., 2019). I modellbyggingsfasen av utviklingen gjennomføres trening, tuning og innledende testing av maskinlæringsmodellen på forhånsdefinerte evalueringskriterier (Nascimento et al., 2019). Basert på resultatene av disse testene tas en avgjørelse på om modellen er god nok til å settes i produksjon.

I det tidligere nevnte CRISP-ML(Q) rammeverket (Studer et al., 2021) trekkes også arbeid med dataen etterfulgt av arbeid med modellen frem som de påfølgende stegene etter forretningsforståelsen er dannet. Arbeidet med data kalles for dataforberedelse og deles opp i valg av data, vasking av data og konstruering av data (Studer et al., 2021). I modelleringen trekkes blant annet definering av kvalitetsmål for modellen, modellvalg, modelltrening og validering av ytelsen frem som viktige momenter.

I Wan et al. (2021) sin studie av hvordan maskinlæring endrer praksiser for programvareutvikling understreker forfatterne fokuset på data som en viktig årsak i utviklingsarbeidet. Artikkelen påpeker at arbeidet med design og bygging av programvaren hovedsakelig omhandler håndtering av data, til forskjell fra tradisjonell programvareutvikling hvor kreativitet og koding utgjør grunnpilaren. Prosessen med å håndtere data er betydelig mer kompleks enn å produsere programvarekode (Amershi et al., 2019).

I en studie av et konsulentfirma i Nord-Amerika som leverer KI-løsninger til kunder, observerte forfatterne at utviklingsprosessen i KI-prosjektene konsulentfirmaet gjennomførte var inspirert av agile metoder som SCRUM og Kanban (Vial et al., 2023). Aspekter fra slik metodikk som kundeinvolvering for kunnskapsoverføring, samt muligheten for å avdekke og adressere problemer tidlig i prosessen ble opplevd som nyttig i KI-prosjektene. Allikevel presiserer forfatterne at det tekniske arbeidet var annerledes (Vial et al., 2023). I artikkelen blir arbeidet med å forbedre KI-løsninger sammenlignet med vitenskapelig arbeid i form av å være svært eksperimentelt med hypoteser som bekreftes eller avkreftes (Vial et al., 2023). Utviklerne som ble intervjuet i studiet opplevde at man ikke vet om noe fungerer i den gitte konteksten før man har prøvd, og at prosessen dermed blir sterkt iterativ.

2.2.3 Modelloverv̊aking

Etter at modellen er satt ut i produksjonsmiljøet, blir den kontinuerlig overvåket (Nascimento et al., 2019). Ny data blir lagt til modellen underveis og utviklerne er nødt til å overvåke ytelsen til systemet og vurdere å gjøre endringer, på bakgrunn av systemets selvdrivende læring (Nascimento et al., 2019; Wang et al., 2019).

2.3 Etablerte utfordringer for KI-prosjekter

Som tidligere nevnt fører de særegne egenskapene til teknologien og utviklingsprosessen til utfordringer i KI-prosjekter. I følge Wan et al. (2021) har ML-systemer et sett med unike utfordringer i tillegg til alle som er gjeldende for vanlige programvaresystemer. Vi vil i dette delkapittelet legge frem noen sentrale utfordringer fra forskningslitteraturen, samt tilhørende mulige håndteringsstrategier.

2.3.1 Kravspesifisering

En sentral utfordring fra litteraturen om KI-prosjekter innebærer utforming og spesifisering av krav (Nascimento et al., 2020). Fra et identifisert forretningsproblem kan man identifisere relevante forretningsmetrikker hos kunden, som igjen brukes til å definere mål, omfang og til slutt krav for prosjektet (Nascimento et al., 2019). Kravene for systemer som er basert på ML er datadrevne og dermed utfordrende å spesifisere i korrelasjon til forretningsmål (Wan et al., 2021). Allikevel understreker forfatterne at forhåndskunnskap om dataen og forretningskonteksten kan føre til viss form for determinisme i kravspesifisering. I en studie av programvareutviklingstilnærminger trekker Bosch et al. (2018) frem data-/utfallsdreven utvikling. Tilnærmingen går ut på å sette krav som kvantifiserbare mål som jobbes mot på en eksperimentell måte, og tilnærmingen anses som passende for datadrevne prosjekter. I følge (Nascimento et al., 2020) omhandler utfordringene tilknyttet kravspesifisering i ML-prosjekter å skaffe forståelse for kundens behov og forretningsproblemet, innhenting av krav, samt å sette krav til dataen. Det trengs derfor nye retningslinjer for å identifisere, beskrive, analysere og håndtere krav (Nascimento et al., 2020).

Nahar et al. (2022) skriver om samarbeidet mellom produktteam og modellteam i ML-baserte utviklingsprosjekter, basert på 45 intervjuer av involverte i ML-prosjekter. Produktteamet defineres som de involverte som er ansvarlige for løsningen og eier dataen (som regel en kunde), mens modellteamet omfatter de involverte med ansvar for å utvikle ML-modellen og dens ytelse. I artikkelen diskuteres blant annet arbeidet med å definere krav for programvareløsninger basert på ML-systemer. I tilfellene der krav til produktet ble spesifisert uten involvering fra personer med kompetanse og innsikt i ML sine kapabiliteter, forpliktet prosjektene seg ofte til urealistiske forventninger til hva teknologien kan oppnå. Det er altså vanskelig å sette realistiske krav til produktet uten å ha forståelse for kapabilitetene til teknologien (Nahar et al., 2022). I tilfellene der krav ble spesifisert uten involvering av personer med kompetanse og innsikt i forretningskonteksten til produktet ble det ofte et for sterkt fokus på tekniske krav, mens krav til hva produktet skulle oppnå forretningsmessig ble neglisjert. Derfor anbefaler forfatterne at kravspesifiseringen for programvare basert på ML-modeller, skjer gjennom tett samarbeid, interaksjon og forhandling mellom produkt- og modellteamet for å utforske hva som er oppnåelig (Nahar et al., 2022). På denne måten kan modellteamet kommunisere kapabilitetene til teknologien for å sette de riktige forventningene hos produktteamet, samt eliminere urealistiske krav tidlig i prosessen. I tillegg kan de involverte som forstår forretningskonteksten teknologien skal fungere som en del av, øke forståelsen for hva som er ønskelig å oppnå med løsningen (Nahar et al., 2022).

Vial et al. (2023) skriver om at konsulentfirmaet han studerte tok utgangspunkt i kundens forretningsmetrikker tilknyttet forretningsproblemet for å levere forretningsverdi til kunden. KI-prosjektene bruker en innledende fase av prosjektet til å skaffe forståelse for forretningsbehov sammen med kunden (Vial et al., 2023). Deretter ble forretningsmetrikker kunden brukte for å måle ytelsen til en forretningsprosess identifisert,

etterfulgt av en vurdering av hvordan metrikkene kan knyttes til mål om nøyaktighet og presisjon i KI-modellen. Allikevel ble det oppfattet som utfordrende å oversette forretningsmetrikkene til mål som brukes for å evaluere ytelsen til en KI-modell (Vial et al., 2023). Konsulentene i firmaet strevde derfor etter å være kontinuerlig bevisste på kundens kontekst, behov og mål for å sikre at alt teknisk eller vitenskapelig arbeid som utføres på KI-prosjektene er rettet mot å skape forretningsverdi for kunden (Vial et al., 2023). Konsulentfirmaet var også opptatt av å evaluere hvorvidt forbedring av ytelsen til en KI-modell faktisk bidrar til økt forretningsverdi for interessenter, og forfatterne av artikkelen påpeker at det ikke alltid er tilfellet. Det er alltid mulig å forbedre den tekniske ytelsen i KI-løsninger, men arbeid mot slike forbedringer burde bare prioriteres dersom det bidrar mot prosjektets forretningsmål (Vial et al., 2023). Ifølge (Passi et al., 2020) trenger forretningssiden vanligvis en datadreven løsning som fungerer «godt nok», ikke perfekt.

I tillegg trekker Vial et al. (2023) frem at konsulentfirmaet forfatterne studerte tar i bruk såkalte «power couples» bestående av en forretningskonsulent og en data scientist. Forretningskonsulenten sitt ansvar er å konkretisere hva løsningen skal oppnå og å tolke forretningsbetydningen av resultatene underveis, mens data scientisten fokuserer på hvordan justeringer i teknologien skal gjennomføres. Paret jobber tett sammen gjennom utviklingsprosessen og diskuterer midlertidige resultater produsert av modellen for å bane veien videre (Vial et al., 2023). Dette gjøres for å sikre at arbeidet er rettet utover mot å skape forretningsverdi for kunden, istedenfor innover på teknologien (Vial et al., 2023). Et slikt samarbeid trekkes frem som en viktig suksessfaktor for konsulentfirmaet.

2.3.2 Konflikter mellom prosessen i KI-prosjekter og agil metodikk

Anvendelsen av agil metodikk i programvareutviklingsprosjekter har blitt en etablert og populær praksis på bakgrunn av rapporterte økninger av suksessraten for slike utviklingsprosjekter (Khalil & Kotaiah, 2017). Vial et al. (2023) gjennomførte en studie av en konsulentvirksomhet som jobbet med utvikling av KI-systemer med relativt stor suksess. I KI-prosjektene deres anvendte de aspekter ved agil metodikk, men erfarte samtidig at det oppsto konflikter mellom etablerte prinsipper i metodikken og den eksperimentelle arbeidsflyten i KI-prosjekter. Konfliktene som oppstår, har konsulentvirksomheten opplevd som utfordrende å håndtere (Vial et al., 2023). Videre trekker forfatterne frem blant annet organiseringen av arbeidsoppgaver, kilder til endring og måling av progresjon som eksempler på slike konflikter.

I agil metodikk er organiseringen av arbeidsoppgaver basert på iterasjoner av fast lengde der utviklingsteamet jobber mot å fullføre oppgaver som ble valgt under et sprintplanleggingsmøte (Vial et al., 2023). Den eksperimentelle arbeidsflyten i KI-prosjekter gjør det derimot betydelig utfordrende å estimere tidsbruken på oppgaver, ettersom utviklerne ikke vet hva som kommer til å fungere og ikke. I tillegg fører midlertidige resultater i eksperimenteringen noen ganger til at oppgavene endrer seg underveis i en iterasjon. Det er derfor utfordrende å sette opp et sett med forhåndsdefinerte arbeidsoppgaver som skal være fullført på slutten av en iterasjon.

Både agil metodikk og arbeidsflyten i KI-prosjekter involverer endringer underveis i arbeidet, men kildene til endringene som skjer er forskjellige (Vial et al., 2023). I agil metodikk er det vanligvis kunden som er driveren for endringer underveis i prosjektet, men i KI-prosjekter oppstår ofte endringer på bakgrunn av midlertidige resultater i eksperimenteringen eller tilgjengeligheten av nye teknikker. Slike endringer kan påvirke prosjektets omfang og i verste fall gjennomførbarheten (Vial et al., 2023).

En annen konflikt som trekkes frem er at det i agil metodikk fokuseres på å levere fungerende og håndgripelige løsninger underveis. Det er ikke nødvendigvis mulig i KI-prosjekter (Vial et al., 2023). Ettersom store deler av arbeidet er eksperimentelt, er det ikke alltid slik at fullførte arbeidsoppgaver resulterer i ekstra funksjonalitet eller et produkt i det hele tatt. Som følge av dette er det utfordrende å skape en opplevelse av progresjon, og å demonstrere progresjonen og verdien av arbeidet som er gjennomført for kunder.

Samtidig belyser forfatterne strategier som har blitt anvendt for å håndtere disse konfliktene (Vial et al., 2023). For å håndtere konfliktene ovenfor trekker Vial et al. (2023) frem 2 strategier konsulentfirmaet i studien har implementert. Forfatterne presiserer at disse strategiene ikke er de eneste måtene å håndtere konfliktene på, men at det er praksiser konsulentvirksomheten de studerte opplevde som nyttige. Strategiene går ut på å ikke være redd for å sparke kunden og å redefinere egen oppfatning av en «ferdig» oppgave (Vial et al., 2023).

For konsulentvirksomheten var det viktig å muliggjøre at de kunne avslutte et prosjekt underveis og på den måten sparke sin kunde, noe som strider med prinsipper fra tradisjonell programvareutvikling hvor leveransen av prosjektet i stor grad blir etablert og avtalt i starten av prosjektet (Vial et al., 2023). I KI-prosjekter kan derimot ikke resultater alltid bli levert etter planen (Vial et al., 2023). Forfatterne forklarer at man ikke kan vite om et ønsket KI-produkt er oppnåelig før man har prøvd (Vial et al., 2023). Resultatene av eksperimenteringen underveis kan avdekke hvorvidt det er mulig å oppnå suksess i prosjektet eller ikke. I tillegg understreker forfatterne at teknologien og teknikker stadig endrer seg i fagfeltet og at det derfor kan være ugunstig for både kunden og konsulentvirksomheten å holde på et dårlig prosjekt (Vial et al., 2023). Derfor gjennomfører virksomheten en vurdering av en kunde sin KI-beredskap før utviklingsarbeidet med løsningen begynner. I denne fasen analyserer de kunden sin datakvalitet, teknologi og infrastruktur, samt deres innstilling til konfliktene oppsummert ovenfor (Vial et al., 2023). De vurderer hvorvidt en kunde er mottakelig for at leveransene underveis ikke alltid vil være fungerende løsninger, at endringer for prosjektets mål og omfang kan oppstå på grunn av andre faktorer enn kunden sine ønsker, samt at gjennomførbarheten av prosjektet er usikker gjennom store deler av utviklingsarbeidet. I tillegg praktiserer de en stage-gate-tilnærming til utviklingsarbeidet hvor prosjektets gjennomførbarhet vurderes basert på etablerte suksesskriterier med jevne mellomrom. I følge Vial et al. (2023) er slike kontinuerlige vurderinger spesielt viktig innenfor KI-prosjekter ettersom stadige endringer ved teknologien og teknikker betyr at det kan være problematisk å jobbe med et dårlig prosjekt over lengre tid.

En annen strategi konsulentvirksomheten anvendte var å redefinere betraktningen av hva det vil si at en oppgave er «ferdig» (Vial et al., 2023). Strategien innebærer å endre definisjonen av hva som er en oppgave, samt hva som burde anses som et resultat (Vial et al., 2023). Istedenfor å etablere arbeidsoppgaver basert på funksjonalitet burde arbeidsoppgavene i KI-prosjekter baseres på hypoteser om hvilke justeringer som kan fungere, og brytes ned til eksperimentene som må gjennomføres for å oppnå et svar på hypotesen. I tillegg var det viktig for konsulentvirksomheten at eksperimenter som ikke resulterte i ønskede resultater, ikke ble ansett som mislykkede. En falsifisert hypotese er et viktig resultat som driver prosjektet fremover ettersom utviklerne kan fokusere på andre hypoteser og eksperimenter (Vial et al., 2023). Forfatterne forklarer også at det ble ansett som viktig å kommunisere og demonstrere falsifiserte hypoteser og tilhørende lærdommer

som progresjon til en kunde. Dette ble gjort i regelmessige sprintgjennomganger og bidro til å øke kundens forståelse for arbeidsprosessen.

2.3.3 Datahåndtering

I kjernen av KI-løsninger står data (Enholm et al., 2022). Datahåndtering utgjør dermed kjernen i utviklingsprosessen av KI-programvare og Nascimento et al. (2020) beskriver oppgavene relatert til datahåndtering i AI/ML-utvikling som utforskning av data, forberedelse av data, samt vasking av data. Videre forklarer forfatterne at KI-programvare egentlig omhandler KI-data og programvareutvikling. I følge Nasciemeto et al. (2020) omfatter utfordringene knyttet til datahåndtering arbeid med innsamling, prosessering, tilgjengelighet og kvalitet. Det tidligere nevnte rammeverket CRISP-ML(Q) inkluderer de samme oppgavene tilknyttet datahåndtering i et ML-prosjekt, men presiserer også at dataforståelse er en kritisk del av håndteringen som burde ligge på plass først (Studer et al., 2021). Rammeverket understreker samtidig at data må forstås i sammenheng med forretning på grunn av at forretningsmål ved en løsning kan bli avledet og endret av dataen som er tilgjengelig. Å skaffe en god forståelse for dataen som er tilgjengelig tilknyttet forretningsproblemet for en KI/ML-løsning betraktes altså som et essensielt steg for datahåndteringen i et prosjekt.

I følge Nahar et al. (2022) er som regel ikke teamet som er ansvarlig for å lage ML-systemet, de som eier og har kjennskap til dataen. Eksisterende datadokumentasjon er også ofte mangelfull og utilstrekkelig for å skape en god nok forståelse for dataen (Nahar et al., 2022). Dette fører til at etableringen av dataforståelse utgjør et sentralt samarbeidspunkt mellom interessenter i prosjektet. I følge Nahar et al. (2022) er behovet for domeneeksperter for å adressere utfordringene tilknyttet dataforståelse regelmessig nevnt i forskningslitteratur, i tillegg til farene ved å bygge et system med utilstrekkelig forståelse av dataen. En av informantene i studien fortalte om at det hadde vært ideelt for utviklerne å tilbringe en eller to uker med en person for å skaffe tilstrekkelig forståelse for dataen. Dataforståelsen og tilgangen til domeneeksperter beskrives som en flaskehals i ML-prosjekter.

Allikevel oppstår det ofte utfordringer på bakgrunn av at domeneeksperter har begrenset tilgjengelighet ettersom de har andre ansvarsområder (Park et al., 2021). En av informantene i studien gjennomført av Nahar et al. (2022) påpekte at selv om en domeneekspert er involvert i er prosjekt, er ikke modellteamet nødvendigvis i kontakt med dem hele tiden. Dette kan føre til at de involverte med teknisk kompetanse feiltolker data og bygger videre på det. I følge Nahar et al. (2022) er det å sikre tilgang til domeneeksperter viktig, samt å planlegge for tilgangen tidlig i prosjektet.

2.3.4 Kvalitet

I litteraturgjennomgangen til (Nascimento et al., 2020) deles kvalitet for KI-systemer opp i en rekke attributter. Av disse blir blant annet forklarbarhet og feil i svarene pekt ut som sentrale utfordringer som kan bidra til å redusere kvaliteten til en KI-løsning. Vi vil her se nærmere på disse og deres implikasjoner.

2.3.4.1 Forklarbarhet og lovverk

Samtidig som kapabilitetene til KI-systemer har økt drastisk de siste årene har også kompleksiteten av slike systemer blitt større. Disse systemene gjør i dag beslutninger som påvirker menneskers liv på liten og stor skala, og det har derfor vokst frem et behov for å forstå hvordan slike beslutninger blir tatt (Arrieta et al., 2020). Dagens

maskinlæringsalgoritmer kan dra nytte av flere millioner parametere, hvor hver av disse blir prosessert og vektet på tvers av hundrevis av lag og filtre. God forståelse for hvordan slike systemer fungerer er derfor blitt et viktig tema for å senke risikoen tilknyttet anvendelsen av KI-systemer, og for å legge til rette for at kunstig intelligens kan være rettferdig og troverdig (Arrieta et al., 2020). Forklarbar KI eller «eXplainable AI» (XAI) betegnes i Arrieta et al. (2020, vår oversettelse) som en *kunstig intelligens som gir detaljer eller grunnlag for hvordan den fungerer og gjør dette enkelt å forstå for et gitt publikum(s. 6)*. Transparens er et annet fagbegrep som brukes i samme kontekst og Arrieta et al. (2020) forklarer at en KI-modell er transparent hvis den kan forståes i seg selv. Altså er transparent-modell det motsatte av en «black-box»-modell hvor virkemåten er skjult for brukere. En utfordring med forklarbarheten av KI-modeller er at modellene er ikke-reduserbare og at forklaringer av løsningene vil kunne være like store som løsningene selv (Nascimento et al., 2020). Forklarbarhet skaper også barrierer for praktisk implementasjon av KI da det store spennet mellom forskningsmiljøer og forretningssektorer gjør det til en større risiko å implementere slike teknikker. Det legges også vekt på at forskningen tilknyttet KI-modeller virker å ha skyggelapper på, og prioriterer ytelse over forståelse, som videre kan hindre nytteverdiene til slike systemer (Arrieta et al., 2020).

Utfordringer med forklarbarhet er resultat av de komplekse teknikkene som brukes for å realisere KI-systemer. Disse teknikkene varierer stort i kompleksitet og det er dermed også et stort spenn i anbefalingene for hvordan forklarbarheten i ulike modeller bør håndteres (Arrieta et al., 2020). På generell basis viser litteraturen til XAI-teknikker som tekstlige/visuelle forklaringer, forklaringer av lokale funksjoner, forenklinger, m.m. Eksempelvis vil en tekstlig forklaring kunne brukes i en stor språkmodell ved å la modellen generere ekstra tekst som beskriver hvordan modellen behandlet inputen for å gi et svar. Slike løsninger basert på forklaring kan dog være ressurskrevende (Arrieta et al., 2020). Litteraturen viser også til mer spesifikke XAI-verktøy som i dag er populære for å øke forklarbarheten i maskinlæringsmodeller, da eksempelvis SHAP og LIME. Disse kan dog problematiseres da de ikke nødvendigvis gir gode resultater på tvers av ulike inputs og kan manipuleres til å skjule avgjørelser i systemet, noe som vil påvirke påliteligheten (Panigutti et al., 2023).

Med den kommende innføringen av EU sin «AI act» (*EU AI Act*, 2023) vil også lovverk og reguleringer kunne være en utfordring som må tas stilling til under utviklingen av KI-systemer, og som har tilknytning til forklarbarhet. Denne forordningen har som hensikt å sikre transparent og sikker utvikling og bruk av KI i EU, og gjennom «AI act» vil alle KI-systemer måtte bli evaluert og bli tildelt et risiko-nivå (Panigutti et al., 2023). Disse risiko-nivåene er: uakseptabel, høy eller minimal risiko. Alle systemer som blir betraktet som høy-risiko vil måtte være transparent nok til at brukeren kan tolke operasjonene som inngår, samt at systemene må inkludere omfattende instrukser for bruk. Slike systemer vil også være pålagt å være under styring av en menneskelig operatør mens de er i bruk (Panigutti et al., 2023). I eksempelet med ChatGPT er ikke denne modellen foreløpig underlagt vurderingen høy-risiko, men dette er enda litt uklart da modeller med generelle kapabiliteter undergår egne evalueringer (*EU AI Act*, 2023; Panigutti et al., 2023). Implementasjonen av «AI act» er derfor ett eksempel på lovverk som vil kunne gi store implikasjoner for utviklingen av KI-systemer. Selv om forordningen ikke er innført er den allerede blitt kritisert, blant annet for å legge for mye vekt på brukeren og ikke gi nok intensiver til leverandører for å legge til rette for utviklingen av sikre systemer (Helberger & Diakopoulos, 2023).

2.3.4.2 Feil i svar og hallusinering

Som vi har sett på tidligere i dette kapittelet består ofte KI-modeller av ikke-deterministiske maskinlæringsalgoritmer, som igjen gir følger for konsistensen og presisjonen for slike systemer (Y. Zhang et al., 2023). En implikasjon av at store språkmodeller bedriver en statistisk gjettelek, er at de ikke nødvendigvis produserer resultater på bakgrunn av fakta eller som er rett for konteksten til inputverdiene, dette fenomenet kalles hallusinering. Hallusinering for LLMer forklares i Zhang et al. (2023) som produksjonen av ulogisk eller ikke-konsistent output i henhold til kildedataen. I artikkelen identifiserer forfatterne også tre ulike former for hallusinering i slik kontekst av en språkmodell; Hallusinering som står i konflikt med brukerinput; Hallusinering som står i konflikt med tidligere informasjon produsert av språkmodellen; Hallusinering hvor språkmodellen ikke produserer informasjon som er konsistent med etablerte sannheter.

Det finnes flere grunner for at det oppstår hallusinering i store språkmodeller. Som vi nå har sett kan hallusinering oppstå på grunnlag av at språkmodellene er ikke-deterministisk, og sekvensielt genererer en token om gangen (Y. Zhang et al., 2023). Hvis modellen velger feil token vil den ikke nødvendigvis rette opp denne feilen, og det kan slik oppstå en snøballeffekt som gir feilaktig respons til brukeren. Forskning peker også på at LLMer har en tendens til å over-evaluere egne ferdigheter, og anta at de har tilstrekkelig kunnskap til å gi et faktuelt svar, selv om dette ikke er tilfellet. En annen kilde for hallusinering kan være mangel på relevant data fra treningsprosessen, og det finnes her en sterk korrelasjon mellom mengden hallusinering og distribusjonen av treningsdata.

For å redusere feil i svarene fra hallusinering peker Zhang et al. (2023) på at det kan gjøres tiltak i de ulike treningsfasene til modellene, samt også når modellene skal trekke sine slutninger. Hovedfokuset for håndtering av hallusinering på tvers av hele livssyklusen til LLMer er dog kvalitet i treningsdata. Et forslag for å minske hallusinering i «fine-tuning»-fasen er manuell sammenstilling av data, da volumet her er relativt lite og man slik vil kunne sikre kvalitet. Vi har også tidligere nevnt RAG som et mulig verktøy for å minske feil i svarene til store språkmodeller (Gao et al., 2024). Zhang et al. (2023) diskuterer også bruken av slike verktøy, som kan hente inn ekstern data, for å redusere hallusinering. Forfatterne er positive til bruken av slike teknikker og forklarer at disse også kan bidra til bedre forklarbarhet da datastrømmen kan spores, men diskuterer også vanskelighetene ved verifiseringen av den eksterne dataen som en mulig svakhet.

Tabell 1 - Etablerte utfordringer i KI-prosjekter

Utfordringer	**Forklaring**	**Håndteringsstrategi**	**Relaterte studier**
Kravspesifisering	Etablering av krav som bidrar mot forretnings-konteksten	Problemforståelse, utgangspunkt i forretningsmetrikker, tett samarbeid mellom interessenter i utformingen, "power couples"	Nascimento et al. (2020) Nascimento et al. (2019) Wan et al. (2020) Nahar et al. (2022) Vial et al. (2023)

Konflikter mellom agil metodikk og KI-prosess	Organisering av arbeidsoppgaver, kilder til endring, måling av progresjon,	Redefinere begrepet "ferdig", stage-gate modell med go/no-go vurderinger, "power couple"	Vial et al. (2023)
Datahåndtering	Arbeidsoppgaver tilknyttet data som er sterkt eksperimentelle	Øke dataforståelsen, Involvering av domeneeksperter	Nascimento et al. (2020) Studer et al. (2021) Nahar et al. (2022)
Kvalitet	Samsvar med etiske krav, Lav grad av hallusinering, med mer	Øke forklarbarhet, Økt datakvalitet, RAG,	Nascimento et al. (2020) Arrieta et al. (2020) Panigutti et al. (2023) Zhang et al. (2023) Helberger og Diakopolos (2023) Gao et al. (2024)

3 Metode

3.1 Kvalitativ metode

I utformingen av metoden for forskningsprosjektet må man også vurdere metoden man skal anvende for innsamling av data for å adressere problemstillingen (Busch, 2013). Det er vanlig å gjøre et skille mellom en kvalitativ og en kvantitativ metode. En kvalitativ tilnærming er mest hensiktsmessig når man skal etablere en dypere forståelse innenfor et spesifikt område ved å studere komplekse og uklare problemstillinger (Busch, 2013). En kvalitativ metode er utforskende og forskeren søker å skaffe innsikt i informantene sine tanker, følelser og opplevelser for å danne en holistisk forståelse av problemstillingen (Choy, 2014). Man ønsker altså å forstå hva informantene mener og tenker, og hvorfor de gjør det.

Vi har av flere grunner valgt en kvalitativ tilnærming i denne oppgaven. For det første er problemstillingen den viktigste drivfaktoren for valget vi har gjort. Problemstillingen vår er bred og utforskende, i tillegg til at oppgaven er basert på en case hos oppdragsgiveren, noe som favoriserer en kvalitativ tilnærming. En annen viktig faktor har vært at utvalget av informanter med kunnskap om fagfeltet har vært begrenset. Studiet baserer seg derfor på å skape en dybdeforståelse av de få informantene tilgjengelig sine opplevelser, tanker og meninger (Tjora, 2019).

Samtidig er det viktig å være klar over at kvalitative tilnærminger har sine begrensninger og ulemper. For det første er kvalitativ forskning preget av subjektivitet (Tjora, 2019) og et fortolkningsbasert ståsted (Busch, 2013). Dataen man samler inn er preget av informantene sin subjektivitet, noe som for øvrig også er tilfellet i kvantitative studier (Tjora, 2019). I tillegg har forskeren sin subjektivitet stor påvirkning i dataanalysen og empirigrunnlaget som dannes, samt i tolkningen av resultatene som gjennomføres til slutt. Vi har derfor vært nøye på å være bevisst over å ikke la våre subjektive meninger påvirke resultatet av dataanalysen. Allikevel er potensielle subjektive påvirkninger noe å ta med i beregningen når man ser på resultatene.

3.2 Casestudie

En casestudie går ut på å studere et fenomen med sterk tilknytting til en bestemt kontekst (Busch, 2013). Casestudier kan karakteriseres som holistiske studier, der forskeren søker å gå i dybden om et fenomen ved å skaffe detaljert innsikt om kompleksiteten av forhold og prosesser, og hvordan disse henger sammen (Oates et al., 2022). Casestudier er dermed godt egnet til å studere et fenomen i en organisasjon, og det er viktig å være klar over at for å forstå fenomenet, må man også forstå konteksten fenomenet oppstår i (Busch, 2013). Oates et al (2022) anser casestudier som særlig passende for å undersøke utvikling, implementering og bruk av informasjonssystemer. En casestudie ble også det vi til slutt anså som mest passende for vår problemstilling.

Etter samtaler med oppdragsgiveren og leting i faglig litteratur opplevde vi at forskning på utvikling av kunstig intelligens var et ungt og begrenset fagfelt med mye tilknyttet usikkerhet. Oppdragsgiveren hadde ansatte som jobbet med et prosjekt med utvikling av

generativ kunstig intelligens. Vi anså det derfor som svært interessant å ta utgangspunkt i å gå i dybden på de involverte sine erfaringer, ettersom vi opplevde eksisterende forskningslitteratur om dette som begrenset og hovedsakelig generell.

I forbindelse med casestudier er det noen utfordringer og begrensninger det er viktig å være klar over. Selv om casestudier studerer et fenomen og dens omgivelser, er det ofte ønskelig å generalisere funnene for å bidra mot det aktuelle forskningsfeltet (Denzin & Lincoln, 2011). Allikevel er det som nevnt viktig å være innforstått med at resultatene fra en casestudie er sterkt preget av fenomenets kontekst, og overførbarheten av forskningen kan dermed være begrenset i situasjoner utenfor denne konteksten. Anvendelsen av casestudier har fått kritikk for å ha dårlig kredibilitet ved generalisering og overførbarhet (Oates, 2022). Allikevel er det ofte faktorer ved en case som er typiske i andre lignende fenomener også. Generalisering kan derfor oppnås gjennom at casen man undersøker deler likheter med andre caser (Oates, 2022).

3.2.1 Vårt casestudie

For vår oppgave har vi valgt konsulentselskapet Kantega som case. Kantega er et IT-konsulentselskap med over 200 ansatte fordelt på kontorer i Oslo, Bergen og Trondheim (Kantega, u.å.a). De tilbyr rådgivende tjenester innenfor AI, analyse og datadrevet innovasjon, brukeropplevelse og tjenestedesign, systemutvikling og arkitektur, elektronisk identifikasjon og digitale lommebøker, samt organisasjons- og forretningsutvikling (Kantega, u.å.b). I tillegg skiller selskapet seg fra andre konsulentvirksomheter gjennom deres unike eiermodell, der alle ansatte eier like mye av virksomheten (Kantega, u.å.c).

Oppgaven vår bygger på et samarbeidet med kontoret i Trondheim, og er basert på et prosjekt de ansatte der jobber med, samt delvis et tidligere pilotprosjekt. Begge prosjektene er sterkt tilknyttet store språkmodeller og anvendelsen av slike. Kantega har blitt valgt av sine kunder på grunnlag av deres tidligere erfaringer med liknende teknologi og deres kompetanse. Siden 2023 har selskapet arbeidet med 4 prosjekter relatert til store språkmodeller, og samme høst vant de anbudet på nok et slik prosjekt tilknyttet utdanningssektoren. Dette prosjektet er hovedfokuset for vår forskning. Prosjektene er for Kantega en del av en bevisst strategisk satsning for å øke kompetanse og få fotfeste i fagfeltet. Kantega sine pågående prosjekter og deres kompetanse på feltet var de største faktorene for at vi så dette som en solid case. Ettersom Kantega er et konsulentselskap og leverer tjenester til andre bedrifter, er også kommunikasjonen med kunde ekstra sentralt. Vi syntes dette er et interessant aspekt i kontekst av KI-prosjekter og tror denne casen kan tydeliggjøre fenomener i samhandlingen mellom kunde og oppdragstaker.

Kunden i caseprosjektet er en offentlig institusjon, og vil videre bli referert til som Gondor i denne oppgaven. Bakgrunnen for prosjektet var at Gondor opplevde søkefunksjonaliteten på sin egen nettside som mangelfull. Informasjon om organisasjonen og deres tilbud lå gjemt, noe som gjorde at brukerne kunne gå glipp av essensielle opplysninger. I tillegg hadde de nettopp gjennomført en oppdatering av egen nettside, for å øke synlighet av data, men IT-sjefen til Gondor opplevde fortsatt at søk var utfordrende. På bakgrunn av dette og et strategisk ønske om å tilnærme seg kunnskap og erfaringer tilknyttet slik teknologi, bestemte Gondor seg for å anskaffe en chatbot bygd på OpenAI sin ChatGPT-4 modell. En av målsetningene var å muliggjøre en annerledes og forbedret søkeopplevelse for eksterne brukere av nettsiden, med mulighet for å utvide denne funksjonaliteten til interne dokumenter senere. Caseprosjektets formål er derfor å utvikle en programvareløsning bygd på en stor språkmodell som kan svare på spørsmål om Gondor basert på informasjonen som ligger på deres nettside.

I tidsperioden denne oppgaven har blitt skrevet, er løsningen fortsatt under utvikling. Prosjektet har gjennomført utvikling av en pilot bygd på begrenset informasjon om Gondor. Fokuset i utviklingsarbeidet er nå å heve løsningens kvalitet mens det gradvis integreres mer data om Gondor.

3.3 Datainnsamling

Dataen i en kvalitativ metode kjennetegnes ofte ved at den er i form av tekst og detaljert (Tjora, 2019). Det finnes mange forskjellige måter å samle inn data på. I dette delkapittelet vil vi redegjøre for valgene vi tok for innsamlingsmetoden og hvordan vi gjennomførte innsamlingen som danner grunnlaget for vår empiri.

3.3.1 Intervjuer

Den vanligste metoden for å generere data i kvalitative studier er gjennom intervjuer (Tjora, 2019). Vi hadde et begrenset antall informanter der ikke alle hadde mulighet til å gjennomføre intervjuene fysisk. Derfor vurderte vi semistrukturerte intervjuer med én informant av gangen som den beste fremgangsmåten for å generere data i vårt forskningsprosjekt. Semistrukturerte intervjuer er egnet i situasjoner hvor formålet er å studere meninger, holdninger og erfaringer (Tjora, 2019). I semistrukturerte intervjuer bruker man åpne spørsmål som gjør det mulig for informanten å reflektere rundt deres egne erfaringer og meninger om et eller flere forhåndsbestemte temaer, med et mål om å skape en relativt åpen samtale mellom informanten og forskeren (Tjora, 2019). På denne måten fasiliterer semistrukturerte intervjuer for at dataen som genereres gjennom svarene fra informanten, har en dybde som bidrar til et holistisk syn på problemet.

Før gjennomføring av intervjuene ble det utarbeidet en intervjuguide for å styre samtalen og trekke ut relevant informasjon fra informantene. I tråd med anbefalinger Tjora presenterer, valgte vi å dele opp intervjuguiden i tre deler; oppvarmingsspørsmål, refleksjonsspørsmål og avsluttende avrundingsspørsmål (Tjora, 2019). Ifølge Myers & Newman (2007) er det viktig å etablere hvem man er før et intervju starter. Vi startet derfor hvert intervju med å introdusere og snakke kjapt om oss selv. Myers & Newman (2007) påpeker at i tillegg at det er viktig å etablere informasjon, på bakgrunn av at dataen som genereres er preget av informanten som person. Derfor inneholdt den første delen med spørsmål i intervjuguiden enkle og konkrete spørsmål om informanten for å samle inn data som kan bidra til å forstå deres perspektiver bedre. Ifølge Tjora (2019) bidrar en slik oppvarmingsdel også til å etablere en trygghet hos informanten i intervjuet. Refleksjonsdelen utgjør kjernen av intervjuet der informanten får mulighet til å gå inn i dybden på forskjellige aspekter ved intervjuets tema (Tjora, 2019). Vi forberedte åtte hovedspørsmål fordelt på to undertemaer med tilhørende mulige oppfølgingsspørsmål. I det første undertemaet ønsket vi å etablere en forståelse for konteksten til casen vi har studert. Den andre delen omhandlet risiko og utfordringer informantene har opplevd i prosjektet. I tråd med anbefalinger fra Myers & Newman (2007) var det viktig for oss å være fleksible i intervjugjennomføringen. Vi fokuserte derfor på å gi informantene mulighet til å tenke høyt og komme med digresjoner underveis for å bidra til å skape en uformell struktur på samtalen, samt for å se etter interessante elementer ved temaet vi ikke hadde planlagt for (Myers & Newman, 2007). Den avsluttende delen med avrundingsspørsmål hadde som formål å bidra til å ta fokuset vekk fra refleksjon og avslutte intervjuet på en positiv måte (Tjora, 2019).

De semistrukturerte intervjuene ble gjennomført både fysisk og digitalt, basert på informanten sitt behov. For alle intervjuene ble det kun gjort opptak av lyd med Microsoft

Teams. Hvert intervju varte i omtrent 45-60 minutter, med unntak av det første som ble gjennomført. Ettersom å gjennomføre intervjuer var en helt ny erfaring for oss ville vi bruke første intervju som en pilottest, basert på anbefalinger fra vår veileder. Det resulterte i små endringer i intervjuguiden basert på spørsmål og formuleringer som ikke resulterte i relevant informasjon. I tillegg erfarte vi viktigheten av å styre samtalen med en balanse mellom å tillate digresjoner, samtidig som man ikke lar informanten spore av og bruke lang tid på digresjoner som ikke er relevante for studiet. Intervjuguiden ble også oppdatert etter gjennomførte intervjuer basert på en vurdering av om punktet for datametning ble nådd. Metning vil si at ny empiri ikke bidro til ny informasjon eller nytt tema i dataen (Guest et al., 2006). Allikevel er det viktig å være bevisst på at man ikke kan være sikker på om en ny informant vil tilføre ny informasjon (Tjora, 2019). Av den grunn gjaldt disse endringene kun konkrete spørsmål med lav grad av refleksjon etter en grundig vurdering av hvorfor en ny informant ikke ville legge til ny informasjon (Tjora, 2019).

Tabell 2 – *deltakerdemografi* nederst i kapittelet utgjør en presentering av relevant informasjonen om våre informanter.

3.3.2 Utvalgsstrategi

I kvalitative studier er det hensiktsmessig å rekruttere informanter basert på et strategisk utvalg (Tjora, 2019). Et strategisk utvalg innebærer å velge ut informanter som av en eller annen grunn har forutsetninger til å produsere refleksjoner basert på egne erfaringer om det aktuelle temaet. I vårt tilfelle har utvalget vært begrenset ettersom oppgaven er basert på en case som omhandler et pilotprosjekt med relativt få deltakere. Teknologien tilknyttet store språkmodeller som er anvendt i casen er også en relativt ny teknologi, som har begrenset antall potensielle informanter med relevant kunnskap om oppgavens tema. Snøballmetoden ble derfor viktig for vårt utvalg av informanter.

Snøballmetoden er en utvalgsstrategi som omhandler å starte med et kontaktpunkt som kan henvise videre til relevante informanter med kunnskap og erfaring om gitt tema (Tjora, 2019). I vårt tilfelle ble de innledende samtalene med kontaktperson for oppdragsgiver brukt til å bestemme oppgavens tematikk. Etter møtet konstruerte vi en e-post som kontaktpersonen videresendte til ansatte i bedriften med erfaring innenfor det gitte temaet, som resulterte i respons fra tre informanter. To av informantene jobber med det aktuelle caseprosjektet, hvor en av dem er prosjektleder. Den tredje respondenten har tidligere erfaring fra et lignende prosjekt bedriften har gjennomført med lik teknologi. Etter intervjuet med prosjektlederen for caseprosjektet, fikk vi kontaktinformasjon til den fjerde informanten som er produkteier hos kunden som løsningen utvikles for.

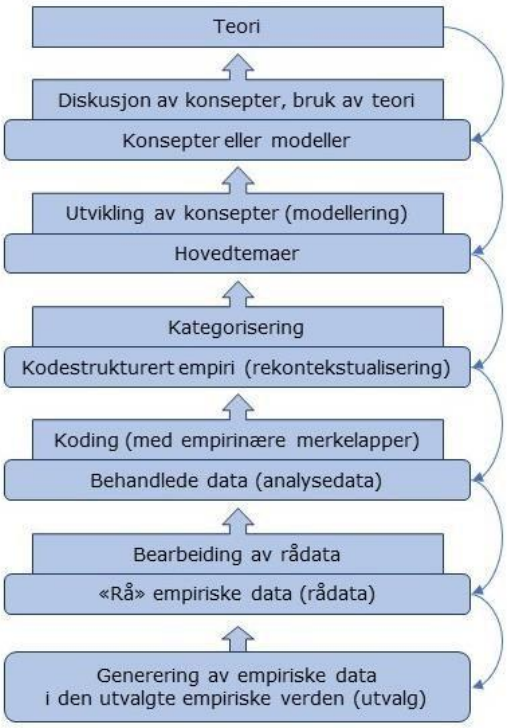
Tabell 2 2 - Deltakerdemografi

Pseudonym	**Intervjuform**	**Rolle**
Théoden	Fysisk	Konsulent. Over 20 års erfaring med utvikling. Utvikler på caseprosjektet, men ikke i direkte kontakt med kunden.
Sam	Digitalt	Sjefskonsulent. Over 15 års erfaring som utvikler og tech-lead. Prosjektleder for caseprosjektet.
Faramir	Fysisk	Skriver en master innen informatikk. Jobbet som utvikler på Kantega sitt første KI-prosjekt som anvendte store språkmodeller, sommeren 2023.

Frodo	Digitalt	Direktør for digitalisering og infrastruktur hos Gondor i over 5 år. Over 10 år med tidligere salg- og markedsføringserfaring hos stor IT-leverandør.

3.4 Dataanalyse

Etter at vi var ferdige med datainnsamlingen og hadde generert empirisk rådata i form av intervjuer, ønsket vi å prosessere og analysere denne dataen. Slikt arbeid gjør det lettere å se nyansene og de gjennomgående tematikkene i intervjuene og bidra dermed til å kunne svare på oppgavens problemstilling. Denne delen er også svært viktig for å sikre kvaliteten i kvalitative studier slik som vår, og studier som ikke gjennomgår en analyse-prosess kan fort mislykkes da de kun blir en samling med anekdoter (Tjora, 2019). Vi valgte å følge stegvis-deduktiv induktiv (SDI) metode for analyse-arbeidet. Metoden er hetet fra Tjora (2019) og presenterer en fremgangsmåte for å stegvis gå fra empirisk rådata til konsepter og teorier. Selv om metoden tilsynelatende kan virke lineær, legger forfatteren vekt på at dette ikke er tilfellet i praksis, og forklarer at man kan gå opp eller ned mellom de ulike nivåene, eller være på flere steg samtidig. Pilene nedover viser til deduktive tester, som har som mål å verifisere oppdagelsene man gjør på de ulike stegene. SDI strategien gav oss en strukturert fremgangsmåte for å arbeide med dataen vi hadde samlet, og lot oss bedre utnytte potensialet i disse.



Figur 2 2 - SDI rammeverket (Tjora, 2019, s. 4, vår oversettelse)

For å behandle rådataen gjennomførte vi manuelle transkribering som var basert på taleopptak fra Microsoft Teams. I transkriberingen passet vi på å ta vare på tenkeord og liknende når vi vasket dataen. Slikt arbeid øker lesbarheten, men ivaretar samtidig essensen av innholdet (Tjora, 2019). Når rådataen var bearbeidet gikk vi videre til å kode intervjuene. Hensikten med denne prosessen er å redusere volumet og legge til rette for

skapelsen av nye idéer på grunnlag av empirien. Vi tok utgangspunkt i empirisk lukkede koder i denne prosessen, altså koder som utelukkende kommer fra dataen og ikke fra litteraturen eller forskningsspørsmålene våre (Tjora, 2019). For å utføre selve kodingen benyttet vi oss av programvaren NVivo (versjon 1.7.1). Programmet lot oss markere ulike tekst segmenter for å gi dem koder og kategorier. For å sikre gode empirisk lukkede koder brukte vi kodetestene fra Tjora (2019, s. 32); For hver av kodene vi produserte gjorde vi en beslutning på om den (1) kunne blitt produsert i forkant av kodeprosessen, samt om den (2) representerte faktisk innhold og ikke kun tema av det som ble snakket om.

Etter at vi hadde kodet intervjuene hver for oss, gikk vi sammen og ble enige om hvilke koder som passet innholdet best. Til slutt utgjorde dette rundt 150 unike koder. Deretter startet arbeidet med å gruppere disse. Denne prosessen var tidkrevende og gikk over flere iterasjoner hvor vi gradvis senket antallet grupper for hver gjennomgang slik at vi til slutt satt igjen med fire grupperinger, slik som anbefalt fra rammeverket vi valgte (Tjora, 2019). Vi hadde også en ekstra gruppe som inneholdt koder vi ikke betraktet som relevante for det videre analysearbeidet, eksempelvis forklaringer av teknikker utviklerne brukte i arbeidet sitt. Videre tok vi med oss de endelige kodegruppene vi hadde lagt inn i konseptutviklingen. I denne fasen er ikke prosessen lenger kun drevet av empirien, men skal ses i lys av teori fra litteraturen, slik at disse sammen kan legge grunnlaget for oppgavens tematiske konseptutvikling. Fra oppgavens forberedende arbeid hadde vi prøvd å få oversikt over relevante studier og teorier som vi nå kunne bruke til å identifisere ulike fenomener fra empirien vår. Vi utforsket videre tematikkene som hadde størst relevans for de empiriske funnene våre og prøvde å se hvordan disse sammenfaller eller differensierer seg fra litteraturen. Å finne ut hva oppgaven vår faktisk omhandler og konseptualiseringen av dette har, i likhet med problemstillingen vår, vært gjennomgående for oppgaven. Det siste steget i SDI-modellen er utformingen av teorier. Denne fasen er svært lik konseptutviklingen, men teorier skal i større grad enn konsepter beskrive fenomener på en ny måte som kan testes (Tjora, 2019). Det er dog viktig å presisere at dette forskningsprosjektet ikke har oppnådd utformingen av teorier. Vi har etablert konsepter som bidrar til å oppdage og forstå fenomenene vi studerer, men ikke nådd det siste steget i Tjora sin SDI-metode med falsifiserbare teorier.

Analysearbeidet var sentralt for utformingen av forskningsspørsmålene våre, og bidro med å tydeliggjøre sammenhengen mellom teorien og empirien i dataen vi hadde samlet.

3.5 Forskningsetikk

Gjennom arbeidet med oppgaven vår har vi behandlet data tilknyttet Kantega, deres ansatte og deres kunder. For å sikre kvalitet og integritet i slike opplysninger har vi etter beste evne etterlevd forskningsetiske standarder. Fra en generell søknad som ble innsendt av studieprogramleder fikk vi godkjenning for datainnsamling gjennom blant annet intervjuer. For hver av respondentene vi kontaktet ble det sendt ut et samtykkeskjema som ga tillatelse for behandling av personopplysninger frem til prosjektslutt. Dokumentet ble signert av hver respondent. Videre gjorde vi lydopptak gjennom Microsoft Teams som ble lagret på NTNU sin Office 365 plattform. Under selve opptakene fulgte vi som nevnt også en intervjuguide hvor vi prøvde å kun registrere nødvendige personopplysninger. I etterkant av intervjuene ble lydopptakene behandlet av verktøyet AutoTekst fra UiO, vi valgte dette verktøyet da det var svært nyttig i arbeidet med transkribering og kun lagret data ved sikre servere hos UiO. Transkriberingene, lydopptakene og liknende vil bli slettet ved ferdigstillingen av oppgaven.

4 Resultater

I dette kapittelet vil vi presentere resultatene fra datainnsamlingen. Resultatene er strukturert i delkapitler som representerer konseptualiseringen av vår kvalitative data til relevante undertemaer. Målet med kapittelet er å belyse hva informantene opplevde som utfordrende i utviklingsarbeidet i caseprosjektet. Gondor benyttes som pseudonym for kunden i caseprosjektet. Tekst med innrykk er utvalgte sitater fra informantene som belyser interessante funn.

4.1 Teknologiens virkemåte

For å avklare hva informantene har opplevd som utfordrende i arbeidet med en chatbot bygd på generativ KI, ble det stilt spørsmål om hva de har opplevd som de største utfordringene. Et gjennomgående tema i alle intervjuene var at informantene opplevde det som utfordrende å forholde seg til teknologien sin ikke-deterministiske og tilfeldige virkemåte. Virkemåten resulterer i variasjon i svarene løsningen generer, og informantene trekker dette frem som utfordrende på forskjellige måter.

> «Den påliteligheten det er avgjørende. [...] Du vet at kalkulatoren din alltid gir deg svaret du vil ha, men du vet ikke om språkmodellen alltid, 100% av gangen, gir deg det svaret du vil ha.» - Faramir

Faramir presiserer at man aldri vil få en garanti om at svaret løsningen generer er 100% rett. Frodo utrykker dette på en annerledes måte og presiserer at man alltid får forskjellige svar på like spørsmål, men at svarene inneholder omtrent lik informasjon hver gang.

> «Den RAG-teknologien som brukes her, den gir jo aldri samme svar. Jeg kan stille 10 like spørsmål og få 10 forskjellige svar hver gang, men svaret er omtrentlig likt hver gang.» - Frodo

Ikke-determinismen i løsningen anses også som en utfordring for arbeidet i utviklingen av KI-løsningen. Både Théoden og Sam opplever at variasjonen i svarene som genereres gjør det vanskelig å måle kvaliteten til løsningen på en god måte. På spørsmål om forskjeller i arbeidet med løsningen i forhold til tidligere systemutviklingsprosjekter de har vært en del av, trekker Sam frem hvordan mangelen på konsistens i svarene er uvant og gjør det vanskelig å definere kvaliteten til svarene som blir generert.

> «Største forskjellen er at det er mye vanskeligere å definere kvalitet. Hva er **godt nok** og hvordan definerer vi test-/akseptansekriterier? Spør man 10 ganger så får man mer eller mindre 10 ulike svar.» - Sam

Utviklerne på prosjektet synes altså at det er utfordrende å definere hvor godt et svar som blir generert er, og dermed vanskelig å finne ut av hva som er godt nok. Allikevel har utviklerne på prosjektet funnet noen rammeverk for å teste svarene løsningen genererer. Théoden og Sam har begge fortalt at de gjennomfører tester på systemet med et rammeverk som genererer en score som adresserer hvor nøyaktig løsningen sin output er. I tillegg til at kvaliteten ved svarene som blir generert er vanskelige å måle, kommer det frem at utviklerne har oppfattet det som utfordrende å gjøre justeringer for å forbedre kvaliteten.

> «[...] Når man gjør justeringer er det også vanskelig å vite om det ble bedre eller ikke. Et **ikke-deterministisk system** er litt **uvant** for en tradisjonell systemutvikler som er vant med at dersom man gir inn "a", så får man ut "b", alltid. Sånn er det ikke her. Gir du inn "a", så kan det hende at dersom du skrur litt så får du ut "b". Kanskje gjør den det en gang, kanskje ikke neste gang. Hvem vet?» - Sam

> «[...] du har kanskje hundre brytere å skru på da. Kanskje du må skru på tre av dem på en måte, og en på en annen måte, for at det i sum skal bli bedre. Det er ikke sånn at du kan ta hver enkelt bryter og skru den til en optimal innstilling.» - Théoden

Sam og Théoden reflekterer altså over hvordan variasjonen i svarene som løsningen genererer fører til at arbeidet med å forbedre presisjonen blir mer eksperimentelt. Informantene diskuterer også hvordan dette bidrar til en opplevelse av en mer uklar arbeidsprosess. Théoden understreker at teknologiens virkemåte har ført til at prosjektet ikke følger progresjonen utviklerne er vant med fra tidligere prosjekter.

> «Vi så at det var **mindre forutsigbart** enn det vi kanskje trodde det var. I et utviklingsprosjekt så tror man gjerne at mengden funksjoner vi har implementert og kvaliteten øker over tid da, men det er ikke alltid tilfellet."- Théoden

At progresjonen i arbeidet ikke samsvarer med tidsbruken på samme måte som forventet, er noe de involverte i caseprosjektet har oppdaget underveis i prosjektet. Théoden forteller også at han ikke er den eneste med denne oppfatningen:

> «[...] Etter et planleggingsmøte i dag morgen, så kom han som vi har snakket med om dette prosjektet tidligere, og han sier at det er så annerledes å gjøre dette prosjektet enn de vanlige prosjektene. For ved et vanlig prosjekt så vet vi at det blir større og bedre. Og til slutt vet vi cirka når vi er ferdig. Nå er det mye mer **famling i mørket**.» – Théoden

Utviklerne opplever altså at det er utfordrende at progresjonen i prosjektet ikke samsvarer med innsatsen som blir lagt inn, i forhold til andre prosjekter de har gjennomført tidligere. Det kommer også frem at informantene mener at det derfor er vanskelig å vite hva som er oppnåelig og når man er i mål med utviklingen. Faramir understreker at man aldri kan garantere at svarene slike løsninger genererer blir helt korrekte og at dette bidrar til uklarhet i kvaliteten på løsningen man sitter igjen med.

> «[...] at det ikke er deterministisk, så du vet aldri om det du skal oppnå blir helt perfekt. Og selv om du får ganske gode resultater, så kan du **ikke garantere at det alltid blir perfekt**. Så kanskje den første risikoen med et prosjekt er at du begir deg på noe du ikke vet hvor bra blir til slutt, hva resultatet egentlig kan bli.» - Faramir

På bakgrunn av språkmodellen sin virkemåte, genererer altså løsningen varierende svar i form av semantikk og innhold. Denne ikke-determinismen gjør det utfordrende for de involverte i caseprosjektet å vite hvilke justeringer som vil fungere for å oppnå forbedringer i outputen og skaper derfor en opplevelse av utydelig progresjon. Basert på informantene sine svar, indikeres det at prosessen med å forbedre preisjonenen blir mer eksperimentell, og innebærer prøving, feiling og læring. Faramir som gjennomførte sitt prosjekt i sommeren 2023 uttrykker en mening om at det er viktig at en kunde legger inn rom for at denne prosessen kan foregå i slike prosjekter.

> «Hvis en kunde eller noen skal ta på seg et prosjekt, så må de legge inn tid både for **prøving og feiling**. De må legge inn tid for at det å skape pålitelighet og en løsning som er god nok, kan ta lengre tid enn man tror.» - Faramir

4.2 Etablering og kommunikasjon av problem, mål og krav

Å etablere forståelse for forretningsproblemet som skal løses, samt definering av mål og krav for prosjekter som involverer KI-teknologi, kan bidra til å sikre klarhet og retning gjennom utviklingsprosessen. Tydelig definerte mål og krav bidrar til at det arbeides mot at prosjektets verdi kan evalueres effektivt. For å avklare hvordan prosessen i forkant av utviklingsarbeidet ble gjennomført i caseprosjektet ble det stilt spørsmål om bakgrunnen og formålet for prosjektet i intervjuene.

Produkteieren forklarer at søkefunksjonaliteten på Gondor sin nettside ikke ble forbedret da de gjennomførte en oppgradering av webløsningen. Derfor bestemte de seg for å anskaffe en løsning basert på OpenAI sin store språkmodell Chatgpt-4 som ville gi en annerledes opplevelse for deres interessenter.

> «[…] Så sa jeg til IT at «supert. Da har vi en bitteliten språkmodell. Vi tar nettsiden også indekserer vi den også definerer vi det som språkmodellen og setter GPT-teknologien fra Microsoft foran og muliggjør en annen søkeopplevelse ved hjelp av det på nettsiden».» - Frodo

Problemet Gondor ønsker å adressere gjennom løsningen som utvikles av Kantega, er altså mangelfull søkefunksjonalitet på nettsiden deres. I tillegg anser produkteieren anskaffelsen av en kunstig intelligent chatbot som en måte å øke kompetansen for den disruptive teknologien innad i IT-avdelingen til Gondor.

> «IT trenger kompetanseheving på GPT. IT må trene seg på det. For vi ser at den GPT-teknologien kommer til å komme overalt i alt vi gjør, i alle siloer vi har, alle fagsystemer vi har, kommer det til å bli tilbudt GPT-funksjonalitet.» – Frodo

Frodo opplever altså at teknologi som OpenAI sin store språkmodell vil få stor innvirkning på alle forretningsprosesser i næringslivet i fremtiden og ønsker derfor å bygge kompetanse blant de ansatte innad i Gondor. Allikevel forklarer Frodo at de ikke har bestemt seg for akkurat hva løsningen skal utgjøre på nettsiden.

> «Så tanken min var at vi skulle prøve å integrere chatboten på weben vår i en eller annen fasong. Enten som en app, eller som en pop-up, eller som en del av den hosta løsningen på weben vi har. Vi har ikke bestemt oss for hvordan den skal fremstå enda, men vi så på muligheten det her kunne appellere veldig til brukerne våre.» - Frodo

Frodo har altså gjort en vurdering på at chatboten kan være appellerende for brukerne av nettsiden og bidra til å forbedre deres forretning, samt den interne kompetansen om teknologien. Det virker derimot som om Gondor ikke har identifisert noen spesifikke mål som de ønsker å forbedre ved å implementere løsningen.

Sam og Théoden opplever også dette som tilfellet og forteller at uklarheten i hvilke mål Gondor har for løsningen er utfordrende i utviklingen.

> «Utlysninga var at Gondor ville ha en GPT-4 løsning. De hadde **ikke helt spesifisert** hva de ville ha eller hva de ville oppnå, men løsningen er at vi lager en chatbot-løsning basert på GPT og RAG. Hvordan den skal leve inni økosystemet, hvilken rolle den får og hva den skal brukes til på sikt, det har dem ikke helt bestemt seg for tror jeg.» – Sam

> «[…] det er ikke sagt så mye på forhånd da, så er det ikke gitt hvilke spørsmål en chatbot skal kunne gi svar på, hva slags type dialog også videre. Det er **ikke gitt noe sånn evalueringskriterier på forhånd**, eller hva som er bra nok. Det er ting som vi på en måte må finne ut i løpet av eller underveis» - Théoden

Utviklerne opplevde altså mangel på informasjon om hva Gondor ønsker at chatboten skal svare på og hvilke evalueringskriterier den kan måles mot. På bakgrunn av dette har ansvaret for oppdelingen av prosjektet blitt lagt på utviklerne. De valgte derfor å sette i gang med utviklingsarbeidet uten tydelig definert omfang og mål, men med mer fokus på de tekniske aspektene ved løsningen. Théoden påpeker at mangelen på konkrete mål fra kunden sin side gjør utviklingsarbeidet utfordrende.

> «Så hadde jeg visst litt sånn tydeligere. Hva som var målet til Gondor. Da hadde det vært enda litt enklere da, ja. [...] jeg har ikke akkurat inntrykk av at de har visst at de skal spare så mye ressurser. Eller få så og så mye omdømme forbedring. Eller hanke inn så og så mange flere kunder. Det er ikke kvantifisert på den måten. **Da hadde det vært enklere**.» - Théoden

Det oppleves altså som utfordrende å jobbe med utviklingen av en løsning hvor det ikke er tydelig hva den skal oppnå for kunden. I caseprosjektet er de eneste rammene for utviklingsarbeidet at løsningen skal muliggjøre søk ved å svare på spørsmål om Gondor basert på informasjon fra deres nettside. Ansvaret for å identifisere mål og krav har i stor grad blitt lagt på utviklerne i prosjektet, noe som har blitt opplevd som svært utfordrende.

4.3 Data og hallusinering

Som nevnt utgjør data kjernen av kunstig intelligens. Da Gondor oppgraderte sin nettside gjennomførte de en opprydning i dokumentene med dataen som lå der. Denne opprydningen var en stor grunn til at produkteieren så en mulighet til å integrere en løsning basert på ChatGPT-4 på nettsiden.

> «[…] kommunikasjon rekte opp hånda og sa at «nå slipper vi løs det her (nettsiden). Vi er fortrolig med at informasjonen som ligger der nå er ren og pen. Vi hadde 60-70 tusen sider som vi har rydda i og nå er det 40 000 med ren og pen informasjon».» - Frodo

Produkteieren var altså av den oppfatning at datagrunnlaget på nettsiden var relevant og oppdatert før Gondor bestemte seg for å anskaffe GPT-løsningen. Allikevel viste det seg at dette ikke var tilfellet da Gondor begynte å teste prototypen i en tidlig fase av prosjektet. De involverte avdekket da at løsningen hallusinerte som følge av feil i datagrunnlaget, spesielt tilknyttet 2 av avdelingene hos Gondor.

> «[…] så har det vært en bråvåkning for 2 av enhetene våre. Grunnen til at de plutselig fikk voldsom stress, **fordi chatboten fant alle feilene** som de visste at lå der, men som de ikke hadde retta når de satte i gang ny web.» - Frodo

De involverte i prosjektet fikk altså erfare hvordan en løsning bygd på en stor språkmodell generer feil når datagrunnlaget er for dårlig. Allikevel var Frodo opptatt av mulighetene dette byr på for å identifisere utdatert og feil data, som ellers kan være vanskelig å oppdage. Dette var noe han uttrykte ved flere anledninger i intervjuet.

> «Plutselig var chatboten **den beste detektoren av feil** i weben. Plutselig har vi et verktøy som finner feil over en lav sko, men som du kanskje aldri hadde funnet hvis du gikk på weben og spurte og lette. Du hadde kanskje ikke visst at det var en feil engang.» - Frodo

Det er derimot ikke alle som deler Frodo sitt optimistiske syn på teknologien innad i Gondor. Sam, Théoden og Frodo nevner alle at det er interessenter internt i Gondor som er skeptiske til chatboten på bakgrunn av hallusineringen, spesielt med tanke på at det potensielt kan negativt påvirke deres omdømme.

> «[…] Det er veldig mange meninger. En del er opptatt av at det kan gå utover omdømmet deres dersom den ikke svarer 100% korrekt og de får masse negativ publisitet.» - Sam

Sam og Théoden forteller også at utviklerne har opplevd hallusinering fra chatboten som utfordrende i selve utviklingsarbeidet. I starten av prosjektet jobbet utviklerne med å utvikle en prototype med utgangspunkt i arbeidet som ble gjort i det tidligere prosjektet Faramir var en del av. De hadde ingen inngående kunnskap om dataen til Gondor og prototypen ble derfor utviklet for å teste språkmodellen og skaffe erfaring med teknologien.

> «Det som skjedde var at det tok veldig kort tid på å få opp en prototyp som vi syntes var god. Men vi visste jo ikke spesielt hva som sto på Gondor sine sider. […] **vi var ikke kjent spesifikt med dem** og visste bare at den svarer godt språklig. Så er det jo litt jobb å finne ut hva det står egentlig i Gondor sine sider om de tingene. Det å vite at det var noe som ble utelatt. Det er **veldig vanskelig for oss å vite for vi har ikke den oversikten**.» - Théoden

Det er tydelig at de opplevde det som utfordrende å vurdere output som ble generert fra denne prototypen ettersom de hadde lite innsikt i og forståelse for dataen om Gondor. Etter hvert i prosjektet har ansatte hos Gondor med kunnskap om organisasjonen blitt trukket inn i prosjektet for å bidra til å vurdere informasjonen som blir generert av løsningen. Théoden forteller at det førte til en realitetssjekk for utviklerne.

> «Når vi fikk med noen **eksperter på tilbudene til Gondor**, **så vi at kvaliteten var litt lavere**, og det var vanskeligere å øke kvaliteten enn vi trodde med de teknikkene vi hadde. Vi så at det var mindre forutsigbart enn det vi kanskje trodde det var.» - Théoden

Utviklerne hos Kantega gikk altså fra å oppleve prototypen som god, til å innse at den inneholdt betydelig feil og mangler etter at personer med mer kunnskap om Gondor ble involvert. Det var altså et tydelig behov for forbedring av kunnskapen om informasjonen for å øke kvaliteten i outputen. I etterkant av anskaffelsen av bedre kunnskap om relevant informasjon i prosjektet presiserer Frodo at løsningen har blitt bedre.

> «[…] før vi tunet litt sammen med kantega, produserte den en del feil. Men jeg opplever at det er **færre feil nå**.» - Frodo

Informantene har altså opplevd at det er enklere å oppdage og minimere hallusinering fra løsningen når involverte med kjennskap til kontekstdataen deltar i prosjektet. Allikevel er det i en praktisk sammenheng svært vanskelig å eliminere hallusinering helt.

4.4 Forklarbarhet og lovverk

Forklarbarhet var en tematikk som også ble drøftet av alle informantene i intervjuene vi gjennomførte. Slik som fremlagt i teorikapittelet handler forklarbarhet innenfor KI om å skape forståelse for hvordan systemene fungerer. Utfordringene knyttet til dette blir diskutert i forbindelse med både utviklingsarbeidet, holdninger hos kunden og hos sluttbrukerne, samt også i henhold til lovverk.

Faramir presiserer at det er viktig å gjøre det tydelig for brukerne av en løsning bygd på store språkmodeller, at de interagerer med et KI-system.

> «Det må aldri være en tvil når du går inn i en AI-samtale, hvis du vil kalle det det. Eller når du interagerer med en AI på en eller annen form, så burde det være ganske tydelig. Uansett hvilken slags AI man interagerer med.» - Faramir

Å tydeliggjøre hva slags teknologi man interagerer med er en del av forklarbarhet og oppleves altså som et viktig moment når en chatbot bygd på store språkmodeller settes i produksjon. I tillegg adresserer flere av informantene at det er varierende kunnskap om teknologien. Særlig er det enighet om at mennesker generelt ikke har god forståelse for hvordan slik teknologi i dag fungerer, og at man derfor må prøve å bevisstgjøre brukeren:

> «[…] forklarbarheten av AI er alltid vanskelig å forklare til Ola Nordmann da. […] vi må fortsette å ta høyde for at en stor del av folk ikke kommer til å vite hvordan det funker. Og

> når du snakker med dem, så kan du i hvert fall prøve å «**disclaime**» det til brukeren.» - Faramir

Frodo opplever også at «disclaimers» vil være en sentral del av løsningen for å øke forklarbarheten til systemet. Han understreker at de vil måtte inkludere et flertalls slike ansvarsfraskrivelser for å forklare hvordan løsningen fungerer og hvordan brukere skal forholde seg til den.

> «Så må det stå at du ikke skal legge igjen epostadresse, personverninfo. Den vil kun klare å svare på spørsmål som er relatert til nettsiden, så dersom du spør om noe annet vil den ikke kunne hjelpe deg.[...] **Det vil komme en haug med disclaimere**. Der har vi ikke landa enda, og folk er veldig opptatt av det.» - Frodo

«Disclaimers» anses altså som et verktøy for å kunne øke forklarbarhet ved løsningen for interessenter som ikke har særlig eksisterende kunnskap om teknologien, eller som en ansvarsfraskrivelse dersom systemet skulle bli brukt utenfor sin hensikt.

En forutsetning for forklarbarhet ved løsningen er derimot at de som arbeider med utviklingen forstår teknologien og hvorfor den genererer outputen den gjør. Allikevel kommer det frem at kompleksiteten til den store språkmodellen og dens ikke-deterministiske natur er vanskelig å forstå seg helt på, selv for utviklerne. Sam diskuterer rundt at konseptet for løsningen i seg selv ikke er så komplisert, men at språkmodellen den baseres på gjør det komplekst.

> «Joda, altså høy kompleksitet. Det er jo **egentlig superenkle ting** på ett vis. Du har et spørsmål, gjør et søk med det, finner dokumentet som matcher, gir de dokumentene til GPT og ber den lage et svar. Det er jo egentlig veldig enkelt, men i og med at det er så mye som ikke er deterministisk her, så blir det jo komplekst av det da.» - Sam

Flere av respondentene snakket om at det finnes teknikker for å motvirke mangelen av forklarbarhet for å håndtere dette, hvor man kan bruke språkmodellene til å selv gi tekstlige forklaringer av egne prosesser. Dette oppleves derimot som vanskelig:

> «Det finnes jo teknikker for å **få chatbotten til å tenke høyt**. For å forklare hva som er bakgrunnen til et svar. Vi ønsker jo å vite hvilke dokumenter chatboten baserer seg på. Vi søker frem, si 10 dokumenter, så blir svarene generert, også er det 3 dokumenter som har betydning. Å få chatboten til å si det selv er ikke så lett. Da ville vi gjort det litt mer forklarbart, hvis en kunne si det.» - Théoden

Selv om det er konsensus blant informantene om at KI-modeller kan være vanskelige å forstå seg på, opplever Faramir at allmennheten ofte ser på ny teknologi som mer komplisert enn den faktisk er, hovedsakelig på grunn av mangel på forståelse.

> «Jeg tror veldig mange gjør det veldig mye mer komplekst enn det egentlig er. Fordi at det er **ukjent og komplekst som ny teknologi**.» - Faramir

Å skape forståelse for systemets virkemåte og at chatboten gir forskjellige svar på like spørsmål er noe som oppleves utfordrende av informantene, både ovenfor brukerne og kunden. Når Sam blir spurt om det var noe de hadde tenkt over i forkant av utviklingen, svarer han:

> «Det vil jeg si at vi var klar over og visste om, men det vi ikke vet er hvordan man forholder seg til det. Og hvordan får man kunden til å forstå det?» - Sam

Sam uttrykker altså at det kan være vanskelig å få kunden til å forstå at svarene ikke blir like. Senere i intervjuet utdyper han med at det er forskjell på kompetanse om teknologien innad hos Gondor.

> «De har jo han som er IT-sjef som er veldig oppegående teknisk for eksempel. Han vet mye om hva det her dreier seg om, men det er jo mange andre tilstøtende som ikke kjenner til

> hvordan et utviklingsprosjekt forløper da, og hva som er mulig og ikke. Så er det store forskjeller på hvor engstelige folk er i forhold til for eksempel feil svar da.» – Sam

Under intervjuene med produkteieren blir det også tydelig at han selv har en del erfaring med bruk av store språkmodeller og hvordan disse fungerer. Han forteller derimot at det er mange ansatte i Gondor som ikke har like mye kunnskap, og som har uttrykt skepsis til løsningen. På spørsmål om hva han har opplevd som utfordrende med prosjektet, understreker han at det har vært vanskelig å få folk til å forstå at løsningen ikke genererer et likt fasitsvar på like spørsmål hver gang.

> «Der har det jo vært masse forvirring. Folk har lurt på hvorfor den ikke gir det samme svaret, skal den ikke det? Det står jo likt. Bare **det å få folk til å forstå** at den ikke gjør det og det er i naturen til GPT-teknologien. Det har vi brukt mye tid på. Og det tror jeg vi kommer til å bruke mye tid på fortsatt.» - Frodo

Aspektet om forståelse er noe Frodo trekker frem flere ganger i løpet av intervjuet. Han uttrykker også at forståelse for løsningen kan sees i tett sammenheng med forventningene interessentene har. Han forteller at løsningen skal testes innad i Gondor etter hvert. I den forbindelse anser han det som viktig å øke forståelsen for hvordan teknologien fungerer blant de ansatte som ikke har så mye erfaring med teknologien fra før av for å tydeliggjøre hvilke forventninger de skal ha til løsningen.

> «[...] Så vi har en jobb å gjøre knyttet til testregimet vårt og å få opp en forståelse om hvordan GPT-teknologien er. For jeg tror det er veldig mange fagmiljøer, særlig på økonomi og regnskap og HR, som er litt firkantet og forventer et svar med to streker under svaret. Så vi må nok si at «hør her, dere er nødt til å akseptere at svaret er omtrentlig likt 10 av 10 ganger, men aldri de samme setningsvalgene, aldri den samme informasjonen.» – Frodo

Samtidig forteller han at han opplever formidlingen av hvordan teknologien fungerer innad i Gondor som en krevende prosess.

> «Jeg kjenner jo at jeg til tider kan bli veldig frustrert av at man ikke… jeg innrømmer at jeg må gi det samme budskapet om og om igjen. Og det er veldig treg materie, man er veldig treglært, enkelte i hvert fall. Så den store utfordringen er nok den pedagogiske utfordringen knyttet til hvordan modellen fungerer, hva man kan forvente. Og jeg sier at man må jobbe med modellen på samme måte som man jobber med for eksempel chatGPT. Da tar jeg nesten for gitt at folk har brukt det. Da kan det plutselig være et stort spørsmålstegn på andre siden av bordet. Fordi folk ikke har brukt det.» - Frodo

Forklarbarhet og transparens har også stor tilknytning til lovverk og regulering, og når informantene ble spurt om de hadde tatt forbehold for slikt svarte alle at dette ikke var noe de hadde tatt hensyn til. Sam svare slik når han ble spurt direkte om den kommende forordningen til EU («AI act»):

> «Jeg må innrømme at det er noe vi ikke har vurdert eller tenkt på» - Sam

Selv om det kom frem at hverken utviklerne eller IT-sjefen hadde tatt hensyn til forordninger som «AI act», trekker de fleste informantene frem personvern og liknende retningslinjer som viktig. Informantene har dog stor tiltro til leverandørene for slike KI-modeller, og tenker at de vil kunne skjermes bak avtaler med disse. Faramir forteller dette om databehandler avtalen de har ved å bruke modellen igjennom Microsoft:

> «Når du bruker en modell gjennom Microsoft, så har du tilgang til akkurat samme modellene, som GPT-4. Men du er underlagt Microsoft sine databehandler da. Det betyr at på samme måte som du bruker Outlook, eller Word, eller SharePoint, som veldig mange bedrifter bruker, så er du **i utgangspunktet dekket**. Og jeg kan ikke tro noe annen enn at da er personvernet ditt dekket.»

5 Diskusjon

I dette kapittelet vil vi diskutere resultatene opp mot den teoretiske bakgrunnen fra kapittel 2 *Teori.* Kapittelet har lik oppbygning som resultatkapittelet for å øke den generelle forståelsen fra leseren sitt perspektiv. Diskusjonen vil bli brukt som utgangspunkt for å besvare de tidligere utledede forskningsspørsmålene.

5.1 Teknologiens virkemåte

Fra resultatene kommer det tydelig frem at alle informantene har konstatert med at KI-løsningen har unike egenskaper i forhold til annen programvare de er vant med fra tidligere. Samtlige informanter vektlegger at det er utfordrende å forholde seg til variasjonen i outputen løsningen genererer, til tross for lik input. Som nevnt i den teoretiske bakgrunnen om generativ kunstig intelligens bedriver modellene i KI-systemer en kompleks gjettelek av hvilken token som skal være den neste for en gitt kontekst. Den ikke-deterministiske oppførselen til løsningen kan derfor kobles til KI-systemet som løsningen baseres på.

Informantene trakk frem at språkmodellen sin ikke-deterministiske og tilfeldige natur gjorde det utfordrende å vite om gjennomførte justeringer resulterte i forbedring av outputen eller ikke. De understrekte hvordan dette fører til at arbeidet med å forbedre nøyaktigheten til løsningen blir i stor grad eksperimentelt. Som nevnt i den teoretiske bakgrunnen som legger frem utviklingsprosessen i KI-prosjekter, bærer utviklingsarbeidet ved KI-løsninger preg av å være sterkt eksperimentelt med et fokus på hypoteser som avkreftes eller bekreftes gjennom prøving og feiling (Wan et al., 2021). Vial et al. (2023) belyste at dette står i kontrast med etablerte prinsipper fra agil metodikk. Utviklerne uttrykker at de er vant til å jobbe med mer deterministisk teknologi hvor lik input fører til lik output og hvor arbeidet som må gjøres for å forbedre løsningen er tydeligere. Det kan dermed virke som om endringen i aktiviteter på bakgrunn av teknologiens virkemåte, oppleves som utfordrende å forholde seg til. Funnet fra empirien kan sees i sammenheng med det Vial et al. (2023) uttrykker som konflikter mellom prosessen i KI-prosjekter og agil metodikk i form av organiseringen av arbeidsoppgaver. Vial et al. (2023) trekker også frem at iterasjoner blir gjennomført på en annerledes måte i KI-prosjekter ettersom det er vanskelig å sette konkrete oppgaver som skal ferdigstilles i løpet av en iterasjon.

På bakgrunn av den eksperimentelle prosessen og utfordringene med å gjøre justeringer for å forbedre outputen, diskuterer også informantene at de opplever at forholdet mellom innsatsen som blir lagt inn og funksjonaliteten og kvaliteten man leverer, ikke nødvendigvis samsvarer. Dette har ført til at utviklerne i stor grad føler at prosessen omhandler «famling i mørket» i forhold til det de omtaler som vanlige prosjekter. Vi mener at en slik opplevelse kan kobles mot det Vial et al. (2023) beskriver som «forskjellige mål på progresjon» i sin sammenligning av agil metodikk og arbeidsflyten i KI-prosjekter. I agil metodikk fokuserer utviklerteamet på å levere fungerende og håndgripelige løsninger etter hver iterasjon, og progresjonen i prosjektet måles på funksjonaliteten som leveres. I KI-prosjekter er derimot arbeidsoppgavene knyttet til eksperimenter og resultatene av disse er ikke nødvendigvis verken et fungerende produkt eller økt funksjonalitet (Vial et al, 2023). Som vi har sett fra empirien kan dette oppleves som utfordrende for utviklerne ettersom progresjonen i arbeidet ikke foregår på samme måte som de er vant til.

Basert på resultatene blir det også tydelig at den eksperimentelle prosessen og vanskeligheten ved å måle progresjonen i utviklingsarbeidet oppleves som utfordrende på grunn av at man ikke vet hvor bra produktet blir til slutt, og at man aldri kan garantere at det blir perfekt. Som nevnt i den teoretiske bakgrunnen er gjennomførbarheten ved et KI-prosjekt ofte usikker gjennom hele utviklingsprosessen (Vial et al.,2023). Dette står til kontrast fra tradisjonelle programvareutviklingsprosjekter hvor usikkerheten gjerne senkes i takt med økt funksjonalitet og tidsbruk. Samtidig tolker vi informantene sine oppfatninger som et fokus på den tekniske ytelsen til løsningen. I teoridelen om utfordringene ved kravspesifisering presenterte vi hvordan tidligere forskning legger vekt på at de tekniske kravene burde sees i sammenheng med forretningsverdien de bidrar til å skape, og at økning i teknisk ytelse ikke alltid tilsvarer økning i levert forretningsverdi (Vial et al., 2023). Passi et al. (2020) presiserer også som nevnt at forretningssiden ikke er avhengig av at løsningen er perfekt, men god nok. Vi mener derfor at gjennom en god kravspesifisering som tar hensyn til forretningsmål vil utviklerne kunne redusere utfordringen med uklarheten i hvor bra løsningen kan bli, ved å skifte fokuset mot hvor bra som er godt nok. Allikevel er det viktig å presisere at det fortsatt kan være utfordrende å vite om målene for det som er godt nok er oppnåelig, og at en slik utfordring bare kan reduseres på denne måten.

Fra resultatene mener vi det er tydelig at teknologiens virkemåte oppleves som en utfordring på bakgrunn av dens sterke påvirkning på arbeidsprosessene i utviklingsarbeidet. De trekker frem hvordan arbeidsprosessen fører til endringer i arbeidsoppgavene, følelsen av progresjon, klarheten i hva som er oppnåelig og hvor lang tid det vil ta å oppnå ønsket kvalitet. Vi mener det er tydelig at det de anser som utfordrende er forskjellene de opplever i utviklingsarbeidet sammenlignet med fremgangsmåten de er vant med fra tidligere prosjekter som følger mer etablert og tradisjonell metodikk. Basert på den teoretiske bakgrunnen kan disse funnene som nevnt kobles opp mot det Vial et al. (2023) omtaler som konflikter mellom prosessen i KI-prosjekter og agil metodikk. Basert på strategiene forfatterne av artikkelen legger frem som håndtering av disse konfliktene vil vi argumentere for at utviklerne i caseprosjektet burde redefinere deres egne oppfatning av hva en «ferdig» arbeidsoppgave og akseptere at de er annerledes i KI-prosjekter. Istedenfor å definere arbeidsoppgaver basert på funksjonalitet burde de baseres på hypoteser om hvilke justeringer som kan fungere (Vial et al., 2023). På bakgrunn av arbeidsoppgavene kan derfor ikke progresjonen måles på samme måte i form av funksjonalitet og håndgripelige produkter (Vial et al., 2023). Ved å prøve ut en justering har man kommet et steg nærmere forbedringer, uavhengig av om justeringen resulterte i forbedret kvalitet i svaret eller ikke. Dette er ettersom man kan bevege seg videre til en annen hypotese om hva slags justeringer som vil forbedre outputen, og burde derfor betraktes som progresjon (Vial et al., 2023).

Informantenes opplevelser av at det er vanskelig å vite hva som er oppnåelig ved utviklingsarbeidet er også en del av arbeidsprosessen i KI-prosjekter. I følge Vial et al. (2023) er gjennomførbarheten usikker gjennom hele utviklingsprosessen. Samtidig understreker forfatterne at det kan være problematisk å jobbe med et dårlig prosjekt over lengre tid på grunn av fagfeltets hurtige utvikling og endringer. Derfor anvender virksomheten de studerer en stage-gate tilnærming med forhåndsdefinerte go/no-go kriterier som prosjektet kontinuerlig vurderes opp mot. Det kan hende at anvendelsen av en slik praksis ville hjulpet med å gjøre prosessen tydeligere for informantene i form av å ha konkrete vurderinger for når et prosjekt ikke lenger er lønnsomt å jobbe med.

Samtidig mener vi at det å bygge kompetanse gjennom erfaringer med slike utviklingsprosjekter vil kunne bidra som en reduserende faktor ettersom det vil skape mer forståelse for teknologien og arbeidsprosessen, og dermed gjøre de mindre uvant for de involverte i prosjektet. En slik begrunnelse er også mye av bakgrunnen for at produkteieren hos Gondor ønsket å gjennomføre prosjektet og anskaffe en KI-løsning.

Det ble også trukket frem en oppfatning om at en kunde må legge til rette for at KI-prosjekter kan foregå på den eksperimentelle og sterkt iterative måten de kjennetegnes ved, og at de er nødt til å forstå at denne utviklingsprosessen kan ta lengre tid enn man tror. Vi velger å trekke denne oppfatningen enda lenger med å argumentere for at det er viktig at kunden forstår hva utviklingsprosessen vil innebære for deres del utover flere aspekter enn tid. Som vi har diskutert er det mange utfordrende aspekter i utviklingsarbeidet i et KI-prosjekt som følger av den unike arbeidsprosessen på bakgrunn av teknologiens virkemåte. At en kunde har forståelse for at arbeidet i prosjektet ikke kan bygges rundt funksjonelle løsninger, at progresjonen dermed må måles på en annen måte, og at det er fare for at et prosjekt ikke kan levere det ønskede resultatet vil være viktig for en god utviklingsprosess. Som nevnt i den teoretiske bakgrunnen påpeker Wan et al. (2021) at problemforståelsesfasen benyttes av utviklere til å kommunisere kapabiliteter og begrensninger ved teknologien til interessenter for å håndtere forventninger. En viktig del av en slik forventningsavklaring er effektene dette har å si for prosessen ettersom de også vil påvirke kunden. Samtidig trakk vi frem at det kan være nyttig å inkludere prinsipper fra agile metoder i den teoretiske bakgrunnen (Vial et al., 2023). Slike metoder har et sterkt fokus på involvering av kunden underveis som kan være fordelaktig for å sikre et tett samarbeid gjennom prosessen og forståelse for progresjonen og prosjektets gjennomførbarhet underveis.

5.2 Etablering og kommunikasjon av problem, mål og krav

Caseprosjektet ble igangsatt på bakgrunn av utilstrekkelig søkefunksjonalitet på Gondor sine nettsider som var under behov for forbedring. Gondor oppfattet derfor den store språkmodellen til OpenAI som en mulighet til å skape en annen søkeopplevelse for Gondor sine brukere. I intervjuet deler han samtidig et syn om at store språkmodeller som ChatGPT kommer til å bli integrert i store deler av Gondor sin verdikjede i fremtiden, og anser prosjektet derfor som en mulighet til å øke kompetansen for slik teknologi hos den interne IT-avdelingen. Det er derimot ikke spesifisert hvordan løsningen skal integreres i bedriften eller konkrete mål for hva den skal oppnå. Det kan dermed virke som at motivasjonen for å anskaffe en slik løsning har vært preget av den voldsomme oppmerksomheten KI har fått basert på teknologiske fremskritt de siste årene (Giray, 2021; Leiter et al., 2023). Dermed har Gondor iverksatt et KI-prosjekt uten tydelige mål for hvordan løsningen skal realisere forretningsverdi.

Liten konkretisering av forretningsproblem, mål og krav fra Gondor er noe utviklerne i prosjektet (Sam og Théoden) har opplevd som utfordrende. Basert på informantenes ytringer, indikeres det at utviklerne ikke har fått noen spesifikasjoner til produktet utover at det skal være basert på ChatGPT-4 og integreres med informasjon fra Gondor sin nettside. Evalueringskriterier og krav er noe utviklerne selv har fått ansvar for å identifisere og formulere, og de diskuterer at dette gjør det utfordrende å navigere utviklingsarbeidet i prosjektet. Forskningslitteratur understreker at krav til en løsning burde etableres på bakgrunn av et forretningsproblem, og er betraktet som en del av et viktig grunnlag for å

skape klarhet og retning i utviklingsarbeidet (Bandi et al., 2023). Aktivitetene tilknyttet å identifisere et forretningsproblem kan som nevnt i den teoretiske bakgrunnen knyttes til den første fasen i KI-prosjekter kalt problemforståelsesfasen (Nascimento et al, 2019). Informantenes svar på spørsmål om formålet med caseprosjektet, indikerer derimot at det ikke har vært gjennomført en slik fase. Fraværet av aktivitetene tilknyttet problemforståelsesfasen har blitt opplevd som utfordrende for utviklingsarbeidet i prosjektet. Våre funn fra empirien indikerer derfor at problemhåndteringsfasen er viktig for å legge et grunnlag som bidrar til å sikre klarhet og retning for resten av prosjektet.

Allikevel er utviklingsprosessen i KI-prosjekter svært eksperimentell og sterkt iterativ på bakgrunn av det datadrevne utviklingsarbeidet (Wan et al., 2021). Vial et al. (2023) diskuterer at resultater fra eksperimenteringen som gjennomføres av utviklerne og endringer i teknologien kan føre til endringer i omfanget og dermed mål og krav for KI-prosjektet. Vi mener allikevel at våre funn indikerer at å identifisere et forretningsproblem i forkant av utviklingsarbeidet ville bidratt som et godt utgangspunkt ved å gi en tydeligere ramme for prosjektet fra start. Imidlertid understreker poenget til Vial et al. (2023) et behov for at slike aktiviteter gjennomføres kontinuerlig gjennom utviklingsarbeidet.

Samtidig mener vi at noe av problematikken Sam og Théoden diskuterer representerer en utfordring ved å utvikle krav og mål for den tekniske løsningen som vil være verdifulle for Gondor. Som nevnt i den teoretiske bakgrunnen er det å oversette et forretningsproblem til tekniske krav en utfordring i seg selv (Vial et al., 2023). Nascimento et al. (2019) og Vial et al. (2023) belyser at et utviklingsteam tar utgangspunkt i eksisterende kvantifiserbare mål av ytelsen til en forretningsprosess. Å ta utgangspunkt i slike mål gjør det enklere å oversette et forretningsproblem til kvantifiserbare mål for den tekniske ytelsen som nøyaktighet og presisjon. Samtidig påpeker Nahar et al. (2022) at kravspesifisering burde skje gjennom et tett samarbeid og interaksjon mellom interessentene i prosjektet, som i mange tilfeller kan minne om forhandlinger. I fraværet av et slikt samarbeid i caseprosjektet har ansvaret for mål og krav falt på utviklingsteamet. Som tidligere nevnt i den teoretiske bakgrunnen identifiserte Nahar et al. (2022) at tekniske team har en tendens til å ha et for sterkt fokus på tekniske krav, og neglisjere forretningssiden av prosjektet. Utviklerne sitt fokus på å utvikle en prototype basert på annen data i caseprosjektet kan understøtte en slik betraktning. Vial et al. (2023) påpeker at det er viktig å ha et fokus på at tekniske krav skal bidra mot å skape forretningsverdi i KI-prosjekter. En slik betraktning vil gjelde for både krav som identifiseres i forkant av et prosjekt i en problemforståelsesfase og krav som må etableres underveis i arbeidet på grunn av endringer i prosjektet.

Videre har arbeidet til Vial et al. (2023) trukket frem at en annen aktivitet som kan sikre at arbeidet som gjennomføres er rettet mot å skape forretningsverdi for kunden, er gjennom et samarbeid i form av såkalte "power couples" bestående av en forretningskonsulent og en med teknisk kompetanse innad i en konsulentvirksomhet sin prosjektgruppe. Et tett samarbeid mellom slike personer har vært en viktig suksessfaktor for konsulentvirksomheten forfatterne studerte. Anvendelsen av "power couples" er spesifikk for casen i Vial et al. (2023) sin studie, men tydeliggjør verdien av tekniske forbedringer sees i sammenheng med forretningsverdien det skaper for kunden, og at både forretningssiden og den tekniske siden burde være representert i teamene som jobber med utviklingen. Fokuset på tekniske krav som kan oppstå når utviklere jobber alene med utviklingen kan føre til arbeid som ikke gir verdi for kunden. Å trekke inn personer med mer forståelse for forretningssiden kan bidra til å motvirke dette (Vial et al. 2023; Nahar

et al., 2022). For caseprosjektet vi har studert mener vi at involveringen av slik kompetanse tidlig i prosessen kunne ha ført til at problematikken angående mangelen på klare mål og krav utformet med kunden hadde blitt adressert.

I følge Wan et al.(2021) kan også forhåndskunnskap om dataen og forretningskonteksten føre til en viss form for determinisme i kravsetting. Studer et al. (2021) trekker også frem dataforståelse som en viktig del av det forberedende arbeidet i datadrevne prosjekter og kobler det slik mot det Nascimento et al. (2019) omtaler som problemforståelsesfasen. Det kan hende at utviklerne i caseprosjektet hadde hatt mer utbytte av å prioritere å skaffe forståelse for den relevante dataen og forretningskonteksten, fremfor å utvikle en prototype uten kontekst, og at et slikt behov hadde blitt tydeligere gjennom et tettere samarbeid mellom aktører som forstår forretningskonteksten og aktørene med tekniske kompetanse, i forkant av utviklingsarbeidet.

5.3 Data og hallusinering

På bakgrunn av Frodo sine ytringer kommer det frem at en opprydning i informasjonen på Gondor sine nettsider, var en viktig motivasjonsfaktor for at han valget om å integrere en KI-løsning basert på OpenAI sin språkmodell ChatGPT på nettsiden. Ansatte internt i Gondor og Frodo antok at denne opprydningen førte til at informasjonen som lå på nettsiden var vasket. På bakgrunn av denne antakelsen ble caseprosjektet iverksatt. Anvendelsen av RAG i prosjektet brukes for å minimere hallusineringen til en språkmodell (Gao et al., 2024), som er en kjent utfordring ved teknologien, ved å supplere svarene som genereres med kontekstdata. Allikevel fikk de involverte oppleve at løsningen hallusinerte på bakgrunn av at informasjonen ikke var like god som først antatt. Informantene trekker alle frem at dette førte til en ekstra utfordring i prosjektet i form av skepsis til løsningen internt i Gondor på grunn av de negative effektene hallusinering i møte med sluttbrukere kunne ha for organisasjonen. Dette funnet tydeliggjør viktigheten av at datakvaliteten er tilstrekkelig i løsninger som baseres på KI-systemer. Som nevnt i den teoretiske bakgrunnen kan hallusinering oppstå som en følge av virkemåten til teknologien, men ved dårlig kvalitet på dataen oppstår feil oftere (Zhang et al., 2023). I tillegg mener vi det belyser viktigheten av å gjøre vurderinger i problemhåndteringsfasen basert på hva som er tilstrekkelig treffsikkerhet for løsningen gitt forretningskonteksten den skal implementeres i. Hvis ikke kan en risikere at utviklingsprosjektet ender i en løsning med utilstrekkelig treffsikkerhet. Slike vurderinger ville også kunne blitt brukt internt i Gondor for å håndtere forventningene og senke skepsisen til de ansatte.

I prosjektet opplevde utviklerne det som svært utfordrende å evaluere svarene løsningen genererte utover semantikken, ettersom de hadde begrenset forståelse for hva slags informasjon som var riktig om Gondor. Å finne ut om svarene var riktige og om viktig informasjon var utelatt ble ansett som utfordrende arbeid uten kjennskap til og forståelse for dataen. Som nevnt i del den teoretiske bakgrunnen er forståelse for dataen som anvendes i KI-løsningen svært sentralt i arbeidet som gjennomføres i prosessen (Studer et al., 2023). I caseprosjektet er det derimot ikke utviklerne som eier og har kjennskap til dataen, men interessenter hos kunden Gondor. Slike interessenter ble dratt inn i prosjektet etter hvert. Dette samarbeidet førte til en virkelighetssjekk for utviklerne tilknyttet deres oppfattelse av kvaliteten i svarene løsningen produserte. Uten involveringen av domeneeksperter for å belyse konteksten til dataen ville det ikke vært mulig for utviklerne i caseprosjektet å identifisere og rette opp i feilene. Å anskaffe bedre kunnskap om dataen ved hjelp av involveringen av domeneeksperter i prosjektteamet er altså helt sentralt for å øke kvaliteten i outputen som genereres fra en slik løsning.

Som nevnt i teorien trekkes dataforståelse og tilgangen til domeneeksperter frem som en flaskehals i KI/ML-prosjekter (Nahar et al., 2022). Vi mener at funnene våre bidrar til å

støtte opp om en slik betraktning. Basert på informantenes ytringer, indikeres det at et samarbeid mellom interessenter med kjennskap til dataen er sentralt for å utbedre ytelsen til løsningen, og kan derfor anses som en viktig faktor i utviklingsarbeidet. Allikevel presiserer tidligere forskning at tilgjengeligheten til domeneeksperter ofte er begrenset på grunn av at de har arbeidsoppgaver og ansvarsområder utover utviklingsprosjektet (Nahar et al., 2022). I et programvareutviklingsprosjekt basert på et KI-system pågår utviklingsarbeidet sterkt iterativt gjennom eksperimentering av hypoteser som avkreftes eller bekreftes (Vial et al., 2023). Studien til Nahar et al. (2020) understreker at det ikke er realistisk å tro at en domeneekspert vil være tilgjengelig hele tiden. Å ha en domeneekspert for å validere ethvert resultat av eksperimentene som gjennomføres i utviklingsprosessen må altså kunne betraktes som urealistisk. Vi mener derfor at en er avhengig av at utviklerne på prosjektet etablerer en tilstrekkelig kompetanse innenfor det relevante domenet, samt forståelse for dataen til Gondor så tidlig som mulig i prosjektet, slik at de kan ta selvstendige avgjørelser basert på eksperimentene uten at det går utover kvaliteten. I følge Studer et al. (2021) er dataforståelse en viktig del av aktivitetene som skjer i forkant av utviklingsarbeidet. Hvordan utviklerne kan anskaffe en slik forståelse er utenfor omfanget av denne oppgaven, men det er tydelig at forståelse for den relevante dataen er viktig for å sikre en god løsning.

Samtidig var Frodo opptatt av at løsningen som ble utviklet i caseprosjektet fant og tydeliggjorde feil i informasjonen som Gondor selv ikke var klar over eksisterte på nettsiden. Han var opptatt av mulighetene dette bydde på for å detektere feil som de kanskje ikke ville funnet ellers. Det innebærer altså feil som selv ikke domeneeksperter har funnet under den nevnte opprydningen. Funnet kan kobles mot Vial et al. (2023) sin studie som presiserer en opplevelse av at man ikke vet om en KI-løsning fungerer i en gitt kontekst før man har prøvd. Samtidig belyser det at datahåndtering er et utfordrende arbeid, selv for de med eierskap og forståelse for dataen. Vi mener derfor at et slikt funn understreker viktigheten av at aktører som har kunnskap om dataen, involveres kontinuerlig i utviklingsprosessen. I den teoretiske bakgrunnen trakk vi frem at konsulentvirksomheten Vial et al. (2023) studerte anvendte en utviklingsprosess som var inspirert av agil metodikk, og nytten av kundeinvolveringen slike metoder fasiliterer for. Å inkludere et slikt prinsipp i utviklingsarbeidet i KI-prosjekter kan altså legge til rette for utbedring av datakvaliteten underveis i prosjektet, på bakgrunn av at løsningen i seg selv bidrar til å tydeliggjøre feil.

5.4 Forklarbarhet og lovverk

Som vi så på i teorikapittelet er forklarbarhet og transparens fundamentale attributter for å sikre kvalitet i systemer som innehar KI-modeller (Nascimento et al., 2020). Slikt arbeid omhandler som nevnt å legge til rette for forståelse av systemene for ulike publikum (Arrieta et al., 2020). Å sikre forklarbarheten av slike systemer underbygger derfor både forretningsmessige, etiske og juridiske mål. I resultatene kommer det frem at informantene fra start av utviklingsprosessen hadde et tydelig fokus på bevisstgjøringen av brukerne til systemet, men foruten om dette var det ikke et veldig stort fokus på forklarbarhet som helhet. At brukerne vet at de interagerer med et KI-system er sentralt for forklarbarhet, men fra teorien vet vi at dette ikke er en dekkende løsning for utfordringene forklarbarhet medbringer (Arrieta et al., 2020). Forslaget som har blitt diskutert i caseprosjektet for å sikre bevisstgjørelse og øke forståelse var «disclaimers». Selv om slike ansvarsfraskrivelser trolig vil kunne tjene målet om bevisstgjørelse, er det vår mening at en slik løsning ikke bidrar stort til å gjøre systemet forklarbart. Skal KI-systemene forstås må det tilrettelegges for at brukerne kan tolke de interne prosessene som foregår. Dette er ikke tilfellet med «disclaimers», som kun setter retningslinjer for hvordan systemet bør brukes. Selv om det kan argumenteres for at en løsning med

«disclaimers» ikke er dekkende for utfordringene tilknyttet forklarbarhet, kan dette problematiseres da det i dag ikke nødvendigvis finnes gode tilgjengelige løsninger for et slikt prosjekt.

Vi har sett fra den teoretiske bakgrunnen at XAI-teknikkene som er omtalt i litteraturen har sine begrensninger. Da KI-modeller ikke er reduserbare, vil det ikke være gunstig å lage forklaringer som omfatter hele den interne prosessen for en KI-modell (Nascimento et al., 2020). Dette er heller ikke mulig i et prosjekt som caseprosjektet da de kun bygger en ferdig modell, og modeller slik som ChatGPT-4 ikke er særlig transparente. Vi så også at populære XAI-verktøy som SHAP og LIME har mulige svakheter i form av pålitelighet (Panigutti et al., 2023), og da slik pålitelighet er noe flere av informantene ser på som essensielt vil også kunne problematiseres. En annen løsning som ble diskutert av Théoden i resultatene og som vi så på teorien, var tekstlige forklaringer av svarene til modellen, hvor språkmodellen uttrykker sine egne resoneringsmønstre til brukeren (Arrieta et al., 2020). En slik løsning blir dog uttrykt som ressurskrevende i både empirien og teorien, og vil muligens ikke være hensiktsmessig i caseprosjekt med tanke på omfanget. På bakgrunn av litteraturen vi har funnet er det vårt inntrykk at litteraturen tilknyttet forklarbarhet (XAI) har splittede definisjoner og manglende metoder for god håndtering av forklarbarhet. I vårt litteratursøk har vi heller ikke funnet teori tilknyttet metoder for enklere bevisstgjøring av kunden gjennom eksempelvis bruken av «disclaimers» eller liknende, og det virker som fokuset til fagfeltet hovedsakelig omhandler tekniske løsninger. I følge Arrieta et al. (2020) er det et gap mellom forskning og industri innenfor fagfeltet om forklarbarhet. Dette gapet kan vanskeliggjøre anvendelsen av XAI-teknikkene som blir foreslått fra forskningen og derfor gjøre det krevende å sikre kvalitet i form forklarbarhet i prosjekter som caseprosjektet.

Et funn vi betrakter som særlig interessant er informantenes unnværende fokus på juridiske retningslinjer. Slik vi så i empirien var det ingen av respondentene som hadde tatt den kommende implementasjonen av «AI act» i betraktning, på tross av at forordningen vil kunne ha implikasjoner for kvaliteten til systemet de arbeidet med (Helberger & Diakopoulos, 2023). Som diskutert ovenfor kom det allikevel frem at de hadde gjort vurderinger angående bruken av «disclaimers» for personvernhensyn. Brukes slike ansvarsfraskrivelser til å gi instrukser for bruk, vil det kunne skjerme kunden hvis systemet skulle bli pålagt retningslinjer for høy-risiko-systemer etter en eventuell ny risiko-evaluering av ChatGPT. Skulle systemet bli vurdert som høy-risiko vil Gondor måtte allokere ekstra ressurser til forvaltningen av systemet ettersom det da trengs en menneskelig operatør for å overse chatboten (Panigutti et al., 2023). De potensielle virkningene av «AI act» demonstrerer effektene som slike retningslinjer vil kunne ha på KI-systemer. Vi tror derfor det vil ha stor verdi å etablere hvilke potensielle lovverk og reguleringer KI-prosjekter vil kunne måtte forholde seg til. Vi ser det videre hensiktsmessig at slike avklaringer skjer tidlig i prosjektet, eksempelvis i problemforståelsesfasen da det her vil kunne ses i sammenheng med de øvrige forretningsmessige målene til prosjektet.

5.5 Forbehold og svakheter ved vår forskning

En svakhet ved oppgaven kan være det empiriske datagrunnlaget. På bakgrunn av begrensningen i antall personer tilknyttet caseprosjektet vårt, og at vi ikke hadde kapasitet til å ta for oss flere prosjekter, fikk vi ikke intervjuet like mange som vi kunne ønsket. Alle informantene vi intervjuet kan også betraktes som teknologientusiaster fra et relativt homogent utvalgt, hvor tre av disse var utviklere fra oppdragsgiver. Det kan derfor argumenteres for at flere informanter fra ulike forretningskontekster kunne bidratt til å

belyse problemstillingen bedre. Ettersom systemet fortsatt er under utvikling mens denne oppgaven ferdigstilles har vi ikke fått undersøkt implementasjonen av dette, og vi har ikke fått data fra brukere av systemet. Slik data ville gjort oppgaven mer generaliserbar, og kunne gitt mer holistisk innsikt for utviklingen av et KI-system.

KI-landskapet og litteraturen som var tilgjengelig for oss under arbeidet med oppgaven kan også betraktes som en svakhet i forskningen vår. Som tidligere diskutert er fagfeltet vi har forsket på i stor vekst og det kommer ny litteratur tilnærmet hver dag, men det finnes fortsatt store gap i den etablerte forskningen. Dette har blant annet påvirket våre valg for litteraturen i teorigrunnlaget vårt. Eksempelvis gjorde vi en tilpasning ved å inkludere en prosessmodell som hovedsakelig var utviklet for rene maskinlæringsprosjekter, og ikke utviklingen av et system basert på en KI-modell slik vi har undersøkt. Selv om det er store likheter mellom slike utviklingsprosjekter, ser vi også at det vil kunne være nyanser i arbeidsprosessene som differensierer seg. Den lave metningen i relevant litteratur gjorde det også vanskeligere å utforme relevante beste praksiser. Det bør derfor tas forbehold og gjøres vurderinger for relevansen av studiets funn og konklusjoner.

6 Bærekraft

Programvaresystemer og teknologi er i dag en av de største bidragsyterne for sosial og økonomisk aktivitet (Becker et al., 2016). Hvordan slike systemer utvikles vil derfor ha påvirkning for det globale bærekrafts dilemmaet vi står ovenfor, og avhengig av hvordan man arbeider vil det kunne være en del av problemet eller løsningen. Bærekraftig utvikling defineres av Penzenstadler (2012, s. 5) som *å møte dagens behov uten at det går ut over fremtidige generasjoners mulighet til møte deres behov*, og tar for seg fire dimensjoner: menneskelige, sosial, økonomiske og miljø. I diskusjonen for hvordan bærekraft kan fremmes av teknologi blir ofte kun de umiddelbare effektene diskutert, mens langtids effektene ikke blir tatt i betraktning (Becker et al., 2015). I denne oppgaven ønsker vi derfor også å se på hvilke umiddelbare og systematiske bærekraftseffekter som vil kunne utarte seg på bakgrunn av mer effektive utviklingsprosesser for KI-systemer.

KI-modeller slik som de vi har sett på tidligere i denne oppgaven er avhengige av store mengder data og maskinkraft for å kunne trenes opp og trekke sine slutninger, systemer bygd på slike modeller kan derfor være svært ressurskrevende. *World Economic Forum* anslår at dersom den enorme fremveksten av KI-systemer fortsetter slik vi ser i dag, vil slike systemer bruke like mye elektrisitet som Island i 2028 (*How to Manage AI's Energy Demand — Today and in the Future*, 2024). Effektivisering av KI-utviklingsprosjekter vil kunne redusere mengden med eksperimentering og testing som er nødvendig for å oppnå de ønskede målsetningene for prosjektet. Dette vil igjen kunne føre til at færre modeller trengs å trenes opp, og man kan slik redusere energibruken av KI. Behandling av data er et annet aspekt slik effektivisering vil kunne gi implikasjoner for. Ved å skape bedre kjennskap til dataen som blir brukt i realiseringen av KI-modeller vil utviklere og domeneeksperter lettere kunne danne seg oversikt over disse. Dette gjør det videre lettere å fjerne data som danner grunnlag for bias og liknende (Enholm et al., 2022). Å redusere slike bias bidrar også til å redusere ulikheter og er fundamentalt for at alle skal kunne ta i bruk slike KI-systemer på likt grunnlag.

Utviklingen av KI gir uvurderlige muligheter for å forbedre menneskers liv på tvers av mange områder (Berente et al., 2021) og det store potensialet til KI vil kunne skape store systematiske endringer. De potensielle langtids effektene ved å fremme effektiv og strukturert utvikling av KI-modeller er derfor også svært store. Mer effektiv tilnærming vil kunne fremskynde potensialet til KI og vi vil slik kunne se forskjeller på stor skala. Et eksempel kan komme fra helsesektoren, der KI-modeller vil kunne bidra til å diagnostisere sykdommer raskere og mer nøyaktig, og dermed forbedre prognoser og redusere kostnader. Dette kan føre til økt livskvalitet og lengre levetid for mange mennesker globalt.

7 Konklusjon

I denne oppgaven har vi tatt utgangspunkt i et caseprosjekt som omhandler utviklingsarbeidet med å lage en KI-løsning basert på OpenAI sin store språkmodell ChatGPT-4. Utviklingsarbeidet gjennomføres av en konsulentvirksomhet på oppdrag fra en kunde som ønsker å integrere teknologien med informasjon fra sin nettside. Studiet har søkt å øke forståelsen for hvordan arbeidsprosessene i utviklingen av KI formes av egenskaper ved teknologien. Videre har vi ønsket å utforske hvilke sentrale utfordringer som oppstår på bakgrunn av dette samspillet og hvordan disse utfordringene kan håndteres, om det i det hele tatt er mulig. For å kunne diskutere denne problematikken, identifiserte vi et behov for å først adressere karakteristikkene ved teknologien som blir anvendt i vårt caseprosjekt, etterfulgt av hva som kjennetegner utviklingsprosessen for slike systemer. Til slutt så vi det nødvendig å utforske hvilke utfordringer som oppstår på bakgrunn av disse faktorene.

Det ble tydelig at KI-systemer har egenskaper som gjør arbeidsprosessen i utviklingsarbeidet krevende og ført til en rekke utfordringer som må adresseres i utviklingsarbeidet. For å drøfte disse fenomenene nærmere har vi tatt utgangspunkt i to forskningsspørsmål: FS1: *Hvilke utfordringer oppleves i utviklingsarbeidet av et system bygd på store språkmodeller?* og FS2: *Hvilke aktiviteter kan bidra til å håndtere disse utfordringene?*

> **Konklusjon av FS1:** Hvilke utfordringer oppleves i utviklingsarbeidet av et system bygd på store språkmodeller?

Gjennom studien har vi identifisert sentrale utfordringer ved utviklingsarbeidet tilknyttet (1) teknisk virkemåte (2) etablering av problem, mål og krav, (3) data og hallusinering og (4) forklarbarhet og transparens.

At KI-løsningen genererer ulik output for lik input gjør det utfordrende for utviklerne på prosjektet å vite hvilke justeringer som forbedrer den tekniske ytelsen til løsningen. Dermed blir utviklingsarbeidet svært eksperimentelt og gjør arbeidsstrukturen og -oppgavene mer utydelige. Dette har ført til en opplevelse av at det er utfordrende å estimere og måle progresjonen i utviklingsarbeidet. Det gjør det også vanskelig å vite hvor bra en løsning kan bli og om den ønskede kvaliteten er oppnåelig. Alle disse aspektene står i kontrast til mer etablerte programvareutviklingsmetoder som agil metodikk. Slike opplevde utfordringer kan brytes ned til at utviklingsarbeidet som gjennomføres er meget uvant for de involverte, til tross for høy teknisk kompetanse.

Når KI-prosjekter ikke har tydelig definerte mål og krav med bakgrunn i et forretningsproblem, oppleves utviklingsarbeidet som utfordrende å navigere for de involverte. I tillegg er det utfordrende å oversette forretningsmål til tekniske krav som adresserer hva som er tilstrekkelig oppnåelse av nøyaktighet og presisjon ved en KI-løsning.

Det sterkt iterative utviklingsarbeidet kan føre til stadige endringer av prosjektets omfang, mål og krav underveis i prosjektet. I KI-prosjekter er også datahåndtering en viktig og krevende aktivitet. Vi har identifisert at utviklingsarbeidet blir utfordrende når prosjektteamet ikke har tilstrekkelig forståelse for dataen og forretningskonteksten. I tillegg er utilstrekkelig datakvalitet fra oppstarten av utviklingsarbeidet en faktor som fører til hallusinering fra teknologien og som kan senke effektiviteten av prosjektet.

Ettersom KI-systemer kan være vanskelige å tolke og forstå, diskuterte informantene muligheten for å bevisstgjøre brukeren for hvilket type system hen interagerer med og videre skape passende forventninger. Dette viste seg å være utfordrende. Vi så også fra teorien at slik bevisstgjøring kun delvis bidrar mot forklarbarhet og at det vil være vanskelig å tilrettelegge for god forklarbarhet med dagens tilgjengelige metoder. I tillegg identifiserte vi en mulig utfordring tilknyttet lovverk og reguleringer gjennom den kommende forordningen til EU som omhandler kunstig intelligens.

Alle disse utfordringene gjenspeiler egenskaper ved teknologien og måten disse former utviklingsprosessen i KI-prosjekter.

Konklusjon av FS2: Er det mulig å håndtere disse utfordringene, eventuelt hvordan?

For å adressere aspektene som oppleves utfordrende i et KI-prosjekt har vi identifisert flere aktiviteter basert på anbefalinger/håndteringsstrategier fra tidligere forskning og våre funn fra datainnsamlingen.

Det vil er sentralt at et KI-prosjekt følger en sterkt iterativ utviklingsprosess på bakgrunn av det eksperimentelle utviklingsarbeidet. En slik utviklingsprosess krever en unik oppfatning hos de involverte av hva en ferdig oppgave i et slikt utviklingsprosjekt er og hva som fører til progresjon i arbeidet. Ved å anerkjenne at arbeidsoppgavene omfatter testing av hypoteser i større grad enn utvikling av funksjonalitet og funksjonelle løsninger, vil en sikre at utviklingsarbeidet fokuserer på det sentrale for å dra prosjektet fremover. I tillegg er det sentralt å anerkjenne at en falsifisert hypotese er progresjon. Som nevnt kan slike falsifiserte hypoteser som ikke gir ønskelige resultater, i tillegg til raske endringer i teknologien og teknikker, føre til at prosjektet ikke viser seg å være gjennomførbart underveis. Det er derimot ikke noe en kan vite før en har prøvd og det er derfor viktig å innføre aktiviteter i prosjektet som sørger for at gjennomførbarheten vurderes kontinuerlig. En måte å gjøre dette på kan være gjennom en stage-gate-tilnærming hvor forhåndsdefinerte go/no-go kriterier vurderes av prosjektteamet. Å avslutte et ugunstig prosjekt kan være fordelaktig for alle involverte i et ungt fagfelt som utvikler og endrer seg raskt. Disse endringene i utviklingsprosessen er naturligvis også noe som må kommuniseres til sentrale interessenter. Det er en del av kommunikasjonen og forventningshåndteringen som burde gjennomføres i forkant av utviklingsarbeidets start, men også noe som understreker behovet for kontinuerlig kommunikasjon og samarbeid mellom prosjekteierne og prosjektteamet.

Behovet for aktiviteter tilknyttet etablering og kommunikasjon av mål og krav for løsningen er sentralt for å tydeliggjøre rammene for utviklingsarbeidet. Slike mål og krav burde være forankret i et forretningsproblem slik at KI-prosjektet bidrar til verdiskapning for en kunde. Derfor krever slike aktiviteter et tett samarbeid og god kommunikasjon mellom personer som forstår forretningskonteksten og personer som forstår teknologiens kapabiliteter og begrensninger. Et slikt samarbeid kan føre til at utviklingsarbeidet med forbedringen av den tekniske ytelsen bidrar mot forretningsmål i større grad. Samtidig kan manglende forståelse mellom den tekniske siden og forretningssiden gjøre samarbeidet og kommunikasjonen utfordrende. Det kan derfor være fordelaktig for det tekniske teamet å inkludere en forretningskonsulent som bidrar med å se sammenhengen mellom de tekniske kravene og forretningsmål. God etablering av mål og krav i forkant av utviklingsarbeidet vil ha en verdi ved å sikre en klarhet og retning for prosjektet fra start. Allikevel preges utviklingsarbeidet av å være eksperimentelt og stadige endringer i omfang kan oppstå underveis. Et tett samarbeid mellom og sterk deltakelse fra forretningssiden og den tekniske siden av prosjektet kan derfor anses som sentralt. Prinsipper fra agil metodikk som kundeinvolvering kan bidra til å oppnå slik deltakelse.

For å håndtere utfordringene tilknyttet datahåndtering og hallusinering ved løsningen ser vi et sterkt behov for aktiviteter med formål om å øke dataforståelsen i prosjektteamet. Å

involvere domeneeksperter med kunnskap om innholdet i dataen og forretningskonteksten i prosjektteamet har vist seg å være en egnet strategi for å forbedre kvaliteten av løsningen. Allikevel kan tilgjengeligheten til slike fagfolk være begrenset, og dermed også deres deltakelse i arbeidet. Å planlegge i hvilke aktiviteter av utviklingsarbeidet det er kritisk at domeneekspertene involveres er derfor et viktig arbeid som burde gjennomføres i forkant av utviklingsarbeidet. I tillegg ser vi derfor et behov for aktiviteter som bidrar til at utviklerne kan tilegne seg kunnskap om og forståelse for dataen som er relevant for løsningen.

På denne måten konkluderer denne studien på bakgrunn av caseprosjektet vi har studert, at det kan være mulig å håndtere de opplevde utfordringene i utviklingsarbeidet av et system bygd på store språkmodeller. Vi har identifisert mulige strategier for slike håndteringer. Hovedfokuset ved disse håndteringene omhandler behov for forståelse for den nye arbeidsprosessen, sentrale aktiviteter i forkant av utviklingsarbeidets begynnelse, samt aktiv deltakelse gjennom hele arbeidsprosessen fra forskjellige aktører med forståelse for forretningskonteksten og for teknologien.

7.1 Videre forskning

Som diskutert tidligere i denne oppgaven er fagfeltet for kunstig intelligens preget av stor aktivitet og stadig ny litteratur. Etter å ha utforsket deler av fagfeltet ser vi klare behov for videre forskning, særlig på tvers av disipliner. Gjennom denne oppgaven har vi studert tverrsnittet av den tekniske anvendelsen av KI og arbeidsprosessene som er nødvendige for å realisere slik teknologi. Vi føler at det her mangler litteratur. Begrensningene vi gjorde ved å hovedsakelig fokusere på store språkmodeller og utfordringer i utviklingsarbeidet plasserte oss i en nisje av fagfeltet som er svært dagsaktuell. Vi ser for oss at etterspørselen av spesialiserte systemer som anvender KI-modeller som ChatGPT vil fortsette sin enorme vekst, og at det derfor også trengs mer særskilt litteratur med utarbeidede beste praksiser. Oppgavens introduksjon belyste også beskrivelser fra andre forskere om et slikt gap i litteraturen, men på mer generell basis. Vi har gjennom vårt litteratursøk også fått oppleve fagfeltets silotenking, og vi støtter Berente et al. (2021) sin oppmuntring om at det trengs forskning med mer holistisk tilnærming til KI.

Vårt arbeid med å etablere opplevde utfordringer i KI-prosjekter har også gitt oss innsikt i problematikkene ved å håndtere slike. Vi ønsker derfor til slutt å trekke frem forklarbarhet som en sentral utfordring vi mener trenger tydeligere retningslinjer og metoder for å kunne bane vei for enklere utvikling av KI-løsninger med høy kvalitet. Forklarbarhet er som diskutert fundamental for rettferdigheten og troverdigheten til kunstig intelligente systemer. Det er vårt inntrykk er at det i dag ikke finnes XAI-metoder som har tilstrekkelig pålitelighet, og at det derfor er nødvendig å utbedre disse metodene for å sikre forklarbarhet i KI-systemer.

Referanser

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B. & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bandi, A., Adapa, P. V. S. R. & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, *15*(8), 260. https://doi.org/10.3390/fi15080260

Becker, C., Betz, S., Chitchyan, R., Duboc, L., Easterbrook, S. M., Penzenstadler, B., Seyff, N. & Venters, C. C. (2016). Requirements: The Key to Sustainability. *IEEE Software*, *33*(1), 56–65. https://doi.org/10.1109/MS.2015.158

Berente, N., Gu, B., Recker, J. & Santhanam, R. (2021). *MANAGING ARTIFICIAL INTELLIGENCE*.

Boehm, B. (2006). A view of 20th and 21st century software engineering. *Proceedings of the 28th International Conference on Software Engineering*, 12–29. https://doi.org/10.1145/1134285.1134288

Bosch, J., Olsson, H. H. & Crnkovic, I. (2018). *It Takes Three to Tango : Requirement, Outcome/data, and AI Driven Development*. 177–192. https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-64660

Bouwman et al. (2005). *Information and communication technologies in organizations*.

Busch, T. (2013). *Akademisk skriving: for bachelor- og masterstudenter*.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q. & Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Chen, C., Seff, A., Kornhauser, A. & Xiao, J. (2015). DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2722–2730. https://doi.org/10.1109/ICCV.2015.312

Chowdhary, K. R. (2020). Natural Language Processing. I K. R. Chowdhary (Red.), *Fundamentals of Artificial Intelligence* (s. 603–649). Springer India. https://doi.org/10.1007/978-81-322-3972-7_19

Choy, L. T. (2014). The Strengths and Weaknesses of Research Methodology: Comparison and Complimentary between Qualitative and Quantitative Approaches. *IOSR Journal of Humanities and Social Science*, *19*(4), 99–104. https://doi.org/10.9790/0837-194399104

Cormen, T. H. & Leiserson, C. E. (2022). *Introduction to Algorithms, fourth edition*.

Denzin, N. K. & Lincoln, Y. S. (2011). *The SAGE Handbook of Qualitative Research*. SAGE.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

Enholm, I. M., Papagiannidis, E., Mikalef, P. & Krogstie, J. (2022). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, *24*(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

EU AI Act: first regulation on artificial intelligence. (2023, 8. juni). Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M. & Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey* (arXiv:2312.10997). arXiv. http://arxiv.org/abs/2312.10997

Giray, G. (2021). A software engineering perspective on engineering machine learning systems: State of the art and challenges. *Journal of Systems and Software*, *180*, 111031. https://doi.org/10.1016/j.jss.2021.111031

Guest et al. (2006). *How Many Interviews Are Enough?* https://doi.org/10.1177/1525822X05279903

Helberger, N. & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, *12*(1). https://doi.org/10.14763/2023.1.1682

How to manage AI's energy demand — today and in the future. (2024, 25. april). World Economic Forum. https://www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future/

Kantega. (u.å.-a). *Kantegas verden*. https://www.kantega.no/dette-er-oss

Kantega. (u.å.-b). *Menneskene og teknologien du trenger for å lage de aller beste løsningene*. https://www.kantega.no/hva-kan-vi-hjelpe-med

Kantega. (u.å.-c). *Velkommen på jobb hos Kantega. 100% eid av oss ansatte*.

Khalil, M. A. & Kotaiah, B. (2017). Implementation of agile methodology based on SCRUM tool. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2351–2357. https://doi.org/10.1109/ICECDS.2017.8389872

Leiter, C., Zhang, R., Chen, Y., Belouadi, J., Larionov, D., Fresen, V. & Eger, S. (2023). *ChatGPT: A Meta-Analysis after 2.5 Months* (arXiv:2302.13795). arXiv. https://doi.org/10.48550/arXiv.2302.13795

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Marks, G. (u.å.). *Klarna's New AI Tool Does The Work Of 700 Customer Service Reps*. Forbes. Hentet 21. mai 2024, fra https://www.forbes.com/sites/quickerbettertech/2024/03/13/klarnas-new-ai-tool-does-the-work-of-700-customer-service-reps/

Myers, M. D. & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nadkarni, P. M., Ohno-Machado, L. & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Nahar, N., Zhou, S., Lewis, G. & Kästner, C. (2022). Collaboration challenges in building ML-enabled systems: communication, documentation, engineering, and process. *Proceedings of the 44th International Conference on Software Engineering*, 413–425. https://doi.org/10.1145/3510003.3510209

Namvar, M., Intezari, A., Akhlaghpour, S. & Brienza, J. P. (2023). Beyond effective use: Integrating wise reasoning in machine learning development. *International Journal of Information Management*, *69*, 102566. https://doi.org/10.1016/j.ijinfomgt.2022.102566

Nascimento, E., Ahmed, I., Oliveira, E., Palheta, M. P., Steinmacher, I. & Conte, T. (2019). Understanding Development Process of Machine Learning Systems: Challenges and Solutions. *2019 ACM/IEEE*

International Symposium on Empirical Software Engineering and Measurement (ESEM), 1–6. https://doi.org/10.1109/ESEM.2019.8870157

Nascimento, E., Nguyen-Duc, A., Sundbø, I. & Conte, T. (2020). *Software engineering for artificial intelligence and machine learning software: A systematic literature review* (arXiv:2011.03751). arXiv. https://doi.org/10.48550/arXiv.2011.03751

Oates, B. J., Griffiths, M. & Mclean, R. (2022). *Researching Information Systems and Computing*.

Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., Scalzo, S., Mazzini, G., Sanchez, I., Soler Garrido, J. & Gomez, E. (2023). The role of explainable AI in the context of the AI Act. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1139–1150. https://doi.org/10.1145/3593013.3594069

Park, S., Wang, A. Y., Kawas, B., Liao, Q. V., Piorkowski, D. & Danilevsky, M. (2021). Facilitating Knowledge Sharing from Domain Experts to Data Scientists for Building NLP Models. *26th International Conference on Intelligent User Interfaces*, 585–596. https://doi.org/10.1145/3397481.3450637

Passi, S. & Sengers, P. (2020). Making data science systems work. *Big Data & Society*, *7*(2), 2053951720939605. https://doi.org/10.1177/2053951720939605

Penzenstadler, D. B. (u.å.). *Sustainability in Software Engineering*.

Schmidt, R., Möhring, M. & Zimmermann, A. (2020). *Value Creation in Connectionist Artificial Intelligence – A Research Agenda*.

Sokolaj, U., Grundstrom, C. & Martul, A. (2023). *Addressing Uncertainty in AI Tool Development in Healthcare Through End-User Involvement*.

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S. & Müller, K.-R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Machine Learning and Knowledge Extraction*, *3*(2), 392–413. https://doi.org/10.3390/make3020020

Tjora, A. (2019). *Qualitative Research as Stepwise-Deductive Induction*.

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vial, G., Cameron, A.-F., Giannelia, T. & Jiang, J. (2023). Managing artificial intelligence projects: Key insights from an AI consulting firm. *Information Systems Journal*, *33*(3), 669–691. https://doi.org/10.1111/isj.12420

Wamba-Taguimdje, S.-L., Fosso Wamba, S., Jean Robert, K. K. & Tchatchouang, C. E. (2020). *Influence of Artificial Intelligence (AI) on Firm Performance: The Business Value of AI-based Transformation Projects*.

Wan, Z., Xia, X., Lo, D. & Murphy, G. C. (2021). How does Machine Learning Change Software Development Practices? *IEEE Transactions on Software Engineering*, *47*(9), 1857–1871. https://doi.org/10.1109/TSE.2019.2937083

Wang, H., Huang, J. & Zhang, Z. (2019). *The Impact of Deep Learning on Organizational Agility*.

Warwick, K. & Shah, H. (2016). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence*, *28*(6), 989–1007. https://doi.org/10.1080/0952813X.2015.1055826

Zhang, C. & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, *23*, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F. & Shi, S. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models* (arXiv:2309.01219). arXiv. http://arxiv.org/abs/2309.01219

Zoph, B., Vasudevan, V., Shlens, J. & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8697–8710. https://doi.org/10.1109/CVPR.2018.00907

Vedlegg

Vedegg 1: NSD

Vedlegg 2: Intervjuguide

Guide I: Utviklere/konsulenter i Kantega (1 time hver)

Husk at oppfattet risiko/trusler er forskjellig ut ifra hvilken rolle man har.

Introduksjon (formål: samle informasjon om intervjuobjektet og få signert samtykkeskjema)

1. Kort gjennomgang av samtykkeskjema
2. Kjapt presentere hva vi ser på og hva intervjuet skal handle om
3. Spørsmål fra informanten før vi setter i gang?
4. Kan du beskrive hva slags erfaring og bakgrunn du har fra tidligere? (pass på at svaret ikke blir altfor langt)
    a. Utdanning
    b. Tidligere jobb
5. Kan du beskrive din rolle i det aktuelle generativ KI-prosjektet?

Spørsmål angående AI-prosjektet (formål: skaffe generell informasjon om prosjektet som grunnlag. Etablere hvordan de jobbet med prosjektet)

1. Kan du beskrive hva oppdraget fra kunden er for KI-prosjektet, altså hva det er de ønsker at dere skal utvikle? (formål: generell info om prosjektet. Etablere hva slags behov løsningen skal dekke)
    a. Hva er det kunden ønsker at løsningen skal gjøre?
    b. Ønsker de en løsning som er tenkt å automatisere arbeidsoppgaver/-prosesser eller for å støtte dem?
    c. Kan du ta oss gjennom hva Kantega ønsker å oppnå med å gjennomføre AI-prosjektet? (skaffe erfaringer?)
2. Kan du beskrive hva slags teknologier dere har valgt for å løse oppdraget? (formål: etablere teknologivalg for å løse behovet. Om det faktisk er generativ AI, grove tekniske trekk ved løsningen)
    a. Stikkord: OpenAI og grounding? RAG?
    b. Er dette teknologier dere har erfaring med fra tidligere.
3. Kan du ta oss gjennom hvordan utviklingsprosessen i AI-prosjektet gjennomføres / ble gjennomført på et overordnet nivå?
    a. Hva slags rammeverk følger dere for prosessen, eventuelle tilpasninger ved rammeverket? (SCRUM eller lignende)
    b. Teamstruktur
        i. Hvilke roller består teamet av?
        ii. Hvordan tas avgjørelser om videre arbeid?
        iii. Løkmodell?

iv. Er strukturen lik som i andre utviklingsprosjekter? Hvorfor (ikke)?

c. La du merke til noen forskjeller i hvordan dere jobber på dette prosjektet i forhold til hvordan du har jobbet med tidligere utviklingsprosjekter som ikke omhandlet generativ AI?
    i. Hvorfor (ikke)?
    ii. Ble slike justeringer fremarbeidet i forkant av utviklingen eller ble de manifestert underveis?

Risiko og tiltak

Spørsmål angående risiko og tiltak (formål: etablere hvordan de planlegger og hvilke risikoaspekter de er bevisste på i forkant)

1. Hva innebærer risiko for deg i et slik prosjekt? (hva tenker du når du hører ordet risiko?)
2. Kan du rangere risikoene på dette arket fra den du mener er størst til minst for AI-prosjektet du er en del av?
    a. Er det noen utfordringer i forbindelse med prosjektet som du mener mangler på arket? Hvordan ville du i så fall rangert de(n)? Hvorfor?
    b. Er det noen som er ekstra betydelige i utviklingen av AI i forhold til mer "vanlig" programvareutviklingsprosjekter?
    c. Endret det seg underveis?
3. Kan du beskrive hvordan dere identifiserte risiko og planla håndterende tiltak i forberedelsene til prosjektet?
    a. Er det en stor del av forberedelsene til et slikt prosjekt?
    b. Ble det brukt noen formelle verktøy/rammeverk for å identifisere og håndtere risikoen?
        i. I så fall: hva tar rammeverket for seg?
    c. Hvilke utfordringer ble i så fall identifisert?
    d. Hvordan planla dere å håndtere risikoen?
    e. Hvordan sørget dere for at dere fulgte disse planene under prosjektet?
        i. Ble det lagt en plan for tiltak dersom disse oppsto?
4. Hvilke erfaringer har dere tatt med dere fra tidligere AI-prosjekter Kantega har gjennomført? (ting som gikk galt osv.)
    a. Hvordan har dere fått tak i disse erfaringene? (formål: identifisere om de har systematisk fokus på læring av andre prosjekter, eller mer uformell overføring)
    b. Preventiv forhindring

5. Kan du gi eksempel på problemer som har oppstått / oppsto underveis i utviklingsarbeidet? (formål: identifisere ting som faktisk gikk galt. Finne ut om de stemte med det som var planlagt for, og hvordan de håndterte problemet).
    a. Var dette problem du/dere var klar over at kunne inntreffe?
    b. Hvilke tiltak ble gjennomført for å håndtere problemet? Ble tiltakene dokumentert?
    c. Er det noen av problemene som utgjorde en stor trussel for prosjektets suksess?
    d. Har du gjort noen erfaringer som du vil gjøre annerledes for et senere prosjekt med generativ AI? (stor forskjell på hva ulike interessenter mener at bidrar med verdi)
    e. På hvilke måter dokumenterer dere slike erfaringer for at andre kan dra nytte av dem?

Avslutning

Spørre om det er noe mer intervjuobjektet ønsker å si før båndopptakeren slåes av? Minne om rettigheter jfr. Samtykkeskjema. Spørre om informanten vet om noen andre personer vi kan intervjue om temaet? Runde av på en positiv måte. Takke for bidraget.

Vedlegg 3: Intervjuguide produkteier

Guide 2: produkteier (1 time hver)

Husk at oppfattet risiko/trusler er forskjellig ut ifra hvilken rolle man har.

Introduksjon (formål: samle informasjon om intervjuobjektet og få signert samtykkeskjema)

6. Kort gjennomgang av samtykkeskjema
7. Kjapt presentere hva vi ser på og hva intervjuet skal handle om
8. Spørsmål fra informanten før vi setter i gang?
9. Kan du beskrive hva slags erfaring og bakgrunn du har fra tidligere? (pass på at svaret ikke blir altfor langt)
    a. Utdanning
    b. Tidligere jobb
10. Hva er din stilling hos Gondor?
11. Hva er dine arbeidsoppgaver?
    a. Hva er din rolle/involvering i det aktuelle generativ AI-prosjektet?
    b. Har du jobbet med prosjekter som involverer kunstig intelligens tidligere?

Spørsmål angående AI-prosjektet (formål: skaffe generell informasjon om prosjektet som grunnlag. Etablere hvordan de jobbet med prosjektet)

4. Hva er bakgrunnen for at Gondor vil gjennomføre dette prosjektet AI-prosjektet? (formål: generell info om prosjektet. Etablere hva slags behov løsningen skal dekke)
    a. Hva ønsker dere å oppnå?
    b. Hva ønsker dere at Kantega skal utvikle?
    c. Hvilke arbeidsoppgaver er løsningen tilknyttet?
    d. Hva ønsker dere å oppnå med løsningen?
    e. Ønsker dere en løsning som er tenkt å automatisere arbeidsoppgaver/-prosesser eller for å støtte dem?
    f. Hvilke delmål har dere satt for prosjektet?
    g. Hvilken kundegruppe er det løsningen vil være rettet mot?
    h. Er det i tråd med deres digitale strategi?
5. Kan du beskrive hvordan AI-prosjektet ble planlagt?
    a. Hvor lang tid tok det fra ide til beslutning om å gjennomføre?
    b. Hvordan fungerte prosessen med å ta det fra ide til realisering?
    c. Måtte prosjektet godkjennes av ledelsen?
        i. Hvordan fungerte i så fall denne prosessen med godkjenning?
    d. I hvilken grad ble det utviklet prosjektinitieringsdokument eller lignende?

6. Hvorfor ble prosjektet godkjent og tatt videre fra idé til gjennomførelse?
    a. Ble det vurdert å benytte annen teknologi enn generativ AI for løsningen?
    b. Hvis det er en prototype, hva skal til for at dere gjennomfører en fullskalert implementering?
7. Hvordan er innstillingen til de ansatte hos Gondor for prosjektet og en eventuell full implementering av en slik løsning?
    a. Hvem er positive?
    b. Hvem er negative?
    c. Gjennomføres det tiltak for å endre innstillingen til de som er negative?
8. Hvordan involveres Gondor i utviklingsprosessen?
    a. Hvordan er samarbeidet med Kantega strukturert?
    b. Hvor mange ansatte hos dere er involvert i prosjektet?
    c. Hva er rollene og arbeidsoppgavene til de involverte fra Gondor?
    d. Sammenlignet med andre utviklingsprosjekter dere har outsourcet, er det noen forskjeller tilknyttet deres involvering i AI-prosjektet?
        i. Hvorfor (ikke)?
    e. Hvordan opplever du at samarbeidet med Kantega går?

Risiko og tiltak

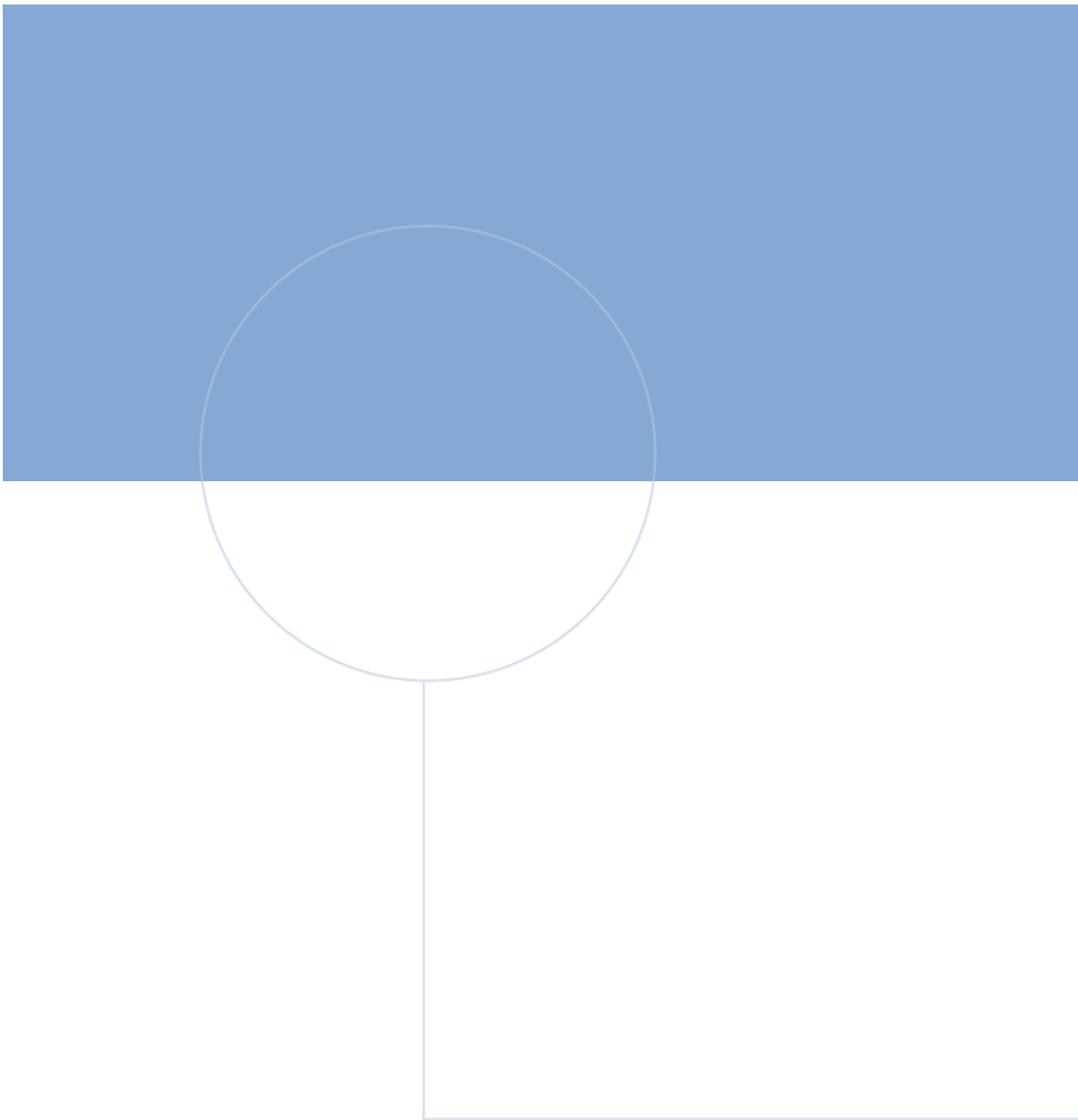
Spørsmål angående risiko og tiltak i forkant av prosjektet (formål: etablere hvordan de planlegger og hvilke risikoaspekter de er bevisste på i forkant)

6. Urelatert til prosjektet i seg selv, hva tenker du på når du hører ordet risiko?
7. Hva mener du er de største utfordringene med å implementere løsningen i full skala for Gondor?
8. Hvilke av disse risikoene mener du er viktig for dere å ta hensyn til? Hvilke mener du er mest kritisk?
    a. Personvern
    b. bias og diskriminering
    c. høy kompleksitet ved løsningen og dermed lav forståelse/forklarbarhet
    d. Mangel på overensstemmelse med nåværende og fremtidige lovverk og reguleringer
    e. høy grad av automatisering og lav grad av menneskelig kontroll.
9. Kan du beskrive hvordan dere identifiserte potensiell risiko og planla håndterende tiltak i forberedelsene av prosjektet?
    a. Hvor mye tid brukte dere på dette?

    b. Ble det brukt noen formelle verktøy/rammeverk for å identifisere og håndtere risikoen?
    c. Sammenlignet med andre IT-prosjekter du har vært en del av, hvilke tilpasninger har blitt gjort i planleggingen og gjennomføringen av AI-prosjektet med tanke på risiko?
    d. Hvilke utfordringer ble identifisert?
    e. Hvordan planla dere å håndtere risikoen?
    f. Hvordan sørger dere for å følge disse planene under prosjektet?
10. Hva slags forventninger/innstilling har sluttbrukerne/kunden til å interagere med en slik generativ AI-løsning?
    a. Hvordan har dere identifisert deres forventninger og innstilling til løsningen?
    b. Hvordan planlegger dere å sikre at løsningen oppfyller brukernes forventninger og behov?
11. Hvordan vil dere vurdere om løsningen er god nok ettersom det er snakk om et ikke-deterministisk produkt?
    a. Måles suksess kvantitativt gjennom formelle benchmarks?
    b. Vil dere kjøre brukertester?
12. Hva var dine forventninger til AI-prosjektet og gjennomførbarheten ved en slik løsning, og har det endret seg underveis?
    a. Hvorfor (ikke)?
    b. Har dere justert på suksesskriterier/delmål ved prosjektet underveis?

Avslutning

Spørre om det er noe mer intervjuobjektet ønsker å si før båndopptakeren slåes av? Minne om rettigheter jfr. Samtykkeskjema. Spørre om informanten vet om noen andre personer vi kan intervjue om temaet? Runde av på en positiv måte. Takke for bidraget.



NTNU

Kunnskap for en bedre verden