

MASTEROPPGAVE

MGMO5900

Mai 2024

«*Det er jo bare Google som snakker*» En kvalitativ studie av elevers erfaringer med språkmodeller i naturfag.

«*It's just Google talking*» A qualitative study on students' experiences with large language models in science education.

Akademisk masteroppgave

30 stp. Oppgave

21 618 ord

Emil Mikalsen

OSLOMET

OsloMet – storbyuniversitetet

Fakultet for lærerutdanning og internasjonale studier

Institutt for grunnskole- og faglærerutdanning

Abstract

The integration of artificial intelligence, specifically large language models (LLM), into educational settings, has sparked both interest and concern. This thesis explores the experiences of a sample of tenth-grade students' usage of language models in their science education, focusing on their perceptions, usage patterns, and the educational implications of this technology.

Through a qualitative approach, focus group interviews are supplemented by survey data. This study investigates both how students use language models in learning science, and how they perceive the usefulness, ease of use, credibility, and engagement of this technology. This study also evaluates students' AI literacy, which is discerned through their discussions about their interactions with, and perceptions of the technology. Interview data are analyzed through content analysis.

The findings indicate that most students perceive LLMs primarily as advanced search engines that provide highly specific and comprehensible answers. LLMs are described to offer significantly more ease of use than Google, making this technology less cognitively demanding. However, the extent of using this technology for cheating, varies among students. While a minority engages in direct cheating through copy-pasting, most students break down tasks into smaller parts, with varying degrees of their own effort. This makes the boundaries between cheating and information retrieval unclear. Notably, AI-literacy appears to mediate usage, leading to more sophisticated and critical engagement with the technology.

These findings further emphasize the importance of critical thinking and scientific literacy, while highlighting the need for fostering AI-literacy. Science education is argued to have an important role in fostering this new literacy, through scientific literacy, computational thinking, and technology education.

Sammendrag

Bruken av kunstig intelligens, spesifikt språkmodeller, i skolen, har blitt et viktig tema i den offentlige samtalen. Denne masteroppgaven undersøker et utvalg elever på 10. trinnns erfaringer med å bruke språkmodeller som verktøy i naturfag. Det legges vekt på elevenes opplevelser med og forståelse av denne teknologien noe som drøftes opp mot naturfagets rolle i utviklingen av KI-literacy.

Studien har en kvalitativ tilnærming der fokusgruppeintervjuer av elever suppleres med en spørreundersøkelse. Intervjudataene analyseres med innholdsanalyse. Studien undersøker hvordan elever anvender språkmodeller som et verktøy i naturfag, deres oppfatning av teknologiens nytteverdi, brukervennlighet og troverdighet, samt deres opplevelse av absorpsjon.

Studiens funn viser at de fleste elevene oppfatter teknologien som mer avanserte søkemotorer som gir svært presise svar som er enkle å forstå. Språkmodellene beskrives som vesentlig mer brukervennlige enn Google, noe som gjør bruken av dem mindre kognitivt krevende. Det varierer til hvilken grad språkmodellene benyttes til juksing. Mens en liten andel jukser gjennom klipp og lim, bryter majoriteten oppgavene ned i mindre deler, med ulik grad av egen innsats. Dette gjør grensene mellom juksing og informasjonssøk uklare. KI-literacy ser ut til å påvirke hvordan teknologien brukes og å henge sammen med mer avansert bruk preget av høyere egeninnsats og kritisk refleksjon rundt bruken.

Funnene understreker viktigheten av kritisk tenkning og scientific literacy, og påpeker behovet for utviklingen av KI-literacy. Det argumenteres for at naturfaget spiller en viktig rolle i utviklingen av dette, gjennom scientific literacy, algoritmisk tenking og kjerneelementet teknologi.

Innhold

Abstract	2
Sammendrag	3
Innhold	4
Forord	6
1. Innledning.....	7
1.1. Bakgrunn	7
1.2. Mål og forskningsspørsmål	8
1.3. Forklaring av noen domenespesifikke begreper innenfor KI.....	9
2. Teori	9
2.1. Kunstig intelligens.....	9
2.2. KI literacy	11
2.3. Utfordringer knyttet til KI i samfunnet og skolen.....	13
2.4. Dagens språkmodellens kunnskap om naturfag.....	15
2.5. KI i skolen	17
2.6. Rammeverk for teknologiaksept	20
3. Metode.....	23
3.1. Metodologi	24
3.2. Kontekst	24
3.3. Utvalg	26
3.4. Datainnsamling.....	30
3.5. Analyse av intervjuene	33
3.6. Analyse av spørreskjema.....	38

3.7.	Kvalitetsvurderinger	39
3.8.	Etikk	42
4.	Resultat	43
4.1.	Spørreskjema	43
4.2.	På hvilken måte har et utvalg elever på 10. trinn tatt i bruk språkmodeller som et verktøy i naturfag?	45
4.3.	Hvordan påvirker oppfatninger av språkmodellenes nytteverdi, brukervennlighet, kredibilitet og opplevelsen av absorpsjon, elevers bruk av språkmodeller?	46
4.4.	Hvilken forståelse av språkmodeller gir elevene uttrykk for i sine samtaler om bruken av denne nye teknologien?	51
5.	Diskusjon	52
5.1.	Elevenes bruk av språkmodellene	53
5.2.	Hvordan påvirker elevenes bruk av språkmodellene	55
5.3.	KI-literacy og elevenes forståelse av teknologien	58
6.	Konklusjon	62
	Referanser	63
	Vedlegg 1: Intervjuguide	68
	Vedlegg 2: Spørreskjema	70
	Vedlegg 4: Godkjenning fra Sikt	72
	Vedlegg 5: Samtykkeerklæring	73

Forord

Arbeidet med denne masteroppgaven har vært svært givende. Jeg er takknemlig for at jeg har kunnet fått undersøke noe så nytt og aktuelt som bruken av språkmodeller. Med den kunnskapen jeg har opparbeidet meg i løpet av dette prosjektet, føler jeg meg bedre rustet til å møte utfordringene moderne teknologi bringer inn i skolen.

Først og fremst vil jeg si tusen takk til min veileder Aase Marit T. Sørum Ramton. Takk for engasjementet ditt, hjelpen underveis, motivasjon og kritiske innvendinger. Jeg vil også takke Per Øyvind Sollid for konstruktive bidrag underveis i prosessen. Takk til Kristina Johnsdatter Andreasen for å gjøre dette prosjektet mulig. Ikke minst takk til naturfagsseksjonen på OsloMet for fine og lærerike år.

Sist, men ikke minst vil jeg takke min kjære samboer Katja. Jeg er utrolig takknemlig for at du har støttet og hjulpet meg underveis i prosessen

1. Innledning

Den mest omtalte formen for kunstig intelligens er språkmodeller som ChatGPT. Denne formen for kunstig intelligens har nå fått stor plass i skolen, og med denne studien vil jeg tematisere bruken av språkmodeller i naturfag. Mer spesifikt har jeg undersøkt dette fra elevenes perspektiv, da det er noe de selv har tatt i bruk og introdusert i skolen. For å kunne belyse dette, vil jeg beskrive hva KI er og utfordringer knyttet til KI i skolen og samfunnet. Deretter ønsker jeg å belyse hva vi vet om KI i skolen fra tidligere.

For å undersøke elevers bruk av språkmodeller, har jeg utviklet et rammeverk inspirert av tidligere teori om bruk og aksept av teknologi. Dette rammeverket tar hensyn til elevenes oppfatninger av teknologiens nytteverdi, brukervennlighet og ekspertise, i tillegg til opplevelsen av å bli absorbert i teknologien. Dette beskrives nærmere i kapittel 2.6. Etter å ha presentert mine resultater vil jeg drøfte mine funn og knytte dette til skolens praksis.

1.1. Bakgrunn

Kunstig intelligens (KI) er ikke et nytt fenomen, det er heller ikke språkmodeller. Men mot slutten av 2010-tallet kom fremveksten av de store språkmodellene (LLM). GPT-3 modellen til OpenAI ble lansert i 2020 under begrenset tilgang, men i november 2022 ble ChatGPT, drevet av GPT-3.5-modellen, tilgjengelig for alle. *Medietrender Ung 2023*-rapporten som ble sluppet i juni 2023, viste at 18% av unge mellom 15 og 24 brukte det til skolearbeid ukentlig (Zulic et al., 2023). Mange lærere bekymret seg for hvordan elevene brukte ChatGPT for å unngå å gjøre skolearbeidet selv (LNU, 2023), og ChatGPT ble et stort tema i den offentlige debatten. Elevenes bruk av ChatGPT gjorde det vanskelig for lærerne å gi tradisjonelle oppgaver, da disse kunne bli løst av ChatGPT på svært kort tid.

Kunnskapsdepartementet og KS (2023, s. 27-29) omtaler KI som et akutt problem, men også som en mulighet for å forbedre kvaliteten i skolen. De mener at skolen må ha kunnskap om hva KI er, når KI er i bruk og hva KI kan og ikke kan gjøre. Skolens bruk av KI må være velbegrunnet, da det byr på utfordringer med tanke på skolens dannelses- og utdanningsoppdrag. Samtidig er det en annerkjennelse av at skolen i begrenset grad kan kontrollere elevers bruk av KI. De mener også at KI kan bidra til å tilby tilpasset opplæring til elevene. Kunnskapsdepartementet og KS (2023, s.

28) etterspør mer kunnskap om KI i skolen: «Kunnskapsgrunnlaget om pedagogiske muligheter og utfordringer med KI i opplæringen er lite. Kompetanseutvikling og forskning på dette området må styrkes, slik at skolesektoren er forberedt på de omfattende endringene vi står overfor.».

En dansk ekspertgruppe satt ned av Børne- og Undervisningsministeriet (2024) legger til grunn at generativ KI i dag er en del av skolehverdagen, og mener at fokuset bør flyttes fra juks, til hvordan det kan brukes for læring. De sier også at elevene må få forstå teknologien og lære å bruke den på en reflektert måte. Medlemmene i ekspertgruppen uttaler seg utenfor sitt mandat i en kronikk i avisen Alltinget (Vedersø et al., 2024). Her skriver de: «Det bør være enhver elevs ret at modtage kvalificeret undervisning i forståelsen og brugen af generativ kunstig intelligens.».

De mener at elevene må sikres teknologiforståelse, slik at de blir mindre sårbare for manipulasjon, for å sikre elevene like muligheter og en moderne dannelse. Dette vil etter min mening være like relevant i en norsk kontekst.

1.2. Mål og forskningsspørsmål

Denne masteroppgaven tar sikte på å utvide forståelsen av kunstig intelligens (KI) i skolen, ved å utforske elevers erfaringer med språkmodeller i naturfag. Studien undersøker hvordan og hvorfor elever har tatt i bruk denne teknologien, samt deres forståelse av den. Vi vet at språkmodeller blir brukt av svært mange elever, noe som gir god grunn til bekymring. Imidlertid kan språkmodeller også by på nye muligheter for å støtte elevenes læring. Gitt at det er elevene som i stor grad har stått for språkmodellers inntog i skolen, mener jeg at det er viktig å forstå bruken av teknologien ut fra deres perspektiv. Dette vil kunne bidra med forståelse, slik at en bedre kan imøtekomme utfordringene og dra nytte av mulighetene.

Mine forskningsspørsmål er som følger:

1. På hvilken måte har et utvalg elever på 10. trinn tatt i bruk språkmodeller som et verktøy i naturfag?
2. Hvordan påvirker oppfatninger av språkmodellenes nytteverdi, brukervennlighet, kredibilitet og opplevelsen av absorpsjon, elevers bruk av språkmodeller?
3. Hvilken forståelse av språkmodeller gir elevene uttrykk for i sine samtaler om bruken av denne nye teknologien?

1.3. Forklaring av noen domenespesifikke begreper innenfor KI

Innenfor dette området er det mange domenespesifikke begreper. Jeg vil gi en kort oppklaring av noen av dem. GPT står for *generative pre-trained transformer*, og er et rammeverk for utviklingen av KI modeller. Det blir brukt for å blant annet utvikle språkmodeller. OpenAI er et selskap som utvikler KI, og stod bak den første GPT-modellen (GPT-1) (Radford et al., 2018). Når jeg i denne studien snakker om GPT-modeller refererer jeg til OpenAIs språkmodeller, ikke alle GPT-modeller. OpenAIs ChatGPT, Microsofts Copilot og Snapchats MyAI er ikke egne modeller, de er programmer drevet av GPT-modellene. Gjennom disse programmene fremstår interaksjonen som en samtale, og de blir dermed ofte omtalt som samtaleroboter. I noen deler av studien vil det være viktig å skille mellom disse, men i elevenes beskrivelser er det ofte mye overlapp og de omtales som språkmodeller.

2. Teori

Teorikapittelet starter med å presenter kunstig intelligens, med fokus på teknologien bak språkmodellene. Deretter beskrives KI-literacy og utfordringer knyttet til KI i samfunnet og skolen. Fokuset flyttes så til hvor gode dagens språkmodeller er på naturvitenskaplig kunnskap og hva vi vet om bruken av KI i utdanning.

2.1. Kunstig intelligens

Kunstig intelligens (KI) kan defineres som systemer som tar i bruk strukturerte eller ustrukturerte data for å gjennomføre et bestemt mål (Kommunal- og moderniseringsdepartementet, 2020, s. 9). Dette omfatter svært spesifikke modeller som for eksempel sjakkcomputere, til modeller som kan anvendes på et bredt utvalg studier som GPT-4. KI-systemer kan utvikles på en rekke forskjellige måter, men metoden som kalles maskinlæring (ML) er mest brukt (Kommunal- og moderniseringsdepartementet, 2020, s. 9-11), og det er denne metoden som er brukt for å utvikle GPT-modellene. ML kjennetegnes ved at modellen selv utvikler reglene basert på data, i motsetning til menneskeskapt algoritmer (Kommunal- og moderniseringsdepartementet, 2020, s. 11). Et eksempel på ML kan være å lære en datamaskin å skille mellom spam-e-post og vanlig e-post. For å gjøre dette kan en maskin trenes på e-poster som er markert som spam eller ikke, for å utvikle en modell som kan skille mellom nye e-poster. Maskinen lærer å gjenkjenne mønstre og

ord som ofte forekommer i spam-e-poster, som 'gratis', 'heldig vinner', og 'bankkonto', og utvikler en modell basert på denne læringen for å identifisere nye e-poster som enten spam eller ikke.

Dyplæring er en avansert form for maskinlæring som benytter seg av kunstige nevrale nettverk med flere lag, og er egnet til å utføre komplekse studier (LeCun et al., 2015). I stedet for å kun se på nøkkelord, kan en slik modell samtidig se på nøkkelord, grammatikk, og andre kontekstuelle forhold som tilsier at en e-post er spam.

Språkmodeller som GPT-modellene er basert på slike kunstige nevrale nettverk. Disse trenes opp ved å prosessere enorme mengder tekst. Teksten brytes ned i «tokens», en token kan være et tegn, en del av et ord eller et helt ord. Modellene trenes ved at de blir gitt en rekke med tokens fra treningsdataene (for eksempel en halv setning) og vurderer sannsynlighetene for påfølgende tokens. Om den blir gitt teksten «Roser er», vurderes sannsynligheten for at alle mulige tokens følger. Sannsynlighetene modellen gir, blir vurdert opp mot fasiten i treningsdataene og modellen blir justert slik at den i neste omgang kommer nærmere fasiten. Dette gjøres igjen og igjen med hele datasettet, og kalles Pre-training fasen. Slik utvikles det en modell for struktur og mønster i språk. Modellene finjusteres deretter med menneskelig tilbakemelding. Dette kan for eksempel foregå ved at modellen gir fem ulike svar på et spørsmål, som mennesker rangerer i rekkefølge etter kvalitet. Før modellene gjøres tilgjengelige for allmennheten, jobbes det med å begrense modellenes evner til å reprodusere fordommer, uttale seg rasistisk eller andre måter modellene kan brukes til å skade andre. Den ferdige modellen blir deretter gjort tilgjengelig via et brukergrensesnitt. Det er store forskjeller i hvordan modellene blir brukt, men jeg vil fokusere på brukergrensesnittene som er ment for å samtale med modellene, som OpenAIs ChatGPT, Microsofts Copilot eller Snapchats MyAI. Felles for disse er at de også blir gitt en slags personlighet av utviklerne, ChatGPT skal for eksempel være «hjelpsom» og «ærlig» (Fraser, 2023). Kombinasjonen av en kraftig GPT-modell, et brukergrensesnitt som ser ut som en chat og utviklernes valg av personlighet, lager en illusjon som får ChatGPT til å fremstå som en karakter (Fraser, 2023).

GPT-modellene framstår intelligente, da de kan svare korrekt på mange komplekse spørsmål. Likevel er det viktig å understreke at disse modellene kun baserer seg på språklige sammenhenger. De har ingen ekte forståelse av verden, kun av mønster og struktur i språket. Når man spør den hva som skjer når et eple løsner fra et tre, svarer den korrekt, at det faller til

bakken, men ikke fordi den har opplevelser eller en forståelse av tyngdekraften slik jeg har. Den svarer at eplet faller til bakken, fordi den er trent på tekst som beskriver dette eller liknende fenomener, og kan gjenkjenne og reprodusere relevant informasjon basert på opplæringen. Det er imidlertid mer avansert. GPT-4 kan gi fornuftige svar på spørsmål uten at modellen er trent på tekst som omhandler dette fenomenet, for eksempel: «Here we have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner.» (Bubeck et al., 2023, s. 11). Eksempler som dette gjør at enkelte mener at GPT-4 til en viss grad kan kalles en AGI (artificial general intelligence), da den ser ut til å en fersk fornuft (Bubeck et al., 2023). Hvorvidt den kan kalles en kunstig generell intelligens skal jeg ikke gå noe dypere inn på, men det viser at GPT-4 kan gi gode korrekte svar på spørsmål som ikke dukker opp i treningsdataene. Imidlertid forekommer det at GPT-modellene hallusinerer, det vil si at de finner på fakta eller kilder (Elstad, 2023). Slike hallusinasjoner kan være informasjon om dyr eller planter som ikke finnes, dødsårsaken til personer som fremdeles er i live osv. Jeg skal komme mer tilbake til hvorvidt vi kan stole på svarene fra GPT-modellene.

2.2. KI literacy

Literacy-begrepet har utvidet seg fra å omhandle evnen til å lese og uttrykke seg ved hjelp av skriftlig språk, til både et bredt generelt literacy-begrep, og som en rekke fagspesifikke literacy-begreper. Den generelle literacy-begrepet omhandler lese- og skriveferdigheter som muliggjør forståelse, kommunikasjon og aktiv deltakelse i samfunnet (Blikstad-Balas, 2023). De ferdighetene som er nødvendige for å kunne gjøre dette innenfor et gitt fagområde, omtales som fagspesifikk literacy (Long & Magerko, 2020, s. 2). Dette vil være alt fra forståelse av fagspesifikke begreper til praksiser og tenkemåter (NOU 2015: 8, 2015, s. 28).

KI literacy er kompetanse i å forstå og kritisk evaluere KI-generert innhold, samt kompetanse i å effektivt kunne samhandle med og ta i bruk KI (Long & Magerko, 2020, s. 2). For å ha disse kompetansene trenger en forståelse for hva KI er, hvordan KI virker og de etiske utfordringene knyttet til KI. Jeg vil utdype innholdet i disse slik det bli beskrevet av Long og Magerko (2020). Avslutningsvis vil jeg trekke frem noe av kunnskapen vi har om unges KI literacy.

For det første må en vite hva KI er, og kunne gjenkjenne når KI er i bruk (Long & Magerko, 2020, s. 3). Elever må altså kunne forskjellen på hva som er KI og hva som er andre former for

informasjonsteknologi. Dette forutsetter en forståelse av hvor bredt KI-begrepet er (Long & Magerko, 2020, s. 3-4). KI blir brukt i mange ulike teknologier, men hva slags KI som brukes og hvilken rolle KI har, varierer mye. Det er ikke tilstrekkelig å kun vite at en applikasjon bruker KI. Virtuelle personlige assistenter som Apples Siri bruker KI til å gjøre tale til tekst og for å tolke intensjonen bak brukerens forespørsel (Sarikaya, 2017). Om en spør Siri om hva avstanden mellom jorda og månen er, brukes KI til å tolke spørsmålet. Siri har en verktøykasse av ulike algoritmer for å respondere på forespørselen, i dette tilfellet et Google-søk. Svaret Siri gir, kommer dermed fra et Google-søk. Dette er i motsetning til generativ KI som ChatGPT, som genererer selve svaret. I tillegg til å forstå hva KI er, trenger elevene kunnskap om hva KI kan gjøre, både nå og i fremtiden (Long & Magerko, 2020, s. 4). Å kjenne til hva slags oppgaver ulike KI-systemer er egnet til å gjøre, er viktig for å effektivt ta i bruk samt for å kritisk evaluere, hvorvidt et KI-system er egnet til å utføre en bestemt oppgave.

KI literacy innebærer også en forståelse av hvordan KI virker, noe som kan støttes av både scientific literacy og algoritmisk tenking (Long & Magerko, 2020, s. 4-6). En slik forståelse krever kjennskap til hvordan informasjon blir representert for en datamaskin. For eksempel det at språkmodeller ser på relasjoner mellom ord i tekster de blir trent på. En bør også ha kjennskap til hvorfor KI responderer på en bestemt måte. Språkmodeller responderer på en tekst, ved å vurdere hvilke ord som er mest sannsynlig at kommer etter, på bakgrunn av treningsdataene. I tillegg blir kjennskap til grunnleggende konsepter som maskinlæring og nevralt nettverk trukket frem. I tillegg kan kjennskap til utvikling av KI, for eksempel hvilke treningsdata som er brukt og hvilke valg utviklerne har tatt, bidra med en forståelse for hvilke faktorer som påvirker resultatene (Long & Magerko, 2020). Kjennskap til hvordan teknologien virker, bør gi elevene kompetanse til å reflektere rundt hvorfor et KI-system ga en bestemt respons (Long & Magerko, 2020).

Kunnskap om, og evnen til å identifisere etiske utfordringer knyttet til KI er et annet aspekt ved KI literacy (Long & Magerko, 2020, s. 6-7; Ng et al., 2021). Etiske utfordringer blir løftet frem i kapittel 2.3. Her vil jeg trekke frem noen av de etiske utfordringene beskrevet av Long og Magerko (2020) som jeg anser som spesielt relevante i forbindelse med språkmodeller. Dette er kjennskap til personvernsutfordringene som følger av den massive datainnsamlingen til KI-selskaper, kjennskap til misinformasjon som følger av KI, og hvordan data- og utviklerbias

påvirker KI, samt utfordringene som følger av ugjennomsiktig KI, hvor en ikke får innsyn i hvorfor et KI responderer slik det gjør.

Barn kan være spesielt utsatt for å ha overdreven tillit til KI og å overvurdere KI-systemers intelligens (Long & Magerko, 2020, s. 8-9). De har en tendens til å personifisere KI-systemer mer enn voksne, og ofte oppfatter de disse systemene som vennlige og pålitelige. Brukergrensesnittet til KI-systemer kan forsterke denne personifiseringen, for eksempel gjennom bruk av ansiktsslignende ikoner eller instruksjoner som får systemet til å fremstå mer menneskelig. Videre har forskning vist at barn har en tendens til å vurdere et systems intelligens basert på hvor effektivt det løser en oppgave, heller enn å ta hensyn til mindre synlige faktorer som systemets underliggende strategier (Long & Magerko, 2020).

2.3. utfordringer knyttet til KI i samfunnet og skolen

Kommunal- og moderniseringsdepartementet (2020, s. 57-58) trekker frem fire problemstillinger knyttet til KI: stordata og dataminimering, datakvalitet, mangel på transparens og autonomi. Jeg vil også trekke frem overdreven tillit og avhengighet som utfordringer.

2.3.1. Stordata og dataminimering

Bruken av KI innebærer flere personvernsutfordringer, spesielt relatert til håndteringen av store mengder data og prinsippet om dataminimering. KI kan nyttiggjøre seg av svært store og komplekse datasett, noe som skaper et insentiv for å samle inn så mye data som mulig. Dette behovet for data må da sees opp mot prinsippet om å begrense mengden innsamlede data til kun det som er nødvendig og formålstjenlig. Når det gjelder bruk av KI i skolen er det viktig å nevne at skoleeier har ansvar for håndtering av persondata som samles inn, og bruk av KI-verktøy som ChatGPT krever en databehandlingsavtale. Lærere kan med andre ord ikke be elevene om å bruke verktøyene, på grunn av personvernsutfordringene, med mindre skoleeier har en databehandlingsavtale.

2.3.2. Datakvalitet

Den tilsynelatende intelligensen til språkmodellene kommer fra å prosessere store mengder tekst, og kvaliteten på disse tekstene er avgjørende for å sikre pålitelige språkmodeller. Utvalgsskjevhet

i treningsdataene vil dermed kunne bli videreført i de endelige modellene. En modell trent på data som avviser menneskeskapt klimaendring, kan dermed videreføre denne oppfatningen (Farrokhnia et al., 2023). Imidlertid settes det restriksjoner i de tilgjengelige GPT-modellene for å minske videreføring av skjevheter som eksisterer i treningsdataene, men det er ikke perfekt (OpenAI, 2023, s. 47-50). Et ytterligere problem i skolen er at dataene kan mangle innhold relevant for en norsk kontekst, for eksempel kulturelle aspekter og norsk læreplan (Farrokhnia et al., 2023).

2.3.3. Mangel på transparens

GPT-modellene er lite transparente, vi kan i liten grad forklare hvorfor modellen produserer en bestemt respons (Hassija et al., 2024). Slike KI-modeller blir ofte omtalt som sorte bokser, og et eget fagfelt kalt forklarbar KI har vokst frem for å imøtekomme denne utfordringen (Kommunal- og moderniseringsdepartementet, 2020, s. 58). Transparente modeller blir anbefalt når en skal introdusere elever til KI, slik at de kan utvikle egen forståelse og KI literacy, og kritisk evaluere svarene fra modellen (Long & Magerko, 2020).

OpenAI implementerer restriksjoner i GPT-modellene for å motvirke bias, hindre misbruk, unngå skadelig innhold og andre viktige etiske prinsipper (OpenAI, 2023). Imidlertid er detaljene rundt disse restriksjonene ikke fullstendig transparente, og de er laget av et amerikansk selskap i en amerikansk kontekst. Restriksjonene som settes bør ifølge Kunnskapsdepartementet og KS (2023, s. 28) være tilgjengelige for lærere og elever.

2.3.4. Autonomi

KI-systemer demonstrerer en grad av autonomi som overgår tradisjonelle formler og algoritmer, og dette reiser viktige spørsmål om ansvar og kontroll. En KI kan settes i en posisjon til å treffe beslutninger og starte prosesser helt selvstendig. Det er svært viktig å være bevisst utfordringene som ligger i å flytte autonomi vekk fra mennesker og over til KI.

2.3.5. Overdreven tillit

Overdreven tillit til KI er et problem som oppstår når modellene stadig presterer bedre og legitimiteten økes (OpenAI, 2023, s. 59-60). Det er grunn til å tro at legitimiteten økes ytterligere ved at flere aktører tar teknologien i bruk, som for eksempel læreren, NDLA og universitetet.

Personer med lav KI- literacy og personer som bruker KI på fagfelt de selv har lite kunnskap, er ekstra sårbare for overdreven tillit (Passi & Vorvoreanu, 2022, s. 4-5). Når KI introduseres bør brukeropplevelsen tilpasses brukerens KI-literacy og fagkunnskap for å motvirke overdreven tillit (Passi & Vorvoreanu, 2022, s. 13). Menneskeliggjøring av KI er et hinder for utvikling av KI-literacy, og kan bidra til overdreven tillit (Ding et al., 2023, s. 4). Derfor er det viktig å utvise varsomhet når man instruerer en språkmodell til å innta en viss rolle, stil eller personlighet (OpenAI, 2023, s. 59-60). Slike instruksjoner kan være å be språkmodeller om å oppføre seg som en hjelpsom medelev, en oppmuntrende lærer, eller en sokratisk samtalepartner, i tillegg har ChatGPT, Copilot og MyAI innebygde instruksjoner brukeren ikke kan kontrollere (Fraser, 2023).

2.3.6. Avhengighet

Samtidig som en kan utvikle overdreven tillit til KI, vil en også kunne bli avhengig av KI, noe som kan føre til at en mister eller ikke utvikler viktige ferdigheter (Farrokhnia et al., 2023; OpenAI, 2023, s. 59-60). Filosof Einar Duenger Bøhn (2024) problematiserer bruken av generativ KI i skolen, da det kan være et hinder for dannelse. Han bruker Google Maps som en parallell, det gjør det veldig enkelt for oss å finne den mest effektive veien, men til gjengjeld mister vi evnen til å selv navigere. På samme måte er det en fare for at elevene overlater høyere ordens tenking til KI, og dermed ikke utvikler disse ferdighetene (Farrokhnia et al., 2023). Om KI blir en ytre autoritet som det stoles blindt på i skolen vil dette stå i en direkte motsetning til skolens dannelsesoppdrag (Duenger Bøhn, 2024).

2.4. Dagens språkmodellers kunnskap om naturfag

Før jeg skal prøve å oppsummere hva vi vet om bruken av samtaleroboter i skolen, vil jeg redegjøre for GPT-modellenes evner innenfor naturfag. Det er ingen helhetlige undersøkelser på dette i naturfag, men det er flere undersøkelser innenfor spesifikke fagområder. Det bør merkes at flere av studiene jeg refererer til og flere av studiene referert til i litteraturgjennomgangene til Lo (2023) og Farrokhnia et al. (2023) ikke har gjennomgått fagfelleevaluering.

Litteraturgjennomgangen til Lo (2023) har sammenfattet forskning på ChatGPT-3.5 i utdanningskontekster. Der konkluderes det med at ChatGPT presterer fremragende innenfor overordnede ferdigheter som kritisk tenking og høyere ordens tenkning, men andre mener at ChatGPT sliter (Farrokhnia et al., 2023). Videre presterer ChatGPT bra innenfor programmering,

mens i matematikk og medisin konkluderes det med at ChatGPT-3.5 ikke presterer på et tilfredsstillende nivå. Imidlertid er det viktig å påpeke at disse vurderingene er gjort med utgangspunkt i høyere utdanning, og det kan tenkes at prestasjonene er tilstrekkelige på ungdomsskolenivå. Lo (2023) sin vurdering av medisin bygger på studier som viser at ChatGPT-3.5 nesten klarer å bestå medisinske eksamener (Fijačko et al., 2023; Kung et al., 2023). Noe liknende finner vi innenfor fysikk, der presterer Chat-GPT-3.5 godt nok til å bestå et introduksjonsemne i fysikk, men blir holdt tilbake av dårlige matematikkferdigheter (Kortemeyer, 2023). Innenfor kjemi evner Chat-GPT-3.5 å forstå symboler og problemer, men ikke å løse dem, ferdighetene blir dermed sett på som svake (Clark, 2023). Dette gjør seg gjeldende selv på videregående-nivå i Vietnam (Xuan-Quy et al., 2023). Imidlertid presterer GPT-4 bra i kjemi, når den kobles opp mot andre kjemiverktøy (Bran et al., 2023). Dette viser at det er et potensiale i å innlemme andre verktøy i språkmodellene.

Undersøkelser gjort av OpenAI (2023), viser at GPT-4 modellen presterer vesentlig bedre enn GPT-3.5. GPT-4 presterer i toppsjiktet for alle naturfagsrelevante Advanced Placement-eksamener, men her er det brukt en avansert ledetekst, few-shot (OpenAI, 2023). Few-shot-leadetekster innebærer at man gir et par eksempler på liknende problemer før en ber språkmodellen løse problemet (Brown et al., 2020). Sammenlikninger av matematikkferdighetene til GPT-3.5 og GPT-4, viser at den nye modellen presterer vesentlig bedre også her (Bubeck et al., 2023; Frieder et al., 2023). GPT-4 kan ifølge Bubeck et al. (2023, s. 37-39) modellere scenarioer fra virkeligheten, og presterer ifølge Frieder et al. (2023) på et bachelornivå i matematikk.

Likevel er det viktig å merke seg at selv om disse modellene stadig forbedres, har de fremdeles en tendens til å hallusinere (OpenAI, 2023, s. 46) og de begår grunnleggende feil mennesker ikke hadde gjort (Kortemeyer, 2023; OpenAI, 2023). Selv om modellene evner å bestå eksamener, vil de kunne produsere misinformasjon som kan medføre misoppfatninger hos elevene (Ding et al., 2023). Videre vil kvaliteten på språkmodellens respons avhenge av hvordan spørsmålet er stilt (White et al., 2023), noe som kan medføre at elevene får enda dårligere resultater.

2.5. KI i skolen

KI i skolen har vært diskutert lenge, men det var ikke før i slutten av 2022 at tematikken virkelig begynte å bli en del av den offentlige samtalen. Jeg vil nå vise til noe av forskningen som er gjort på bruken av samtaleroboter i skolen fra tiden før ChatGPT, og forskning som er gjort etter. Det bør igjen nevnes igjen at deler av forskningen på GPT-modellene ikke har gjennomgått fagfelleevaluering.

2.5.1. Før ChatGPT

Alemdag (2023) har gjennomført en metaanalyse av empirisk forskning på samtaleroboter før ChatGPT. Det er viktig å nevne at mange av studiene inkludert her er fra høyere utdanning, og at de bruker ulike samtaleroboter basert på ulik teknologi. Flere av samtalerobotene er gamle og mindre avanserte enn GPT-modellene. Analysen finner at samtaleroboter generelt sett har en medium, men signifikant, positiv innvirkning på læring (Alemdag, 2023), imidlertid finner en annen metaanalyse ikke denne effekten på grunnskolenivå (Wu & Yu, 2024). Effekten er størst når en ikke har tilgang til undervisning og når forelesning er undervisningsformen (Alemdag, 2023). Samtaleroboter reduserer stress sammenliknet med forelesninger, da læringen kan foregå i et tempo tilpasset studenten, noe som igjen øker motivasjonen (Yin et al., 2021). Imidlertid har elever i grunnskolen dårligere selvreguleringsferdigheter, og er mer utsatt for å gi opp om responsene er utilfredsstillende, trolig blir dette ytterligere forsterket hos elever som presterer dårlig på skolen (Chiu et al., 2023). På den andre siden er det en fordel av elevene ikke frykter å fremstå som uintelligente når de spør samtaleroboter om hjelp, noe de kan oppleve i forbindelse med lærere eller medelever (Rokhayani et al., 2022; Wu & Yu, 2024).

2.5.2. Bruk av GPT-modellene i utdanning

GPT-modellene er nye og det er begrenset med forskning på bruken av dem i utdanning, spesielt i autentiske kontekster (Yan et al., 2024). Jeg vil derfor trekke frem forskning fra mange ulike land, på tvers av både fagområder og utdanningsnivåer. Jeg vil starte med å trekke frem forskning knyttet til ulike måter GPT-modellene kan brukes, på, dette er som samtalepartner, vurdering, og kunnskapskilde. Deretter vil jeg se på hva vi vet om GPT-modellene og kritisk tenking.

Avslutningsvis ønsker jeg å trekke frem hva vi vet om hvorfor personer ønsker å bruke verktøyet i læringsarbeidet.

Samtaler med ChatGPT kan fremstå som samtaler en har med et ekte menneske, og dette har bruksområder i og utenfor undervisning (Farrokhnia et al., 2023). Den kan brukes til både å samtale om en tekst en gir til ChatGPT eller informasjon ChatGPT selv har produsert, samt som en samtalepartner i seg selv. Som samtalepartner kan ChatGPT brukes til rollespill og andre pedagogiske spill, om den blir instruert til å innta en viss rolle (Jeon & Lee, 2023). Elever kan også trene sine argumentasjonsferdigheter ved å be ChatGPT innta en posisjon og argumentere mot eleven (Farrokhnia et al., 2023). Et eksempel på dette er å be ChatGPT argumentere for at partikkelforståelsen av fotonet er den rette, mens eleven får i oppgave å finne feilene i argumentasjonen og skrive motargumenter (Bitzenbauer, 2023). ChatGPT kan også bli brukt som en assistent elevene kan henvende seg til, med den fordel at den kan svare umiddelbart, i og utenfor undervisningen (Farrokhnia et al., 2023; Jeon & Lee, 2023; Li & Xing, 2021).

Elever kan bruke ChatGPT som et verktøy for formativ vurdering. Tilbakemeldingene ChatGPT gir på elevers produkter kan tilpasses elevens behov, noen vil ha behov for å primært bli roset, mens andre vil ha behov for mer kritiske innvendinger (Farrokhnia et al., 2023). ChatGPT gir gode tilbakemeldinger på elevtekster, men noe dårligere enn erfarne lærere med god tid (Steiss et al., 2024). Imidlertid kan ChatGPT levere tilbakemeldingene umiddelbart, noe som gjør tilbakemeldinger mer effektive, og nytten er kanskje større på utkast enn ferdige tekster (Steiss et al., 2024). Tilbakemeldinger fra ChatGPT kan fremme positive følelser og øke elevens motivasjon til å revidere egen tekst, samt kvaliteten på arbeidet (Meyer et al., 2024). I naturfag kan en lærer gi elevene tilgang til en samtalerobot ment for å gi tilbakemeldinger på forskningsrapporter. Lærer kan da instruere den til å respondere med en viss tone og be den legge vekt på argumentasjon, tekstoppbygging og vitenskapelig innhold. Eleven kan levere et utkast til denne samtaleroboten, bruke tilbakemeldingene til å revidere og deretter levere til læreren. Andre muligheter for formativ vurdering kan være å be den lage spørsmål ment for å vurdere elevens kunnskaper, elevene kan svare på disse og bli vurdert av ChatGPT (Bitzenbauer, 2023; Jauhiainen & Guerra, 2023). I norskfaget opplever VGS-elever at Chat-GPT kan bistå med revideringen av en tekst, men den er langt fra perfekt og de opplever et behov for å være kritisk til det Chat-GPT kommer med (Bueie, 2024).

Flere studier har sett på å bruke ChatGPT som en kunnskapskilde for elever, både som en kilde i seg selv og som et verktøy for å tilpasse en gitt tekst til elevens nivå. ChatGPT brukt som

kunnskapskilde kan gjøre det lettere og mer effektivt å finne relevant informasjon, noe som vil kunne gjøre at elever bruker med tid på å prosessere og bruke informasjonen, og mindre tid på å lete etter den (Farrokhnia et al., 2023). ChatGPT vil også kunne kontekstualisere og tilpasse informasjonen den genererer til elevens faglige nivå, samt preferanser som tone og struktur (Farrokhnia et al., 2023; Kasneci et al., 2023). I tillegg til å selvstendig kunne produsere informasjon kan ChatGPT bli gitt en tekst og tilgjengeliggjøre den for elever (Jauhiainen & Guerra, 2023; Jeon & Lee, 2023). I studien til Jauhiainen og Guerra (2023) blir ChatGPT brukt til å dynamisk tilpasse informasjonen til elevens kunnskapsnivå og majoriteten av elevene rapporterte selv at de lærte mye og likte denne undervisningsformen. Alle lærerne intervjuet av Jeon og Lee (2023) mente at ChatGPT var egnet til å tilpasse tekster etter elevenes nivå. Imidlertid er det utfordringer med å bruke ChatGPT som en kilde til kunnskap. Språkmodeller som ChatGPT evner ikke å vurdere troverdigheten til treningsdataene og mangler en forståelse av meningen bak ordene den bruker, dette fører til at de ikke er feilfrie og kan reproducere fordommer (Farrokhnia et al., 2023).

På den ene siden kan ChatGPT være til hinder for kritisk tenkning, men det kan også bli brukt som et verktøy for å utvikle kritisk tenking. Det er mange misoppfatninger rundt ChatGPTs intelligens, og mange stoler blindt på, eller klarer ikke å identifisere når ChatGPT gjør feil (Ding et al., 2023). Når ChatGPT enkelt kan gi svar eller løse oppgaver, vil dette trolig kunne senke motivasjonen til å selv oppsøke og kritisk evaluere informasjon (Kasneci et al., 2023). Imidlertid vil intervensjoner kunne føre til at elever blir mer kritiske til ChatGPT. Bitzenbauer (2023) fikk elever til å generere en naturvitenskapelig tekst med ChatGPT, for så å be elevene finne feil, rette opp feilene ved bruk av kilder, og sammenlikne både ledeteksten gitt til ChatGPT og svarene med medelever. Etter undervisningen var de 53 elevene mer positive til bruken, men de var mer kritiske og mente at man i mindre grad kan bruke teknologien uten kjennskap til hvordan den fungerer.

Selvrapporing fra studenter viser at de som opplever stor arbeidsmengde og tidspress i større grad bruker ChatGPT (Abbas et al., 2024). Studenter som har positive holdninger til ChatGPT ønsker i større grad å ta verktøyet i bruk, og oppfattet nytteverdi er det som i størst grad påvirker holdninger (Albayati, 2024; Liu & Ma, 2024). Albayati (2024) finner at oppfattet brukervennlighet, tillit og holdninger blant de rundt en, også påvirker studentenes holdninger til

ChatGPT. Liu og Ma (2024) finner at oppfattet brukervennlighet påvirker oppfattet nytteverdi, men at det ikke kan predikere holdninger. Duong et al. (2023) finner at oppfattet nytteverdi og brukervennlighet påvirker bruk og intensjon om å bruke. De finner også at opplevelsen av at ChatGPT hjelper med å delta i faglige samtaler, fører til at intensjon om å bruke verktøyet i større grad resulterer i faktisk bruk.. Utenfor utdanningskonteksten viser undersøkelser at intensjon om å bruke også blir påvirket av oppfattet troverdighet, og i hvilken grad en oppfatter ChatGPT som interaktivt og dermed menneskelignende (Camilleri, 2024).

2.6. Rammeverk for teknologiaksept

For å undersøke hvordan elevene har tatt i bruk denne nye teknologien har jeg sett til forskning på hva som påvirker brukere til å ta i bruk teknologi, og tilpasset dette til min studie for å utvikle et rammeverk. Rammeverket er inspirert av teknologiakseptmodellen (TAM) (Davis, 1989), og utvidet med kognitiv absorpsjon (Agarwal & Karahanna, 2000; Lowry et al., 2013) og kredibilitetsteori (Assaker, 2020). TAM blir vanligvis brukt i kvantitative studier, men TAM og modeller basert på TAM brukes også i kvalitative studier (Menon & Shilpa, 2023; Vogelsang et al., 2013). Ved å ta utgangspunkt i faktorer som tidligere har vist seg å være av betydning for bruk og aksept av teknologi, har jeg utviklet spørreskjemaet, intervjuguiden og analysert intervjuene.

Teknologiakseptmodellen (TAM), utviklet av Davis (1989), er en modell som har som formål å forstå og forutsi hvordan brukere aksepterer og bruker ny teknologi. Modellen foreslår at brukeres beslutning om å bruke ny teknologi hovedsakelig drives av to faktorer: den oppfattede nytteverdien og oppfattet brukervennlighet.

Den oppfattede nytteverdien er et mål på i hvilken en bruker tror at den nye teknologien vil hjelpe med å utføre oppgavene mer effektivt. Den oppfattede brukervennligheten er et mål på i hvilken en bruker tror at teknologien vil være enkel å bruke, uten anstrengelser. Modellen ble utviklet i en tid hvor teknologi fikk større og større plass i folks hverdag, og en ønsket å forstå hvilke faktorer som påvirket hvorvidt folk tok i bruk ny teknologi (Marangunic & Granić, 2015). TAM har ytre motivasjon som utgangspunkt, og fokuserer i liten grad på indre motivasjon eller gleden forbundet med bruken (Lowry et al., 2013, s. 618-619). TAM blir brukt som en selvstendig modell, men det er også vanlig å utvide den med andre relevante faktorer basert på hva en skal undersøke (Marangunic & Granić, 2015). Ulike faktorer brukt for å utvide modellene vil kunne

ha en direkte innflytelse på intensjon om å bruke, men de kan også ha sin innflytelse gjennom å påvirke nytteverdi eller brukervennlighet (Marangunić & Granić, 2015).

Hvorvidt elevene har tillit til at språkmodellene gir troverdige svar anser jeg som spesielt relevant i forbindelse med elevenes adoptering av denne teknologien. Det kan argumenteres for at det ikke er behov for å utvide TAM da teknologiens troverdighet kan ses på som en del av dens nytteverdi. Imidlertid viser forskning på ChatGPT som læringsverktøy at for noen kan teknologien oppleves som nyttig, selv når den oppleves utroverdig (Ding et al., 2023). For å se på hvordan tillit til teknologien påvirker bruk har TAM blitt utvidet på mange ulike måter (Wu et al., 2011). Hva som blir målt i studier som har inkludert tillit varierer, og det er ikke klart hvilken effekt det har (Wu et al., 2011). I sin studie på Wikipedia har Meseguer-Artola et al. (2016) inkludert *oppfattet kvalitet* som et konstrukt for å undersøke bruk. En av tilnærmingene til tillit som Wu et al. (2011) finner er å se på tillit som både tillit til kvaliteten, men også tillit til intensjonen. En liknende todeling finnes hos Assaker (2020), som har brukt kredibilitetsteori og utvidet TAM med to konstrukt, *troverdighet* og *ekspertise*, for å undersøke bruk av brukergenerert innhold på internett. Etersom utfordringene med språkmodeller går på både hvorvidt de er faktisk korrekte samt hvorvidt de viderefører skjevheter og fordommer, mener jeg en slik todeling er fornuftig. Jeg har tatt inspirasjon fra Assaker (2020) og utvidet modellen min med *troverdighet* og *ekspertise*.

Indre motivasjon og absorpsjon er en pådriver for bruk av teknologi, spesielt teknologi som sosiale medier og spillbaserte læringsplattformer, men det gjør seg gjeldene for all teknologi (Agarwal & Karahanna, 2000; Lowry et al., 2013). For å få et mer helhetlig bilde av teknologiaksept, har det vist seg fordelaktig å aspekter som reflekterer indre motivasjon mer presist. En slik utvidelse er *kognitiv absorpsjon*, som har blitt introdusert av Agarwal og Karahanna (2000). Kognitiv absorpsjon er satt sammen av de fem ulike konstruktene, kontroll, nysgjerrighet, glede, fokusert innlevelse og tidsdissosiasjon. Agarwal og Karahanna (2000) har vist at kognitiv absorpsjon har innflytelse på brukerens oppfattelse av teknologiens nytteverdi og brukervennlighet, men også direkte på intensjon om å bruke, noe som viser til at kognitiv absorpsjon er en helhetlig del av teknologibruk. Flow-teori og absorpsjon er det teoretiske grunnlaget for kognitiv absorpsjon (Agarwal & Karahanna, 2000). Flow-teori viser til en tilstand hvor en er fullstendig oppslukt av en aktivitet, ofte kjent som flyttilstand. På denne måten blir

kognitiv absorpsjon brukt for å måle indre motivasjon og engasjement med teknologien. Jeg har inkludert de fem konstruktene for kognitiv absorpsjon, for å kunne undersøke hvilken rolle disse har for elevenes adopsjon av teknologien. Det faktumet at mange elever har tatt teknologien i bruk på eget initiativ, gjør det ytterligere relevant å inkludere disse.

I mitt endelige rammeverk har jeg ti konstrukter, hvor to kommer fra kredibilitetsteori og fem kommer fra kognitiv absorpsjon. For å operasjonalisere bruk har jeg tatt valgt intensjon om å bruke. Hele oversikten over rammeverket og en beskrivelse av hvert konstrukt finnes i tabell 1.

	Konstrukt	Beskrivelse
(Davis, 1989; Lowry et al., 2013)	Intensjon om å bruke	I hvilken grad en person har tenkt til å bruke teknologien i fremtiden.
TAM (Davis, 1989)	Oppfattet nytteverdi	I hvilken grad en person oppfatter at teknologien bidrar til bedre og å effektivisere arbeidet.
	Oppfattet brukervennlighet	I hvilken grad en person oppfatter teknologien som enkel å bruke, uten anstrengelser.
Kredibilitetsteori (Assaker, 2020)	Troverdighet	I hvilken grad en person oppfatter at teknologien er objektiv og sannferdig.
	Ekspertise	I hvilken grad en person oppfatter at teknologien har ekspertise og kompetanse.
Kognitiv absorpsjon (Agarwal & Karahanna,	Kontroll	I hvilken grad en person oppfatter å ha kontroll over interaksjonen med teknologien, om den responderer slik en ønsker.

2000; Lowry et al., 2013)	Nysgjerrighet	I hvilken grad teknologien er i stand til å stimulere en persons nysgjerrighet.
	Fokusert innlevelse	I hvilken grad en person har oppmerksomheten rettet mot oppgaven, og til hvilken grad forstyrrelser blir ignorert.
	Glede	I hvilken grad en person oppfatter at teknologien er morsom å bruke og hvorvidt en føler glede og tilfredsstillelse.
	Tidsdissosiasjon	I hvilken grad en person mister evnen til å oppfatte tidsforløpet når en bruker teknologien.

Tabell 1: Rammeverket brukt i denne studien. Øverst er grunnkonstruktene fra TAM, etterfulgt av mine utvidelser fra kredibilitetsteori og kognitiv absorpsjon.

3. Metode

Denne masteroppgaven tar sikte på å utvide kunnskapen om hvordan og hvorfor elever i naturfag har tatt i bruk språkmodeller, samt deres forståelse av teknologien. Gitt at det er elevene som i stor grad har stått for teknologiens inntog i skolen, anser jeg det som svært relevant å undersøke dette fra deres perspektiv. For å utforske dette, valgte jeg en kvalitativ tilnærming.

Jeg gjennomførte et undervisningsopplegg i to parallellklasser som hadde til hensikt å gi alle elevene noe kunnskap om språkmodeller og erfaring med å bruke Copilot som et verktøy for å tilegne seg kunnskap. Det var dog tekniske utfordringer som førte til at ikke alle elevene fikk gjennomført hele undervisningsopplegget, men alle elevene hadde noe erfaring med språkmodeller fra tidligere. I etterkant av undervisningsopplegget ble elevene gitt ett anonymt spørreskjema, og fire elevgrupper ble valgt ut til fokusgruppeintervju. Fokusgruppeintervjuene ble gjennomført med en intervjuguide og tatt opp og transkribert, disse dataene utgjør hoveddelen av min studie. Kvalitativ innholdsanalyse, inspirert av Hsieh og Shannon (2005) ble brukt for å

analysere intervjudataene. Spørreskjema ble brukt som et supplement til intervjuene for å kunne si noe generelt om hele elevgruppen

Dette kapittelet starter med en begrunnelse for valget av kvalitativ metode, etterfulgt av en beskrivelse av utvalget og konteksten. Jeg vil beskrive hvordan og hvorfor jeg har brukt spørreskjema og fokusgruppeintervjuer for å samle data. Deretter vil jeg beskrive hvordan dataene er analysert, og drøfte oppgavens troverdighet og etiske aspekter.

3.1. Metodologi

Gitt studiens natur og mål, har jeg valgt en kvalitativ tilnærming for å få en dypere forståelse av elevenes bruk av, erfaringer med og forståelse av språkmodeller. Kvalitative tilnærminger sikter på å gå i dybden på og forstå sosiale fenomener som ikke lar seg måle med tall (Thagaard, 2018). Tilnærmingen egner seg dermed til å undersøke bruken av språkmodeller i naturfag, da det er et område med lite forskning. Ved en kvalitativ tilnærming mister jeg imidlertid muligheten til å generalisere funnene til elever generelt, og studiens overførbarhet må derfor vurderes enkeltvis. Tidligere forskning på og teori om teknologiaksept, beskrevet i kapittel 2.6 (Agarwal & Karahanna, 2000; Assaker, 2020; Davis, 1989; Lowry et al., 2013) har blitt brukt som inspirasjon for studiens teoretiske rammeverk. For å undersøke hvilken KI-literacy elevene gir uttrykk for har jeg tatt utgangspunkt i de overordnede temaene beskrevet av Long og Magerko (2020).

3.2. Kontekst

Prosjektet ble gjennomført over to dager i hver klasse. Den første dagen bestod av en time hvor jeg introduserte elevene til prosjektet og en kort undervisningsøkt om hvordan GPT-modellene virker. Den andre dagen skulle elevene få erfaring med å bruke Copilot, men i klasse to fikk ikke elevene selv brukt Copilot. På slutten av økten svarte elevene på spørreskjema og i de to påfølgende timene tok jeg ut elever til to fokusgruppeintervjuer.

Undervisningen om GPT-modellene

Et 30 minutter langt undervisningsopplegg om teknologien bak GPT-modellene ble gjennomført den første dagen. Dette ble gjort fordi elevene i neste økt selv skulle bruke Copilot som et læringsverktøy, noe som krever at elevene har kunnskap om hvordan bruke Copilot og en

forståelse for nødvendigheten av å være kritisk til Copilot (Long & Magerko, 2020; White et al., 2023). Undervisningen ble gjennomført ved hjelp av PowerPoint-presentasjon. Elevene var aktive underveis gjennom å diskutere med læringspartner samt å ta del i helklassesamtaler. Hvordan språkmodeller virker ble undervist om ved hjelp av skjermbilder jeg tok av OpenAIs Playground (OpenAI, 2024). Her vises det er liste med tokens (deler av ord) og sannsynligheten for at de følger etter teksten «Roser er». Deretter ble elevene vist en tekst, hvor slutten av teksten var skjult. Elevene ble bedt om å gjette hva det stod. Dette ble brukt for å demonstrere hvordan modellene trenes. Etter å ha demonstrert hvordan modellene trenes, oppsummerte jeg kort at modellene fremstår kunnskapsrike, men at de kun ser på strukturer i språk. For å tydeliggjøre dette brukte jeg et håndplukket eksempel med de to ledetekstene, «vann er» og «dihydrogenoksid er». Dette ble brukt for å vise hvordan ordbruk kan påvirke resultatene og at det som kan være sannsynlig ikke behøver å være korrekt, da svaret på den sistnevnte var «dihydrogenoksid er ekstremt farlig for mennesker».

Elevene ble vist noen mulige måter å ta i bruk Copilot. Her la jeg vekt på at svaret en får er et produkt av spørsmålet en stiller, dermed er det viktig å være presis. I tillegg kan en be om at svarene skal tilpasses til sitt eget nivå ved å gi kontekst som for eksempel klassetrinn. For at elevene skulle kunne ta i bruk samtaleelementene disse modellene tilbyr, ble de anbefalt å samtale med Copilot, ved for eksempel å stille oppfølgingsspørsmål. Jeg tipset også elevene om at når de føler at de har forstått noe, kan de gi sin forklaring til Copilot og be om innspill.

Undervisningsopplegget for klasse en

Undervisningsopplegget dag to, varte i 60 minutter. Her ble elevene vist en kort video som viste syv ulike vitenskapelige fenomener, og de valgte selv et av disse. Elevenes oppgave var å finne ut årsaken bak fenomenet, ved hjelp av Copilot. Mot slutten av timen skulle de forklare det til klassen.

Et av klippene i videoen var av en mann som gjentatte ganger tar hånden kjapt gjennom en strøm av flytende metall, helt uten å skade seg. Elevene ble kun vist slike korte klipp, uten noen forklaring av det vitenskapelige fenomenet. Elevene brukte egne forkunnskaper og beskrev hva de så til Copilot, for å finne navnet på og forklaringen på fenomenet. I dette klippet var målet at elevene skulle finne frem til Leidenfrost-effekten.

Oppsummeringen av hva elevene fant ut, ble gjennomført som en helklassediskusjon i slutten av timen. Her ble noen av elevene bedt om å forklare hva de hadde funnet ut, mens jeg og andre elever stilte spørsmål. Jeg gjorde dette fordi jeg ønsket at elevene skulle ha fokus på å sette seg inn i og forstå fenomenet, og ikke lage et produkt. Om de skulle lage et produkt som en tekst eller presentasjon fryktet jeg at flere kun kom til å plagiere Copilot. Diskusjonen ble også gjennomført fordi jeg ønsket en vurdering i slutten av timen.

Undervisningsopplegget for klasse to

Intensjonen var å gjennomføre det samme undervisningsopplegget i klasse to, men det var tekniske utfordringer som gjorde at mange elever ikke kom inn på Copilot. Undervisningen startet likt, men når det ble klart at mange ikke kom inn på Copilot valgte jeg å gjennomføre et opplegg på storskjerm. Jeg tok frem ChatGPT via OpenAIs Playground (OpenAI, 2024) på storskjerm, og vi valgte å undersøke videoen om leidenfrost-effekten. Elevene kom med innspill til spørsmål som jeg skrev inn. Deretter diskuterte vi svaret og hvordan det bør revideres slik at vi får svar på det vi ønsker å vite. Når vi fikk et svar som virket troverdig og relevant ble dette sjekket opp mot andre kilder. Som en følge av dette gjennomførte jeg ikke vurdering gjennom helklassediskusjon på slutten av timen.

3.3. Utvalg

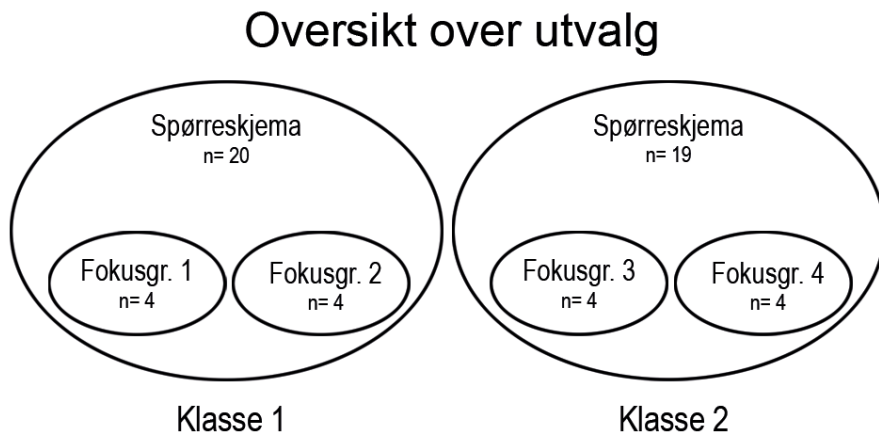
Datainnsamlingen ble gjort i to parallellklasser på 10. trinn ved en skole på Østlandet. Skolen ble funnet gjennom universitetets nettverk. Elevene brukte PC på skolen og de skulle ha tilgang til Copilot gjennom sin skolemail. Klassenes lærere var begge interesserte i KI og så på det som potensielt nyttige verktøy.

I klasse 1 hadde ingen av elevene brukt Copilot før, men alle hadde brukt enten ChatGPT og Snapchat MyAI. Noen elever hadde brukt disse verktøyene mye, mens andre hadde kun litt erfaring. Bruken var på eget initiativ, ikke som en del av undervisningen, men læreren hadde anbefalt elevene å prøve det ut.

I klasse 2 hadde elevene noe erfaring med Copilot og de fleste hadde noe erfaring med bruk av språkmodeller som en del av naturfagundervisningen. Elevene hadde også erfaring med å lete

etter feil i KI-genererte bilder av naturfaglige modeller. Alle elevene i klasse 2 hadde erfaring med bruk av språkmodeller på eget initiativ.

Totalt bestod utvalget av 39 elever, hvor 16 deltok på fokusgruppeintervju. Figur 1 viser en oversikt over utvalgene.



Figur 1: Oversikt over utvalgene i hver klasse

3.3.1. Fokusgruppene

I klasse en valgte jeg ut fokusgrupper på bakgrunn av observasjon av elevene mens de brukte Copilot. I klasse to ble fokusgruppene valgt ut på bakgrunn av helklassediskusjonen og lærerens kjennskap til elevene. Til sammen valgte jeg ut 16 elever til fire fokusgruppeintervjuer, to i hver klasse. Jeg vil først beskrive og begrunne kriteriene for utvelgelse til fokusgruppene, for deretter å beskrive hvordan utvelgelsen ble gjennomført i hver av klassene.

De to fokusgruppene i hver klasse var ment å representere to ytterpunkter på en skala av oppfattet nytte. En fokusgruppe for høy oppfattet nytteverdi og en for lav oppfattet nytteverdi. Jeg ønsket at alle gruppene skulle favne om et bredt utvalg av strategier, altså ulike måter å ta i bruk verktøyene. Rasjonalet bak valget av heterogene grupper med tanke på strategier, er at det vil kunne føre til en bredere diskusjon rundt hvordan språkmodellene kan brukes (Halkier, 2010, s. 30). Hensikten bak å ha homogene grupper med tanke på oppfattet nytteverdi var todelt. For det første ønsket jeg å kunne gå i dybden på hvorfor verktøyet kunne være både nyttig og unyttig. For det andre vil en gruppesammensetning der elevene i større grad deler erfaringsgrunnlag gjøre det lettere for elevene å delta i samtalen (Halkier, 2010, s. 32). Jeg fryktet at en elev som selv

oppfattet verktøyet som unyttig ikke ville være komfortabel med å si dette om alle de andre oppfattet det som veldig nyttig. Som et ytterligere tiltak mot dette bisto lærerne i sammensetningen av fokusgruppene for å sørge for at fokusgruppene bestod av elever som var trygge på hverandre og dermed kunne dele åpent (Halkier, 2010, s. 34).

Utvelgelse i klasse en

Jeg gjennomførte deltakende observasjon (Fangen, 2010), for å få oversikt over hvilke elever som brukte hvilke strategier samt hvor nyttig de oppfattet Copilot, slik at jeg kunne sette sammen hensiktsmessige fokusgrupper. Observasjonen var dermed ikke med i analysen. I forkant av observasjonen utviklet jeg et egnet observasjonsskjema.

Jeg ønsket kjennskap til hvordan elevene brukte Copilot og hvordan de opplevde dette. Perspektivet en får som deltakende observatør lar en få dypere kjennskap til hva informantene holder på med og hvordan de opplever dette, enn det man får gjennom utvendige metoder (Fangen, 2010). Jeg gikk inn i rollen som lærer, og gikk rundt i klasserommet for å se på arbeidet til elevene, spørre og utfordre dem. På denne måten viste alle elevene hva de hadde gjort til meg, og jeg fikk spurt dem når jeg lurte på noe. Når elevene hadde spørsmål, henvendte de seg til meg. Dette ga meg innsikt i hvilke utfordringer elevene møtte på. Imidlertid er utfordringen med den deltakende rollen at det er vanskeligere å systematisk dokumentere funn.

Observasjonsskjema, figur 2, var utformet som et klassekart, hvor hver pult var todelt. Den ene delen bestod av en rekke stikkord relatert til ulike strategier, samt plass til å notere ned nye strategier. Stikkordet «samtale» ble brukt om eleven hadde en dialog med Copilot, og «utdyp», «ris» og «ros» er brukt om eleven ba Copilot utdype sitt svar og om de ga Copilot ris eller ros. Valgene av stikkord er basert på litteratur om bruk av språkmodeller og egen erfaring. Den andre er en linje ment for å representere hvor positive elevene var til å bruke Copilot. På denne måten kunne jeg raskt notere ned observasjoner for hver elev, samt ha oversikt over hvilke elever jeg mangler observasjoner på. Det utfylte observasjonsskjemaet gjorde det lett og oversiktlig å sette sammen fokusgruppene, noe som var nødvendig siden intervjuene skulle foregå kort tid etter.

Figur 2: Observasjonsskjema, hver boks representerer en elev, en del for strategi og en for nytte

Utvelgelse i klasse to

Utvelgelsen av elever til fokusgruppeintervjuene i klasse to ble gjennomført på bakgrunn av elevenes deltakelse i helklassediskusjonen og ved hjelp av lærerens kjennskap til elevene. I klasse to ble gjennomføringen som nevnt annerledes, noe som medførte at jeg ikke kunne velge ut elever på samme måte som i klasse en. Enkelte elever kom med innspill i helklassediskusjonen som jeg anså som interessante. Dette var både med tanke på valg av strategi i spørsmålsstillingen, og en mer kritisk holdning til teknologiens nytte. Utover disse elevene fikk jeg hjelp av læreren til å velge ut elever. Jeg forklarte hensikten med å ha to ulike fokusgrupper til læreren, og hvilke av elevene jeg allerede hadde funnet. På bakgrunn av dette valgte læreren ut resten av elevene.

Klasse 1		Klasse 2	
Fokusgruppe 1	Fokusgruppe 2	Fokusgruppe 3	Fokusgruppe 4
Hans	Lise	Mari	Henrik
Mona	Rikke	Janne	Lars
Ida	Jacob	Thea	Dan
Alina	Jens	Tiril	Pia

Tabell 2: Oversikt over fokusgruppene og tilhørende elever

Hver av fokusgruppene bestod av fire personer, totalt var det 16 elever som ble intervjuet hvorav 10 var jenter og 6 var gutter. Klasse to hadde ifølge lærerne et generelt høyere faglig nivå enn klasse en. Fokusgruppe to og tre hadde et middels faglig nivå og fokusgruppe en og fire hadde et høyt faglig nivå, ifølge lærerne. Tabell 2 viser hvilke elever som tilhørte de ulike fokusgruppene, navnene er fiktive.

3.4. Datainnsamling

I dette delkapittelet vil jeg beskrive intervjuguiden og hvordan jeg har gjennomført og transkribert fokusgruppeintervjuene. Deretter vil jeg beskrive hvordan spørreskjemaet er utviklet og bygget opp.

3.4.1. Fokusgruppeintervju

Fokusgruppeintervjuer kjennetegnes ved at en har seks til ti personer som samtaler om et emne, hvor intervjuerens rolle er å moderere og fasilitere for diskusjon blant deltakerne, heller enn å direkte lede samtalen (Kvale & Brinkmann, 2015, s. 179-180). Imidlertid anbefaler Wibeck et al. (2007) mindre grupper for å gi alle deltakerne større mulighet for å delta og dermed fremme en rikere diskusjon. Derfor valgte jeg fokusgrupper med fire elever. Ved å studere elevenes samtaler om et gitt tema, er tanken å få innsikt i hvordan de forstår og beskriver temaet ut fra deres eget perspektiv (Wibeck et al., 2007). Gruppedynamikken lar elevene dele, diskutere og utfordre hverandre noe som kan få frem varierte gjennomtenkte perspektiver og holdninger som eksisterer i gruppen (Kvale & Brinkmann, 2015; Thagaard, 2018; Wibeck et al., 2007). De livlige diskusjoner gjør fokusgruppeintervjuer bedre egnet enn individuelle intervjuer, til å få frem de emosjonelle aspektene (Kvale & Brinkmann, 2015, s. 179-180), altså de underliggende følelsesmessige reaksjonene knyttet til elevenes opplevelser. Fokusgruppeintervjuer medfører ofte mindre strukturerte samtaler enn individuelle intervjuer (Kvale & Brinkmann, 2015). Det kan ifølge Kvale og Brinkmann (2015) gjøre fokusgruppeintervjuer velegnet til eksplorative studier, da det gir informantene større mulighet til å legge premissene for hva som er relevant. En ulempe med en slik åpen tilnærming er at intervjuene vil kunne bli forskjellige, og resultatene vanskeligere å sammenlikne (Thagaard, 2018). Imidlertid kan den åpne delingen av strategier og erfaringer på elevenes premisser, ha ført til at de reflekterte mer og dermed lærte mer enn de hadde gjort gjennom individuelle intervjuer. Dette er i tråd med de generelle forskningsetiske

retningslinjene utviklet av De nasjonale forskningsetiske komiteene (2014), som sier at forskningen skal komme deltakerne til gode. For å både gi elevene frihet til å samtale om temaet slik de finner det relevant fra deres perspektiv, og sørge for at intervjuene dekket de relevante temaene, benyttet jeg meg av en delvis strukturert tilnærming. Jeg utviklet en delvis strukturert intervjuguide. Intervjuguiden ble utviklet med utgangspunkt i studiens rammeverk, som er inspirert av teknologiakseptmodellen (Davis, 1989), og utvidet med kredibilitetsteori (Assaker, 2020) og kognitiv absorpsjon (Agarwal & Karahanna, 2000; Lowry et al., 2013), slik det er beskrevet i kapittel 2.6. Intervjuguiden ble endret noe underveis blant annet fordi elevene la stor vekt på erfaringene med MyAI og ChatGPT. Den endelige intervjuguiden finnes i vedlegg 1.

Jeg vil gi en kort beskrivelse av strukturen i intervjuene. Introduksjonen bestod av en kort gjennomgang av hensikten med intervjuet, retten til å trekke seg og informasjon om lydopptaket. Deretter startet selve intervjuet med noen enkle åpningsspørsmål om elevenes forhold til språkmodeller. Hensikten her var å bruke enkle spørsmål for å åpne samtalen (Kvale & Brinkmann, 2015). I klasse en ble elevene også spurt om deres erfaringer med å bruke Copilot i undervisningsøkten. Elevene begynte deretter å prate fritt om deres erfaringer, og jeg brukte intervjuguiden som et verktøy for å moderere samtalen. Intervjuguiden inneholdt en liste med punkter under hvert av hovedtemaene, som jeg krysset av på når jeg opplevde at de var besvart. På denne måten ble det tydelig for meg i hvilken retning jeg burde flytte samtalen. For eksempel når en av elevene i fokusgruppe en pratet om at det var vanskelig å få Copilot til å besvare oppfølgingsspørsmål, ba jeg elevene utdype deres opplevelse av å ha kontroll over språkmodellenes responser, på en måte som også åpnet opp for motstridene syn. Etter hvert som tiden gikk ut og punktene ble krysset av, beveget vi oss til avslutningen av intervjuet. For å runde av intervjuet gjenga jeg kort hovedtrekk i elevenes svar og forhørte meg om hvorvidt de var enig i denne beskrivelsen og om de ønsket å legge til noe. Til slutt takket jeg for deltakelsen, spurte om deres opplevelse av intervjusituasjonen og ga elevene muligheter til å stille spørsmål til meg. Dette ble gjort både for å ivareta elevene og for å få muligheter til å forbedre egen intervjupraksis jamfør anbefalingene til Kvale og Brinkmann (2015).

3.4.2. Transkripsjon

Intervjuene ble tatt opp med Nettskjema-diktafon. Jeg brukte transkriberingsfunksjonen i nettskjema, og lastet ned disse transkripsjonene. Disse ble deretter lastet inn i programmet

NVivo, hvor jeg ferdigstilte transkripsjonen med hjelp av lydopptaket spilt av fra nettskjema. Hensikten med transkripsjonen er å oversette de muntlige samtaleene til skriftlig form slik at de kan analyseres (Kvale & Brinkmann, 2015, s. 204-205), og det er viktig for å sikre informantenes personvern. Gjennom å transkribere vil dataene naturligvis bli svekket, da mange aspekter ved en muntlig samtale ikke lar seg direkte oversette til skriftlig form, og jeg som transkribent må ta aktivt stilling til hva som skal prioriteres (Kvale & Brinkmann, 2015, s. 204-205, 208). Jeg har transkribert intervjuene til noe som minner om en mer formell skriftlig stil, da jeg har utelatt overfløydige ord som «*mmh*» og gjentakelser. I transkripsjonene har jeg tatt med følelsesuttrykk som latter og når noe ble sagt nølende, i hakeparenteser. Flere av elevene var usikre på flere sider ved KI, som for eksempel hvordan den virket, og kom med uttalelser som var nølende eller spørrende. For å ikke feiltolke slik utsagn som sikre bestemte utsagn når jeg analyserte dataene, ble dette markert med [nølende]. Noen av elevene fortalte om morsomme opplevelser med språkmodellene, som førte til latter hos både dem selv og resten av elevene. Jeg markerte dette med [latter], fordi det ikke alltid kom frem at elevene oppfattet dette som morsomt i ren tekst. For å sikre personvernet ble alt skrevet på bokmål og elevene gitt fiktive navn. Der hvor elevene nevnte navnet på ting som læreren eller idrettsklubben sin ble disse endret til [lærer] og [idrettsklubb].

Direkte oversettelse fra muntlig til skriftlig språk kan medføre at informantene fremstår forvirret og uintelligente (Kvale & Brinkmann, 2015, s. 213-214). Sitater fra intervjuene gjengitt i studien har derfor blitt omformulert slik at utsagnene er sammenhengende og dermed lettere å lese, samtidig som elevene ikke skal bli satt i et dårlig lys. For eksempel: «*Jeg pleier å skrive svar enkelt og kort. For jeg orker ikke sånn langt. Da pleier man å skrive sånne punkter med det viktigste.*» ville blitt omformulert til «*Jeg pleier å skrive svar enkelt og kort, for jeg orker ikke lange [tekster]. Da pleier den å skrive punkter med det viktigste.*»

3.4.3. Spørreskjema

På bakgrunn av rammeverket jeg utviklet lagde jeg et spørreskjema. Spørreskjemaet hadde til hensikt å supplere resultatene fra intervjuene på en slik måte at jeg kunne si noe om hele elevgruppen, samt å gi elevene mulighet til å uttrykke seg anonymt, noe som kan føre til mer ærlige svar (Nardi, 2018, s. 73). Spørreskjema ble gitt i forkant av intervjuene, av praktiske årsaker. Spørreskjemaet bestod av utsagn som elevene tok stilling til med en femtrinns Likerts-

skala, slik det er gjort av Palos-Sanchez et al. (2022). Likerts-skalaer lar en måle respondentenes holdninger til en gitt påstand, ved at de krysser av på en skala fra «svært uenig» til «svært enig» (Nardi, 2018, s. 80-81). Med en femtrinns skala kan respondentene si seg enig/uenig eller svært enig/uenig, alternativt kan de innta en nøytral posisjon. Jeg valgte en femtrinns Likerts-skala da det lar meg få frem nyanser samtidig som jeg begrenser omfanget til spørreskjemaet.

Spørreskjemaet i sin helhet finnes i vedlegg 2.

Spørreskjemaet ble utviklet på bakgrunn av rammeverket og den tilhørende litteraturen beskrevet i kapittel 2.6. Jeg tok inspirasjon fra Palos-Sanchez et al. (2022) som undersøke elevers opplevelse av å bruke Kahoot. Utsagnene knyttet til *troverdighet* og *ekspertise*, ble utformet etter inspirasjon fra Assaker (2020). For å unngå at elevene kun krysset av på samme svaralternativ, og for å kunne måle hvorvidt elevene faktisk hadde lest og tatt stilling til påstandene, ble liknende utsagn formulert både som positive og negative, i tråd med anbefalinger fra Nardi (2018, s. 84-85). Om man oppfattet at Copilot var et lettvent verktøy å bruke krysser man et sted av på enig og et annet uenig. Studiens omfang og antall deltakere gjør den ikke egnet til å gjennomføre statistiske analyser slik de gjøres i TAM-litteraturen. Jeg mener likevel at resultatene er med på å styrke studiens kvalitet da disse gjør det mulig å si noe om klassene som helhet og ikke kun fokusgruppene.

3.5. Analyse av intervjuene

En kvalitativ innholdsanalysen ble gjennomført etter inspirasjon fra Hsieh og Shannon (2005). De beskriver tre ulike fremgangsmåter, og jeg benytter det de kaller «directed», eller teoridrevet, slik det blir oversatt av Fauskanger og Mosvold (2014). Den teoridrevne tilnærmingen egner seg for å utvide eller validere tidligere teori (Hsieh & Shannon, 2005). For å besvare forskningsspørsmål 1 og 2 tok analysen utgangspunkt i det teoretiske rammeverket for bruk og aksept av teknologi, beskrevet i kapittel 2.6. Rammeverket er inspirert av teknologiakseptmodellen (Davis, 1989), og er utvidet med kredibilitetsteori (Assaker, 2020), samt kognitiv absorpsjon (Agarwal & Karahanna, 2000; Lowry et al., 2013). For å besvare forskningsspørsmål 3, ble de fire overordnede temaene Long og Magerko (2020) bruker for å beskrive KI—literacy, utgangspunktet for koder.

I den teoridrevne tilnærmingen starter analysen med teori. Teorien brukes for å lage kodene som skal brukes for å systematisere datamaterialet (Hsieh & Shannon, 2005). Å utvikle og følge klare retningslinjer for kodingsprosessen er viktig for å sikre analysens reliabilitet (Hsieh & Shannon, 2005).. Åpenhet rundt disse er viktig for at forskningen skal kunne etterprøves, og for at den skal kunne gjøres åpen for kritikk. Jeg vil her beskrive og begrunne valgene jeg har tatt som utgjør retningslinjer for kodingen. Deretter vil jeg kort beskrive selve kodingsprosessen og gi eksempler på de ulike kodene.

Fokusgruppeintervjuer vil, som nevnt kunne være litt uforutsigbare, og elevene selv spiller en viktig rolle i å drive samtalen. Som en følge av dette er ikke hele datamaterialet relevant for å besvare forskningsspørsmålene, og disse ble dermed ikke kodet. Jeg mener det har vært hensiktsmessig å kunne bruke flere koder på samme utsagn. Noen av elevenes utsagn bidrar til å belyse mer enn kun ett konstrukt, og for å kunne få et helhetlig bilde av hvert konstrukt, ble alt som var relevant for et gitt konstrukt kodet deretter. Utsagn som *«Når det gjelder definisjoner, vil jeg heller ha en fra en anerkjent fagperson, enn fra en KI som bare sammenfatter masse data»* fortelle både noe om nytteverdien og synet på teknologiens ekspertise. Lars mener at teknologien ikke er nyttig når det gjelder å definere noe, da teknologien ikke har den ekspertisen fagpersoner har, som en følge av språkmodellens natur. Slike utsagn er dermed med på å gi et bilde av teknologens nytteverdi og dens ekspertise. For å ikke miste konteksten har jeg alltid kodet hele elevbidrag, og der det har vært nødvendig, har jeg inkludert større deler av samtalen. Monas utsagn *«Men nå prater du med MyAI, er jo litt som å prate med et menneske, føler jeg. Det er jo som å snakke med venninna di.»*, kan isolert sett si noe om opplevelsen av å prate med MyAI. Konteksten det blir sagt i gjør det tydelig at dette også er en svakhet ved MyAI, da den ikke fremstår like smart og profesjonell som ChatGPT. Å beholde konteksten var med andre ord viktig for å tydeliggjøre hvordan elevens utsagn bidro til å belyse forskningsspørsmålene.

For å sørge for at jeg ikke gikk glipp av relevante data i kodingsprosessen, ble transkripsjonene kodet i to omganger. Kodingen ble først gjennomført ved at jeg jobbet meg gjennom transkripsjonene og kodet fortløpende. Deretter gjennomgikk jeg transkripsjonene med søkelyset på en og en kode. For eksempel lette jeg gjennom transkripsjonene etter utsagn som var med på å si noe om opplevelsen av kontroll over interaksjonen med teknologien. Ved å lete gjennom

transkripsjonene på denne måten fant jeg ytterligere utsagn som var med på å belyse det gitte konstruktet.

Koding av dataene ble gjennomført i programmet NVivo etter retningslinjene beskrevet ovenfor. Dataene ble først kodet etter rammeverket beskrevet i kapittel 2.6, deretter etter de fire overordnede temaene for KI-literacy beskrevet av Long og Magerko (2020). Tabell 3 og 4 inneholder beskrivelser av kodene og eksempler på tilhørende data. Når kodingsprosessen var ferdig, hadde jeg samlet alle dataene som bidro til en forståelse av hvert av disse konstruktene. Deretter leste jeg jeg gjennom dataene flere ganger for å bli kjent med dem. Analysing av de kodede dataene innebærer også en meningsfortetning. Hensikten med meningsfortetning er å trekke ut meningsinnholdet i en tekst og gjengi det med få ord (Kvale & Brinkmann, 2015, s. 232).

Kode	Beskrivelse	Eksempel
Intensjon om å bruke	I hvilken grad en person har tenkt til å bruke teknologien i fremtiden.	Pia: Jeg tror jeg kommer til å bruke det ganske mye. Mer og mer jo mer den utvikler seg, og jo mer man kan bruke den til.
Oppfattet nytteverdi	I hvilken grad en person oppfatter at teknologien bidrar til bedre og å effektivisere arbeidet.	Lars: Spørsmål med refleksjon sliter den ofte med. Du må ha konkrete fakta og spørsmål. Pia: Jeg synes den er veldig god å kunne forklare. Hvis du stiller de riktige spørsmålene, så kan det hjelpe til med å få en bedre forståelse innenfor temaene.
Oppfattet brukervennlighet	I hvilken grad en person oppfatter teknologien som enkel å bruke, uten anstrengelser.	Pia: Det er jo veldig lett å bare spørre den MyAI på Snapchat, hvis det er noe man lurer på, og få et svar. Og dermed slippe å gå på Google, og lete etter et svar.
Troverdighet	I hvilken grad en person	Mona: Du vet ikke om det alltid er riktig, men

	oppfatter at teknologien er objektiv og sannferdig.	du kommer i hvert fall nærmere å finne ut av noe, enn om du bare tenker på det. Jens: Jeg stoler på det når jeg spør om vitenskapelige spørsmål. Sånn hvorfor det blir til vann, eller is blir til vann. Sånne ting. Da stoler jeg på det.
Ekspertise	I hvilken grad en person oppfatter at teknologien har ekspertise og kompetanse.	Alina: Det skal jo være riktig da, for det er jo et program.
Kontroll	I hvilken grad en person oppfatter å ha kontroll over interaksjonen med teknologien, om den responderer slik en ønsker.	Ida: [Sier noe om at hun bruker MyAI mest i samfunnsfag og KRLE] for jeg prøvde å stille et mattestykke. Men så ga den liksom en forklaring på hva matte er.
Nysgjerrighet	I hvilken grad teknologien er i stand til å stimulere en persons nysgjerrighet.	Jens: Ja, det er vel litt det at vi drev og snakket om hvordan man kunne gjort det selv i så fall. For det å fryse vann ved å slå på en flaske, det hadde vært dritkult å gjøre selv.
Fokusert innlevelse	I hvilken grad en person har oppmerksomheten rettet mot oppgaven, og til hvilken grad forstyrrelser blir	Lise: Appen ville ikke fungere. Det var veldig irriterende. Så mistet vi alt vi hadde søkt på når vi gikk inn og ut. Pia: Jeg er litt motsatt av det. Jeg tenker at om man har en oppgave, og det er en del

	ignorert.	man ikke gidder å bruke lang tid på å lete etter informasjon for. Så kan man heller bruke for eksempel ChatGPT for å få hjelp fort, få svar på et spørsmål fort, så slipper man å bruke lang tid på det som kan være litt kjedelig. Da slipper man å miste fokus og miste motivasjonen til å fortsette med oppgaven. Hvis det går fortere og de kjedelige delene ikke blir like kjedelige og like langtekkelige.
Glede	I hvilken grad en person oppfatter at teknologien er morsom å bruke og hvorvidt en føler glede og tilfredsstillelse.	Ida: Men den hadde emoji'er. Alina: Ja, det var litt lættis. Mona: Ja det var gøy. Ida: Ja, den la liksom på litt emoji'er og utropstegn, det var gøy. Det var litt bedre enn å bare få en kjedelig nettside opp.
Tidsdissosiasjon	I hvilken grad en person mister evnen til å oppfatte tidsforløpet når en bruker teknologien.	Henrik: Jeg pleier ikke å sitte med den så ofte. Det er jo kort tid om gangen. Veldig ofte enten i friminutt eller i timen. Eller noen ganger hjemme når jeg ikke har noe å gjøre.

Tabell 3: Beskrivelse av kodene med eksempler fra datamaterialet.

Kode	Beskrivelse	Eksempel
Hva er KI	Elevene gir beskrivelser av hva KI er, hva som er	Lars: Det blir spennende å se om vi faktisk kan lage ordentlig kunstig intelligens, og ikke

	og ikke er KI.	en database som gjetter.
Hva kan KI gjøre	Elevene gir beskrivelser av hva KI kan gjøre.	Rikke: Men den klarer jo ikke å diskutere.
Hvordan virker KI	Elevene gir beskrivelser av hvordan KI virker.	Henrik: Den er jo ikke bevist, den kan ikke se ting som skjer rundt den.
Hvordan bør KI brukes	Elevene gir beskrivelser av hvordan KI bør eller ikke bør brukes. Inkludert etiske betraktninger.	Pia: Hvis det finnes mer informasjon på internett som støtter en side av en sak, vil det være mer sannsynlig at ChatGPT gjentar denne sidens argumenter

Tabell 4: Beskrivelse av kodene med eksempler fra datamaterialet.

3.6. Analyse av spørreskjema

Dataene fra spørreskjemaet ble importert til SPSS. Som følge av at undervisningen i de to klassene ble forskjellig, gjennomførte jeg Levens test og T-test for uavhengige utvalg. Levens test viste at det var lik varians. Resultatene fra T-testen viste at det ikke var noen signifikante forskjeller mellom de to klassenes responser på noen av punktene i spørreskjemaet. For eksempel viste T-testen for det første spørsmålet om *oppfattet nytteverdi*, en t-verdi på 0,76 (df = 37), noe som gir en p-verdi på 0,45. Forskjellen er dermed ikke signifikant. På bakgrunn av dette slo jeg da sammen de to datasettene. Cronbachs alfa ble anvendt for å evaluere den indre reliabiliteten til hvert av rammeverkets overordnede faktorer (Taber, 2018). Resultatene vises i tabell 5. Verdiene varierer mellom 0,64 og 0,83, noe som tyder på at de korresponderende spørsmålene i stor grad måler det samme. Verdien for oppfattet nytteverdi er noe lavere enn den mye brukte grensen på 0,7, men resultatet må vurderes med hensyn til hva en måler (Taber, 2018). Her vurderer jeg den indre reliabiliteten som tilstrekkelig for denne studien, gitt at de ulike elementene naturligvis vil ha en viss variasjon.

For hvert av konstruktene ble det opprettet en ny variabel, for gjennomsnittet av de konstruktets tilhørende variabler. Resultatene ble presentert som deskriptiv statistikk, for elevenes gjennomsnittlige svar på hvert av konstruktene, med tilhørende standardavvik. På grunn av det

begrensede datagrunnlaget ble det brukt Pearsons korrelasjonstest. Korrelasjonstesten ble brukt for å måle hvordan *intensjon om å bruke* og *oppfattet nytteverdi*, korrelerer med hverandre og resten av rammeverket.

	Cronbachs alfa
Intensjon om å bruke	0.76
Oppfattet nytteverdi	0,64
Oppfattet brukervennlighet	0,81
Kredibilitet	0,71
Kognitiv absorpsjon	0,83

Tabell 5: Viser mål på den indre reliabiliteten til hvert av rammeverkets overordnede faktorer

3.7. Kvalitetsvurderinger

Å vurdere troverdighet i kvalitative studier er komplisert, og det er ingen entydige retningslinjer. Det som er viktig er å gjøre rede for konteksten og valg tatt underveis i prosessen, slik at andre kan ta stilling til hvilken overføringsverdi resultatene har. Jeg har prøvd å tydeliggjøre dette gjennomgående i oppgaven. Jeg vil også komme med mine refleksjoner rundt hvilken påvirkning rekkefølgen på spørreskjemaet og intervjuene, samt forskjellene i undervisningen og utvelgelsen til fokusgrupper, kan ha hatt på studiens resultater. Jeg vil benytte meg av begrepene reliabilitet og validitet, men det må forstås at disse har en annen form i kvalitativ forskning enn det de har i kvantitativ forskning (Kvale & Brinkmann, 2015, s. 275-276).

3.7.1. Reliabilitet

Reliabilitet i kvalitativ forskning går ut på hvorvidt studien kan etterprøves og reproduseres av andre på et annet tidspunkt (Kvale & Brinkmann, 2015, s. 276). Gjennomføring av intervjuene er av stor betydning for studiens reliabilitet (Kvale & Brinkmann, 2015, s. 276). Jeg har prøvd å gi en detaljert beskrivelse av konteksten, selve intervjuprosessen, og vedlagt intervjuguiden.

Fokusgruppeintervjuer blir i stor grad formet av deltakeren, noe som naturligvis gjør intervjuene forskjellige, dette er imidlertid en av fordelene med metoden (Kvale & Brinkmann, 2015). Jeg har prøvd å gjennomføre intervjuene på en slik måte at om andre hadde intervjuet de samme elevene i den samme konteksten, ville de fått like data. Intervjuene startet med åpne spørsmål om elevenes erfaringer, disse ledet til en samtale som i stor grad ble styrt av elevene. Underveis stilte

jeg spørsmål for å oppklare eller for å be om andre elevers synspunkter. Elevenes samtaler dekket i stor grad alle punktene i intervjuguiden, men jeg hadde til tider behov for å prøve å flytte samtaleretningen gjennom spørsmål. For å holde det nært til elevenes verden, brukte jeg ting elevene hadde nevnt tidligere som kontekst for mine spørsmål. Noe som understreker det at intervjuer vil være unike fordi det er interaksjoner mellom mennesker (Cohen et al., 2018).

En detaljert beskrivelse av bearbeidingen av dataene er også av stor betydning for studiens reliabilitet. Jeg har beskrevet hvordan både transkripsjonene og analysene er gjennomført, slik at andre i størst mulig grad vil kunne gjennomføre dette på en tilsvarende måte. Imidlertid vil ens egen forforståelse alltid påvirke hvordan dataene forstås (Cohen et al., 2018).

3.7.2. Validitet

Validitet handler overordnet om hvorvidt studien måler det den skal måle, og kan deles inn i intern og ekstern validitet (Cohen et al., 2018). Studiens forskningsspørsmål har blitt revidert underveis i prosessen, for å imøtekomme tekniske utfordringer og for at de skal samsvare med de faktiske dataene fra fokusgruppeintervjuene. Jeg vil først reflektere rundt studiens interne validitet, altså hvorvidt mine konklusjoner er gyldige avledninger fra mitt datamateriale. Deretter vil jeg reflektere rundt studiens ytre validitet, altså i hvilken grad mine funn har overføringsverdi til andre situasjoner.

Hsieh og Shannon (2005, s. 1283) trekker frem at gjennom en teoridrevet innholdsanalyse møter en dataene med en sterkt bias. Den teoretiske forståelsen som ligger til grunn, kan gjøre det vanskelig å oppdage andre faktorer (Fauskanger & Mosvold, 2014). Det kan tenkes at dersom jeg hadde benyttet med av TAM3 (Venkatesh & Bala, 2008) hadde elevenes datakyndighet og sosial innflytelse hatt stor forklaringsverdi.

Metodetriangulering har gjort det mulig å studere elevene fra ulike perspektiver, noe som bidrar til et mer fullstendig bilde av det som studeres (Cohen et al., 2018). Intervjuene har vært studiens primære data, men spørreskjemaene har kunnet utfylle ved at de undersøker hele elevgruppen. Der intervjuene har bidratt med dybde og forklaringer, har spørreskjema i større grad kunne si noe styrkeforhold. Dataene fra intervjuene og spørreskjema vil kunne både utvide og kontrollere hverandre. Selv om datagrunnlaget til spørreskjema (n = 39) var vesentlig høyere enn intervjuene (n = 16), er det få respondenter og det begrenset hvilke konklusjoner som kan trekkes.

Konstruktvaliditeten er også lavere enn tidligere studier (Davis, 1989), noe som begrenser validiteten av disse dataene.

Denne studien er ikke egnet til å komme med konklusjoner om elever generelt, den sier først og fremst noe om de elevene jeg har undersøkt på det tidspunktet jeg har undersøkt dem på. Imidlertid kan funn fra kvalitativ forskning ha overførbarhet. Gjennom detaljerte beskrivelser av kontekst og funn, kan leseren vurdere hvorvidt det er generaliserbart til en annen gitt kontekst (Cohen et al., 2018). Eksempelvis kan de tykke beskrivelsene av elevenes bruk av språkmodeller potensielt ha overføringsverdi, og bidra til en forståelse av bruken, samt hvilke misoppfatninger en bør sørge for at elevene unngår.

3.7.3. Refleksjoner rundt valg jeg har tatt

I forkant av at elevene skulle bruke Copilot gjennomførte jeg en kort undervisningsøkt om teknologien og hvordan denne kan brukes. Jeg anså dette som viktig, fordi jeg i neste økt skulle få elevene til å ta i bruk teknologien. På bakgrunn av tekniske utfordringer med Copilot og det faktumet at elevene la stor vekt på tidligere erfaringer, ble denne økten en mindre del av studien og elevenes forståelse av teknologien ble viktigere. Det er dermed viktig å merke seg at den forståelsen av teknologien elevene ga uttrykk for, kan ha blitt påvirket av min undervisningsøkt om teknologien.

Klasse to gjennomførte ikke undervisningsopplegget med Copilot på samme måte som klasse en, noe som potensielt påvirket intervjuene. Imidlertid diskuterte elevene fra klasse en primært erfaringer de hadde fra tidligere, noe som begrenser denne effekten. Det førte til ulike måter å velge ut fokusgrupper, noe som resulterte i ulike skillelinjer mellomfokusgruppene i de to klassene. I klasse en var det liten forskjell mellom gruppene, mens i klasse to var det store forskjeller. Fokusgruppe fire hadde vesentlig høyere forståelse for KI enn de andre fokusgruppene, men fremstod også som mer reflekterte og motiverte generelt. De var mer aktive under tavleundervisningen, noe som kan ha forsterket forskjellene. Jeg fikk inntrykk av at elevene i fokusgruppe tre ønsket å delta til dels fordi de dermed slapp ordinær undervisning, noe som kan ha påvirket deres engasjement og ønske om å fremstå reflekterte.

Spørreskjemaet ble gitt før intervjuene av praktiske årsaker. Det er mulig at spørreskjemaets påstander kan ha påvirket intervjuene. Bruken av både positive og negative påstander bidro

imidlertid til at spørreskjemaet var nøytralt. Det er også verdt å merke seg at rekkefølgen av disse uansett ville ha kunnet påvirke hverandre. Ved å ha spørreskjemaet først, sikret jeg at elevenes svar var basert på deres egne erfaringer og ikke påvirket av diskusjoner under intervjuene.

3.8. Etikk

Thagaard (2018) trekker frem informert samtykke, konfidensialitet og konsekvenser av å delta i forskningen som viktige etiske betraktninger for kvalitativ forskning. Jeg vil i dette kapittelet gjøre rede for mine vurderinger knyttet til disse. I lys av at jeg skulle intervjuere elever kreves det godkjenning av Sikt før datainnsamlingen starter. Prosjektet ble dermed meldt inn og godkjent av Sikt (vedlegg 4). På bakgrunn av studiens omfang og elevenes alder, godkjente Sikt at elevene hadde samtykkekompetanse og dermed selv kunne skrive under på samtykkeskjemaet.

Elevene og de foresatte ble i forkant av datainnsamlingen informert om prosjektet av lærerne. I starten av den første timen med elevene introduserte jeg meg selv, og beskrev forskningsprosjektets mål og omfang. Deretter ble elevene informert om at det var helt frivillig for dem å delta og at de hadde rett til å trekke seg. Jeg gjennomgikk hva deltakelse ville innebære for dem, og delte ut samtykkeerklæringer (Vedlegg 5). Elevene fikk da mulighet til å stille spørsmål til både meg og deres lærer. Underveis i prosjektet ble elevene igjen informert om retten til å trekke seg uten konsekvenser. Før intervjuene startet, informerte jeg igjen om formålet med prosjektet, retten til å trekke seg, lydopptakeren, anonymitet og jeg åpent for spørsmål. Etter intervjuene fikk elevene igjen muligheten til å stille spørsmål.

I samtykkeerklæringen elevene skrev under på, blir det beskrevet at studien skal omhandle bruk av Copilot og deres erfaring med bruk av Copilot i en undervisningsøkt. Allerede fra starten av første intervju la elevene svært mye vekt på erfaringer deres erfaringer med ChatGPT og MyAI, utenfor denne økten. Dette førte til en endring av studiens omfang. Som en følge av dette er det viktig å ikke kun forholde seg til at elevene signerte på samtykkeerklæringen, men vurdere dette under hele prosessen (Thagaard, 2018).

For å sikre elevenes anonymitet ble intervjuene som nevnt transkribert til bokmål, elevene fikk fiktive navn og navn på ting som idrettsklubben de tilhørte ble fjernet. For å sikre elevenes personvern i forbindelse med at de brukte Copilot, ble dette grundig avtalt med skolen og deres retningslinjer ble fulgt. På tross av dette var realiteten at flere av elevene Copilot på en måte som

ikke sikret deres personvern, og en elev valgte å bruke ChatGPT. Skolen og kommunen ble derfor informert om at retningslinjene ikke var tilstrekkelige for å sikre elevenes personvern.

Ved å ta i bruk språkmodeller i undervisningen, spesielt som verktøy for å undersøke et fenomen, kan dette være med på å legitimere språkmodellene som kunnskapskilder. Undervisningsøkten om teknologien bak språkmodeller var delvis ment for at elevene skulle bli bedre på å ta verktøyene i bruk på en effektiv måte, men også for å gi de en bedre forståelse av teknologien og dens styrker og svakheter. Gjennom undervisningsøktene der elevene brukte Copilot selv eller som helklasse, ble elevene oppfordret til å bruke andre kilder for å vurdere om svarene var sanne, og gjennom observasjonen leste jeg samtaler med Copilot og pratet med elevene. På denne måten ble det flere ganger vist at modellene kunne gi feilaktige svar. Deltakerne i fokusgruppene fikk også muligheten til å kritisk reflektere sammen rundt teknologiens nytteverdi og bruksområder.

4. Resultat

Målet med denne studien var å utvide kunnskapen om elevers bruk av språkmodeller i naturfag. Jeg vil i dette kapittelet redegjøre for resultatene fra analysen av spørreskjemaet og fokusgruppeintervjuene. Jeg vil først redegjøre for resultatene av spørreskjema. Deretter vil jeg redegjøre for analysen av fokusgruppeintervjuene, disse vil struktureres etter studiens forskningsspørsmål.

4.1. Spørreskjema

De samlede resultatene fra spørreskjemaet presenteres her i to tabeller. Spørreskjemaet hadde som nevnt hensikt å måle elevenes erfaringer knyttet til de ulike delene av studiens rammeverk. Disse ble målt ved at elevene tok stilling til utsagn med en femtrinns Likerts-skala. Elevenes holdninger blir dermed representert med verdier fra 1-5, hvor 3 er nøytralt. Verdiene for kredibilitet og kognitiv absorpsjon er utregnet ved gjennomsnittet av de tilhørende variablene. Tabell 6 viser at elevene forholder seg relativt nøytrale, med liten tendens til å uttrykke sterk enighet eller uenighet. Unntakene er oppfattet brukervennlighet og glede, med verdier på henholdsvis 3,86 og 3,49. Oppfattet brukervennlighet har en verdi på nesten 4, som vil si at elevene i gjennomsnitt oppfatter språkmodellene som enkle å bruke. Imidlertid viser standardavviket at det er ulike oppfatninger.

	Gjennomsnitt	SD
Intensjon om å bruke	3,18	1,05
Oppfattet nytteverdi	3,00	0,75
Oppfattet brukervennlighet	3,86	0,99
Kredibilitet	3,11	0,68
Troverdighet	2,99	0,75
Ekspertise	3,21	0,75
Kognitiv absorpsjon	3,10	0,61
Kontroll	2,99	1,08
Nysgjerrighet	3,00	0,71
Fokusert innlevelse	3,11	0,72
Glede	3,49	0,85
Tidsdissosiasjon	2,97	0,79

Tabell 6: Gjennomsnitt og standardavvik (SD) i elevens svar på spørreskjemaets påstander, basert på en fem-trinns Likerts-skala, hvor 1 er svært uenig, 3 er nøytral og 5 er svært enig.

	Intensjon om å bruke	Oppfattet nytteverdi	Oppfattet brukervennlighet	Kredibilitet
Oppfattet nytteverdi	0,56			
Oppfattet brukervennlighet	0,35*	0,46		
Kredibilitet	0,47	0,72	0,51	
Kognitiv absorpsjon	0,67	0,73	0,51	0,65

Tabell 7: viser resultatene fra Pearsons korrelasjonstest. Alle korrelasjoner er signifikante ($p < 0,01$), med unntak av * ($p < 0,05$).

For å analysere korrelasjoner mellom ulike deler av rammeverket, ble disse utregnet med Pearsons korrelasjonstest. Tabell 7 viser korrelasjoner mellom de overordnede delene av rammeverket, alle verdiene møter kravet til signifikans ($p < 0,05$). Det er viktig å påpeke at korrelasjonene ikke viser kausale sammenhenger, kun samvariasjon. Resultatene viser at det er lineære positive korrelasjoner mellom alle variablene. Funnene viser en moderat korrelasjon mellom *oppfattet nytteverdi* og *intensjon om å bruke* (0,56). Imidlertid er korrelasjonen sterkere for *kognitiv absorpsjon* (0,67). *Oppfattet nytteverdi* korrelerer sterkest med *kredibilitet* (0,72) og *kognitiv absorpsjon* (0,73).

4.2. På hvilken måte har et utvalg elever på 10. trinn tatt i bruk språkmodeller som et verktøy i naturfag?

Elevene sier at de har tatt teknologien i bruk, uten at lærerne har bedt dem gjøre det. De bruker for det meste ChatGPT og MyAI, noen har også brukt Copilot. Mange av elevene sier de har brukt flere av verktøyene. Noen elever har utelukkende brukt MyAI, et par elever begrunner dette med at de ikke får tilgang til ChatGPT på sin skole-PC. Jeg vil først beskrive de mest omtalte bruksområdene jukse og søkemotor, deretter vil jeg kort beskrive andre bruksområder enkelte elever nevner.

Nesten alle elevene sier de i noen grad bruker språkmodellene til å gjøre skolearbeidet for dem. Elevene sier de sparer mye tid ved å bruke språkmodeller, og Alina kaller ChatGPT «*en jukse-maskin*». Et par elever forteller at de gjør dette på en måte som likner Maris: «*Kopierer inn spørsmålet, putter det inn i MyAI, og svaret kopierer jeg inn igjen.*». Imidlertid sier flertallet at de ikke kan bruke oppgaveteksten eller svarene direkte. Hvor mye arbeid de selv må gjøre varierer. I noen tilfeller holder det å redigere svarene, slik at det fremstår som eget arbeid. I de fleste tilfeller bryter elevene oppgaven ned i mindre deler som de ber om svar på. Dette medfører at det i elevenes forklaringer til tider blir en glidende overgang mellom å bruke det til de beskriver som juks, og det de beskriver som å søke etter informasjon.

Å bruke språkmodellene som søkemotorer er noe omtrent alle elevene beskriver at de gjør. Ida forteller: «*Jeg bruker det som om jeg hadde brukt Google, det er bare mye enklere*». De bruker det til å finne informasjon i forbindelse med skoleoppgaver, men også når de lurere på ting i hverdagen. Jacob har brukt ChatGPT til å lære om bilmotorer. Dette gjorde han ved å innlede en samtale slik: «*Jeg sa, hvordan kan man bygge en V12 motor?*». Henrik bruker det for å finne informasjon om «*utbytte av forskjellige typer trening*». Alina forteller: «*Her om dagen gikk jeg tur med hunden min, også spiste den en plante. Da spurte jeg MyAI om planten var farlig*». Noen av elevene bruker språkmodeller på andre måter når de skal finne informasjon. Lars bruker det som et verktøy for å søke på Google: «*Jeg bruker det for å få en grunnforståelse, slik at jeg vet hva jeg bør Google*». Pia bruker det til å forenkle informasjon hun har funnet: «*Hvis man finner en side man synes er litt vanskelig å skjønne, synes jeg det kan være veldig fint å spørre, for eksempel ChatGPT, om å forklare teksten på en enklere måte*».

Noen elever nevner også andre bruksområder. Pia bruker ChatGPT til å få tilbakemeldinger underveis i en skriveprosess: «*Hvis man gir ChatGPT en tekst eller rapport man har skrevet, kan man spørre ChatGPT hva synes du om denne, eller kan jeg få en tilbakemelding på denne teksten. Jeg synes ofte det kan fungere ganske bra.*». I oppstarten av en skriveprosess kan språkmodellene bli brukt for idemyldring, for å gi forslag til hvilke temaer som bør dekkes og hvordan teksten bør struktureres.

4.3. Hvordan påvirker oppfatninger av språkmodellenes nytteverdi, brukervennlighet, kredibilitet og opplevelsen av absorpsjon, elevers bruk av språkmodeller?

Resultatene vil her bli presenter etter oppbyggingen i studiens rammeverk. Dette er for å kunne gi tydelige beskrivelser av alle delene av rammeverket og hvilken innvirkning det har på elevenes bruk. Jeg vil først beskrive oppfattet nytteverdi og brukervennlighet, deretter kredibilitet og avslutningsvis kognitiv absorpsjon.

4.3.1. Oppfattet nytteverdi og brukervennlighet

Elevene legger stor vekt på at verktøyene er veldig nyttige og enkle å bruke. Det er som nevnt ulike oppfatninger om hvor nyttig språkmodellene er til å jukse, men det er allikevel stor enighet om at det ikke finnes noe bedre verktøy. Ulike oppfatninger om nytteverdi ser ut til å først og fremst påvirke hvordan de bruker verktøyene, for noen er det klipp og lim mens andre sier de må gjøre mer av arbeidet selv. Elevene sier at de bruker språkmodellene som søkemotorer fordi, de på veldig kort tid gir ett presist svar på akkurat det de spurte om. Mange av elevene beskriver dette i kontrast til Google, slik som Mona: «*Det [Google] tar mye lengre tid. Du må søke opp, så må du finne en kilde som er troverdig, så må du begynne å lete i en lang tekst, for å finne den lille setningen som gir deg svaret.*». Lise sier: «*Jeg synes også at det er lettere enn å søke vanlig, for [på Google] da får man bare mange kilder, også sier de forskjellige ting. Man vet aldri hva som er riktig*». Elevene sier også at svarene fra språkmodellene har et språk som er mer forståelig, og om det er for vanskelig kan de be om at det forenkles.

Flere av elevene sier at de bruker MyAI oftere enn ChatGPT, fordi den er lett tilgjengelig inne på Snapchat. Elevene i fokusgruppe en diskuterer forskjellen i nytteverdi og brukervennlighet mellom ChatGPT og MyAI. Alina sier: «*Den Snapchat-AI bruker jeg ofte. Sånn plutselig. Når man er på Snapchat hele tiden, trenger man bare å trykke på den*». Elevene i fokusgruppen sier

de bruker den mer til tross for at responsene er av lavere faglig kvalitet, at den er dårligere til å henge med i samtalen og ofte glemmer hele samtalen. Alina kaller MyAI «*et litt noob program*» mens «*ChatGPT er mer seriøs*». Allikevel nevner de at de spør MyAI om bruk av medisiner, sykdom, allergener i mat og hvorvidt en plante er giftig eller ikke. Det er først når det kommer til skoleoppgaver at disse elevene velger å legge inn den ekstra innsatsen det er å gå inn på PC-en for å finne frem ChatGPT.

Det er enkelte områder hvor elevene oppfatter at språkmodellene ikke alltid er nyttige. Her blir det trukket frem at de kan være dårlige på å gi oversikt over store temaer, noe som fører til at elevene spør om spesifikke deler av et tema istedenfor. Det er delte meninger om språkmodellenes evne til å drøfte, og enkelte sier derfor at de ikke bruker språkmodellene når de trenger å få frem ulike sider av en sak. Pia er skeptisk, men hun har imidlertid funnet en metode som fungerer. Ved å be om motargumenter i en samtale og støttende argumenter i en annen, kan hun bruke språkmodellene til å utforske forskjellige synspunkter og selv veie argumentene mot hverandre. Lars opplever at han får litt lavere karakter når han bruker språkmodeller til å forberede seg til vurderinger, og han prefererer å forberede seg på andre måter. Til tross for disse utfordringene, beskriver noen av elevene at tidsbesparelse fungerer som en modererende effekt på nytteverdien. Dette kommer til uttrykk når de opplever tidspress i forkant av prøver eller innleveringer, men også når de ønsker å bruke tid på andre fritidsaktiviteter.

4.3.2. Kredibilitet

Elevene har ulike meninger om teknologiens kredibilitet, og det påvirker elevenes bruk ulikt. For å beskrive hvilken effekt det har på elevenes bruk av teknologien vil jeg ta utgangspunkt i tre ulike oppfatninger av teknologiens kredibilitet. Disse beskrives i rekkefølgen, høy kredibilitet, moderat kredibilitet, lav kredibilitet.

Et mindretall av elevene uttrykker at teknologiens høye kredibilitet er en av grunnene til at de bruker den. Disse elevene forteller i større grad at de leverer språkmodellenes svar som eget arbeid, og de bruker språkmodellene som søkemotorer ukritisk. Alina har høy tillit: «*Det skal jo være riktig da, for det er jo et program.*». Når hun blir møtt med det at modellene kan gjøre feil begrunner hun det med: «*Da har den misforstått spørsmålet. Det er jo en robot, eller et program. Det er ikke et menneske som kødder med deg.*». Hun sier videre: «*Folk stoler sikkert ikke så mye*

på det, siden det er nytt og sånn.». Mari mener MyAI er veldig troverdig fordi *«den søker jo, eller har informasjon fra mange kilder, også samler den alt til ett svar.».* Lise mener det er bedre å spørre språkmodeller enn Google fordi når man søker på Google *«får man bare mange kilder, også sier de forskjellige ting. Man vet aldri hva som er riktig.».*

Majoriteten av elevene uttrykker at teknologien mangler kredibilitet, men sier de bruker den relativt ukritisk. Disse elevene bruker språkmodellene til det de kaller å «jukse» mindre, og de forteller at de i større grad må bearbeide oppgavetekst og responser. De bruker teknologien i større grad som søkemotor og uttrykker at svarene bør sjekkes opp mot andre kilder, men at de sjeldent gjør det. Mona sier: *«Du vet ikke om det alltid er riktig, men du kommer i hvert fall nærmere å finne ut av noe, enn om du bare tenker på det.».* Hun forklarer dette med at man *«tenker kanskje litt annerledes når man får innspilt av andre».* Hun sier videre at hun *«sjekker det opp med internett innimellom»*, og *«hvis jeg virkelig vil vite svaret på noe, så sjekker jeg.».* Noen elever uttrykker at behovet for å etterprøve svarene er mindre om de stiller naturfaglige spørsmål, enn spørsmål om følelser og samfunn. Det blir forklart slik av Jens *«Vitenskap er jo forsket på, det er bevist.»* og Rikke forklarer det slik: *«På internett er det mye mer konkrete svar på hva metall er. Det er mye større likheter på alle nettsider den finner. Så da er svaret sikkert mer sant.».*

Noen få elever uttrykker at de har lav tillit til teknologien, som de begrunner i oppfatninger om hvordan teknologien virker og egne erfaringer. Disse elevene uttrykker at de bruker språkmodeller i mindre grad og på andre måter enn resten av elevene. Lars begrunner at han ikke bruker MyAI på bakgrunn av erfaring: *«Jeg så med en gang at det den sa var helt feil. Om jeg prøvde å skrive noe av det ned fikk jeg helt hodepine, for det den sa ga ikke mening i det hele tatt.».* Lars bruker derimot andre språkmodeller til å bistå han med Google-søk, fordi han ikke stoler på at de er sannferdige. Pia bruker de til å forenkle troverdig informasjon hun selv har funnet, hun beskriver fordelen slik: *«da bruker den informasjonen man har gitt den, istedenfor å bare hente informasjon som man ikke vet hvor kommer fra. Da er det også lettere å oppgi kildene til informasjonen.».* Elevene i fokusgruppe fire har opplevd at språkmodellene kan ta politisk eller ideologisk standpunkt, noe de forklarer ut fra deres forståelse av teknologien. Som en følge av dette mener de at man må være ekstra kritisk til temaer som bærekraftig utvikling og på områder hvor det kan være motsetninger mellom vitenskapelige og religiøse forklaringer. De er

også ekstra skeptiske til å bruke det på områder der de selv har lite forkunnskap, fordi det blir vanskeligere å oppdage feil.

4.3.3. Kognitiv absorpsjon

Kognitiv absorpsjon har blitt undersøkt gjennom fem ulike aspekter, kontroll, nysgjerrighet, fokusert innlevelse, glede og tidsdissosiasjon. Disse fem aspektene vil sammen kunne bidra til at en blir oppslukt i bruken av teknologien (Lowry et al., 2013), i dette tilfellet språkmodeller. Noen av elevene forteller om situasjoner hvor de snakker med språkmodellene kun for underholdning, eller for å utforske hva de kan og ikke kan gjøre. Disse beskrivelsene kan minne om en slik tilstand. Imidlertid er disse opplevelsene i liten grad knyttet til naturfag. Jeg vil derfor presentere elevenes beskrivelser av hvert av de fem aspektene individuelt.

Opplevelsen av å ha en høy grad av kontroll over responsene til språkmodellene ser ut til å være viktig i deres beskrivelser. Elevene bruker ofte Google som en kontrast når de beskriver språkmodellene. Når elevene Googler ønsker de at svaret skal komme tydelig frem i en boks over selve søkeresultatene (en funksjon som kalles *sammendrag av søkeresultater*, og består av et sammendrag av en nettside ment for å gi rask informasjon), men dette er ofte ikke tilfellet. Jens sier «*Kommer det ikke opp på boksen, gidder jeg ikke å søke mer.*». Ved bruk av språkmodeller opplever elevene derimot å være i kontroll over hvordan responsen blir, og de kan be om at svaret blir presentert på en enkel og presis måte. Elevene opplevde mindre kontroll over Copilot enn både ChatGPT og MyAI, og flere sa de i fremtiden derfor ikke ville bruke Copilot. Copilot hadde til vane å gjenta seg selv, endret i liten grad på forklaringer og svarte ikke på oppfølgingsspørsmål.

Kun et par elever uttrykte at interaksjoner med språkmodellene i seg selv stimulerer til faglig nysgjerrighet. Imidlertid beskrev Jacob at han undersøkte flere ting enn det han i utgangspunktet hadde tenkt til når han pratet med ChatGPT om bilmotorer. Jens undersøkte videoen jeg viste av en væske som på sekunder blir til et fast stoff. Han ble nysgjerrig på hvordan han selv kunne gjennomføre dette når han fant ut at det dreide seg om underkjølt vann. Han gir deretter uttrykk for at tilgangen på Copilot gjorde det enklere å skulle følge denne nysgjerrigheten. Jane beskriver også noe liknende «*Jeg blir mer interessert av ChatGPT. Da kan man få svar på det man er interessert i mye forttere, enn om du søker og leter. Da gidder man ikke.*». Resultatene gir ingen

sterk indikasjon på at språkmodellene i seg selv stimulerer til elevenes nysgjerrighet, men det kan tyde på at bidrar til at elevene i større grad følger opp nysgjerrigheten.

Språkmodellenes gir som oftest elevene et konkret og forståelig svar uten distraksjoner, i motsetning til Google. Noen beskriver at dette gjør det enklere å fokusere på skoleoppgaven. Et Google-søk gir mange ulike kilder som kan svært vanskelige å forstå eller de kan være irrelevante. Dette kan gjøre prosessen både vanskelig og kjedelig, noe som fort kan lede til prokrastinering. Lars som er skeptisk til å bruke språkmodellene som søkemotor anerkjenner allikevel denne fordelene i forbindelse med oppgaver som er vanskelige eller kjedelige, og beskriver utfordringen med Google slik: *«Sitter man og søker, kan man begynne å kjede seg veldig fort. Da kan man fort miste fokuset og motivasjonen til å fortsette å jobbe.»*

Språkmodellene derimot gir svar raskt, og Pia forklarer fordelene slik: *«Da slipper man å miste fokus og miste motivasjonen til å fortsette med oppgaven.»* Påvirkningen fra fokusert innlevelse på elevenes bruk kommer også frem på en annen måte. Noen av elevene uttrykte en misnøye mot både MyAI og Copilot fordi de glemte samtalen, noe som førte til at de måtte starte samtalen på nytt. De elevene som også hadde erfaring med ChatGPT uttrykte dermed en preferanse for ChatGPT på bakgrunn av dette.

Bruken av språkmodeller synes å være mer lystbetont for elevene, enn Google. Språkmodellenes ser ut til å ha et mer livlig og lesbart språk sammenliknet med elevenes søkeresultater, eksempelvis Copilots bruk av emoji'er. Språkmodellenes evne til å tilpasse språket og «personlighet» etter elevenes ønske ser ut til å i seg selv være gjenstand for nysgjerrighet og distraksjoner. Imidlertid gjør det også fagstoff mer engasjerende, slik Mona beskriver *«[en faglig forklaring fra språkmodell] blir litt morsommere, enn at du skal få vitenskapsmann på 70 [år]»*. Elevene gir imidlertid ikke uttrykk for at språkmodellen gir dem indre motivasjon til å undersøke naturfaglige temaer, de forklarer først og fremst at språkmodellene gjøre det de allerede må gjøre mer lystbetont. MyAIs tone gjør at den i større grad enn Copilot og ChatGPT fremstår som en venn, og dette får noen av elevene til å ville prate med den for underholdningsverdien. Det er uklart hvilken innflytelse dette har på elevenes bruk av MyAI til naturfaglige samtaler. Bruken av språkmodeller kan også gi elevene glede på andre måter. Både Dan og Pia uttrykker at det kan være gøy å samtale med ChatGPT om temaer de kan mye om, fordi det gøy å se *«hvor dårlig den er»* og *«hvor merkelig svaret den kommer opp med er»*.

Selv om språkmodellene kan holde på oppmerksomheten, uttrykker de ikke at de mister evnene til å oppfatte tidsforløpet i arbeid med naturfag. De lengre og utforskende samtaler handler i liten grad om naturfaglige temaer. Der det fremstår som at elevene i størst grad blir oppslukt er når de utforsker teknologiens muligheter, for eksempel om den kan svare på vanskelige spørsmål.

4.4. Hvilken forståelse av språkmodeller gir elevene uttrykk for i sine samtaler om bruken av denne nye teknologien?

Resultatene fra analysen av intervjuene ved bruke av Long og Magerko (2020). Disse må tolkes i lys av at elevene har ikke blitt bedt om å forklare hvordan språkmodeller virker, eller noe liknende. Det er den autentiske forståelsen elevene gir uttrykk for i samtaler om hva de mener språkmodellene kan og ikke kan brukes til.

Nesten alle elevene beskriver språkmodellene som bedre og automatiserte utgaver av Google, eksempelvis Jens: «*Det er jo bare Google som snakker*». De gir uttrykk for at språkmodellene har agens (evnen til å handle med hensikt og gjennomføre bevisste handlinger). Dette kommer til uttrykk ved at de forklarer at språkmodellene aktivt søker på nett for å finne informasjon som de deretter bruker til å svare på elevens spørsmål. «intelligensen» til språkmodellen ser ut til å først og fremst omhandle evnen til å vurdere og bruke informasjon på internett. Noen av elevene tenker at språkmodellene ukritisk finner en kilde på nett og gjentar dette til dem, andre tenker at informasjon fra flere kilder blir analysert og satt sammen. Enkelte går enda lenger og uttrykker at språkmodellene har tilgang til omtrent alt av informasjon på internett som de kan bruke for å gjøre avanserte analyser og utregninger. Det er korrekt at noen av språkmodellene kan gjennomføre søk på internett for å finne informasjon, men det er ikke slik majoriteten av responsene blir generert. I beskrivelsene er det tydelig at kunnskapen til språkmodellene kommer direkte fra internett, men det er ingen beskrivelser av hva som utgjør evnen til å forstå elevens spørsmål og gjennomføre et søk. Det er enkelte elever som oppfatter språkmodellene identiske som Apples Siri, men gir uttrykk for at de eger seg til litt ulike ting.

Noen av disse elevene sliter med å forstå hvorfor språkmodellene noen ganger gir et feilaktig svar, mens de andre ganger kan svare riktig på liknende spørsmål. Hans uttrykker for eksempel «*det var litt rart at den klarte å gi feil, også har den egentlig riktig svar. Da burde den si det først*». Til dette responderer Alina at språkmodellen har misforstått spørsmålet, som med en

pragmatisk tilnærming kan sies å være delvis sant. Hun, i likhet med de andre elevene, gir imidlertid ikke uttrykk for å forstå et sentralt poeng, det at språkmodellene først og fremst er trent til å fullføre tekst. En evne som kommer av lærte sammenhenger og strukturer i språk fra treningsdataene og menneskelig tilbakemelding. Det kan virke som at noen av elevene tror språkmodellene blir smartere når de søker på nett, noe som viser liten forståelse for at treningsdataene er noe som er gitt av mennesker og ikke noe språkmodellene selv finner. Menneskers påvirkning får ingen plass, og språkmodellene blir dermed i større grad omtalt som agenter.

I fokusgruppe fires beskrivelser kommer det frem et annet syn. De beskriver språkmodeller som i mindre grad har agens, og i større grad er påvirket av mennesker. De beskriver at informasjonen er gitt til modellen på forhånd, modellen selv søker altså ikke etter informasjonen på nett. Likevel viser de ikke en forståelse for at modellen blir trent på tekst, og at treningsdataene er utilgjengelige for modellen. De uttrykker at modellen velger en tekst og baserer seg på den, Lars kaller det «*en oppskrytt database*». De viser også en forståelse av at modellene først og fremst er til for å fullføre tekst, og sammenlikner det med tekstforslagene som gis på mobiltefontastaturet. Dette blir brukt til å forklare at modellene kommer med ukorrekte påstander, og at de kan lures til å si ting som « $1+1=3$ ». De viser en forståelse for at datagrunnlaget kan være av dårlig kvalitet, noe som kan føre til skjevheter og misinformasjon. For eksempel uttrykker Pia: «*Hvis det finnes mer informasjon på internett som støtter en side av en sak, vil det være mer sannsynlig at ChatGPT gjentar denne sidens argumenter.*». De trekker også inn at modellen blir påvirket av utviklerne, og det faktum at modellene er laget av amerikanske selskap. Videre problematiserer de at modellene blir påvirket av brukere. De mener modellene blir påvirket ved at brukerne vurderer språkmodellens svar ved å trykke på enten tommel opp eller ned, samt at modellen lærer av samtalene med brukeren. Elevene som viste størst forståelse av modellene var alle blant de mest kritiske til modellenes kredibilitet.

5. Diskusjon

Hensikten med oppgaven er å bidra med kunnskap om elevers bruk av språkmodeller som et verktøy i naturfag, noe som er viktig for at man kan treffe egnede tiltak for å utnytte teknologiens potensiale og begrense de negative effektene. Funnene mine viser at elevene har liten forståelse

for hvordan teknologien virker og at de bruker den på en snever måte. Primært brukes det til å jukse og som en søkemotor. For de fleste elevene er språkmodeller en ny og bedre Google, den er enklere og mer behagelig å bruke, fordi de kjapt får et konkret og forståelig svar. Det virker som mesteparten av elevene i praksis stoler på språkmodellene i stor grad, selv om de uttrykker at en ikke bør gjøre det.

5.1. Elevenes bruk av språkmodellene

Farrokhnia et al. (2023) argumenterer for at en må utnytte de mulighetene som ligger i bruken av språkmodellene, samtidig som en begrenser de negative effektene. Funnene mine tyder derimot på at elevene i liten grad utnytter mulighetene, og at de bruker språkmodellene på måter som kan ha store negative effekter.

Elevene benytter språkmodeller som søkemotorer fordi disse raskt gir enkle svar, i motsetning til Google som gir mange komplekse kilder som krever mye innsats å vurdere og tolke. Selv om språkmodellene som oftest gir gode resultater innenfor mange av naturfagets fagområder, er de ikke søkemotorer (Fijačko et al., 2023; Kortemeyer, 2023; Kung et al., 2023; OpenAI, 2023, s. 4-11). Språkmodellene har begrenset evne til å vurdere hvorvidt svarene er korrekte, men de kan generere svar som fremstår både overbevisende og korrekte (Farrokhnia et al., 2023, s. 46; Lo, 2023; OpenAI, 2023). Derfor er evnene til å kritisk evaluere KI-generert innhold en sentral del av KI-literacy (Long & Magerko, 2020). Mine funn kan tyde på at dette er ekstra utfordrende i naturfag, fordi flere av elevene har et syn på kunnskap i naturfag som noe sikkert og gitt. Dette kommer for eksempel til uttrykk i Jens sitt utsagn: «*Vitenskap er jo forsket på, det er bevist*». Et slikt kunnskapssyn svekker elevenes evner til kritisk tenking, og de vil i større grad stole på en ytre autoritet (Ferguson & Krange, 2020). Elevene vil bli da blir sårbare for språkmodellenes misinformasjon, skjevheter og fordommer (Farrokhnia et al., 2023). På den andre siden, så blir språkmodellenes evne til å raskt generere informasjon tilpasset et nivå som eleven kan forstå, løftet frem som en mulig styrke (Farrokhnia et al., 2023). Mine funn støtter at språkmodellene har denne evnen, men de viser også at elevene må være mer kritiske, noe som tilsier at vi behøver mer forskning på hvordan dette kan fungere i praksis.

Å ukritisk bruke språkmodellene som søkemotorer vil kunne svekke elevenes kritiske tenking og gjøre dem mindre opplyste. Duenger Bøhn (2024) mener at opplysning handler om å ta ansvar, en

må på selvstendig vis gjøre seg opp en mening etter å selv ha undersøkt. Derimot beskriver majoriteten at de i stor grad gjør det motsatte, de aksepterer en ytre autoritet som Bøhn kaller det. Fordi språkmodellene gjør det så enkelt å finne informasjon, vil elevenes motivasjon til å ta dette ansvaret kunne svekkes (Kasneci et al., 2023), effekten forsterkes trolig om elevene ser på kunnskap som noe sikkert og gitt av en autoritet (Ferguson & Krange, 2020). Dette vil igjen føre til at elevene i mindre grad utvikler evnen til kritisk tenking som beskrevet av Farrokhnia et al. (2023). Funnene underbygger dermed aktualiteten i naturfagslæreplanens mål om at elevene skal utvikle en forståelse for hva som kjennetegner naturvitenskapelig kunnskap, samt å forstå teknologi og hvilke problemer det kan medføre (Utdanningsdirektoratet, 2020). Opplysning og kritisk tenking er viktige demokratiske prinsipper og sentrale mål i skolen (Duenger Bøhn, 2024; Kunnskapsdepartementet, 2017). Behovet for at elever trenes i kritisk tenking slik at de selv kan ta stilling til kunnskap, vært viktig lenge før språkmodellene (Ferguson & Krange, 2020; NOU 2014: 7, 2014, s. 129). Likevel vil det nok være behov for å inkludere KI-literacy i arbeidet med kritisk tenking fremover.

Språkmodellene medfører nye problemer i forbindelse med skriveprosesser. Elevene tar i bruk språkmodeller på ulike måter i skriveprosessen, noen overlater hele prosessen til språkmodeller, andre bruker dem for å få tilbakemeldinger på egen tekst. Det er heldigvis de færreste som overlater hele prosessen til språkmodellene, men det kan til tider være utydelig hvor mye arbeid majoriteten av elevene gjør selv. Det kan være at det som beskrives som informasjonssøk og inspirasjon i virkeligheten er nærere plagiat enn det elevene gir uttrykk for.

Ungdomsskoleelevene jeg har undersøkt ser ut til å anse språkmodellene som mer troverdige og i større grad plagiere dem, enn VGS-elever (Bueie, 2024). Funnene til Bueie (2024, s. 302), indikerer at det er noe annet å ta i bruk språkmodeller for de som har bedre skriveferdigheter og mer fagkunnskap, noe mine funn kan se ut til å støtte. Imidlertid er det andre faktorer som også kan ha innvirkning, som det at Bueie (2024) undersøkte i norsk-faget, og at teknologien kan ha utviklet seg på det året som har gått mellom undersøkelsene. Funnene ser ut til å gi støtte til læreres opplevelser av at språkmodeller blir flittig brukt (LNU, 2023), men til dels nyansere bekymringene rundt plagiat. Språkmodellene kan gjøre det lett for elevene å plagiere uten å bli oppdaget (Farrokhnia et al., 2023; Lo, 2023), noe som kan gjøre det fristende å plagiere det som egentlig skulle være til inspirasjon eller søk etter informasjon. Duenger Bøhn (2024) frykter at elever i takt med språkmodellenes utvikling vil begynne å bruke dem i større grad og tidligere i

prosessen, noe som støttes av bekymringene til OpenAI (2023), om overdreven tillit og avhengighet. Dette medfører at elever også i selve skriveprosessen ender opp med å overlate arbeidet til en ytre autoritet (Duenger Bøhn, 2024). Dette vil være svært problematisk i forbindelse med å skrive eksempelvis forsøksrapporter, ettersom en viktig del av skriving i naturfag er å kunne «beskrive observasjoner og erfaringer» (Utdanningsdirektoratet, 2020).

Når tenkearbeidet blir overlatt til språkmodellene, vil ikke elevene utvikle evnen til høyere ordens tenking (Farrokhnia et al., 2023), og så lenge elevene aksepterer og er avhengige av en ytre autoritet, vil de heller ikke kunne bli dannede (Duenger Bøhn, 2024). Det finnes imidlertid måter å ta teknologien i bruk, uten å akseptere den som en ytre autoritet. Duenger Bøhn (2024) trekker frem at språkmodellene kan gi tilbakemeldinger på elevenes eget arbeid underveis i skriveprosessen, men anser det som usannsynlig. Imidlertid var det nettopp dette elevene som Bueie (2024) undersøkte, mente var mest nyttig, og Pia bekreftet dette. Andre trekker frem at KI-literacy og høyere ordens tenking kan utvikles gjennom å revidere tekster generert av språkmodeller, eller å argumentere med dem (Bitzenbauer, 2023; Farrokhnia et al., 2023). Det er dog grunn til å tro at elevene ikke utvikler nødvendig KI-literacy for å bruke teknologien på en kritisk måte, dersom skolen ikke aktivt arbeider med det (Bueie, 2024), da mine funn viser at språkmodellene først og fremst er «jukse-maskiner» og søkemotorer.

5.2. Hvordan påvirker elevenes bruk av språkmodellene

Jeg vil først drøfte funnene relatert til nytteverdi og brukervennlighet, for så å trekke inn *kredibilitet* og *kognitiv absorpsjon*. Sammen ser disse faktorene ut til å dekke mange ulike aspekter ved elevenes intensjon om å bruke, så vel som elevenes oppfatning av hva språkmodellene kan brukes til. Imidlertid vil det kunne være mange andre faktorer jeg ikke har fått undersøkt gjennom mitt rammeverk.

Teknologiakseptmodellen tar utgangspunkt i at *oppfattet nytteverdi* og *oppfattet brukervennlighet* er av størst betydning for brukeres *intensjon om å bruke* teknologi (Davis, 1989). I spørreskjemaet er korrelasjonen mellom disse henholdsvis moderat ($r = 0,56$, $p < .05$) og svak ($r = 0,35$, $p < .05$), noe som er lavere enn funnene til Davis (1989). Korrelasjonene er derimot høyere med *kognitiv absorpsjon*, og jeg vil senere drøfte dette forholdet. Tidligere forskning viser at nytteverdi og brukervennlighet er av stor betydning for bruk av språkmodeller (Albayati, 2024;

Camilleri, 2024; Duong et al., 2023; Liu & Ma, 2024). Analysen av intervjuene viser at majoriteten av elevene beskriver teknologien som en veldig enkel og brukervennlig søkemotor, og i varierende grad en nyttig «jukse-maskin». Det er dette de beskriver som årsaken til at de bruker teknologien i naturfag. Analysen av spørreskjema viser derimot at elevene inntar en mer nøytral holdning til nytteverdi. Forskjellen kan skyldes at de ulike kontekstene, spørreskjema og intervju, fanget ulike aspekter ved, og nyanser i, opplevd nytteverdi. Det er også mulig at elevene som ikke ønsket å delta på intervju generelt sett opplevde språkmodellene som mindre nyttige, og at fokusgruppene dermed ikke er representative for klassene. Spørreskjemaet var anonymt, derfor var det ikke mulig å sammenlikne fokusgruppens responser opp mot klassene, men det ville vært av interesse å kunne utforske hva denne forskjellen skyldtes. Samtidig kan det ikke utelukkes at forskjellene er påvirket av mine fokusområder og flyten i gruppeintervjuene, da disse primært handlet om hva elevene har brukt språkmodellene til, heller enn hva de *ikke* har brukt det til. Det kan tenkes at dette gjør at et syn på språkmodellene som nyttige, kommer best fram.

Analysen av intervjuene viser at synet på språkmodellenes kredibilitet blir påvirket av hvordan de tror teknologien virker, noe som igjen påvirker hva de bruker den til og hvor nyttig den er til ulike formål. Elevene som er minst kritiske ser ut til å oppfatte teknologien som mest nyttig, da den kan løse de fleste oppgaver omtrent helt av seg selv. Tidligere forskning viser at barn primært vurderer «intelligensen» til KI basert på prestasjoner, fremfor underliggende mekanismer (Long & Magerko, 2020, s. 8). Dette gjør det lett å vurdere *kredibiliteten* til språkmodeller som høy, og oppfattelse av høy kredibilitet ser ut til å henge sammen med stor *opplevd nytteverdi*. Samtidig er dette årsaksforholdet vanskelig å fastslå, da en opplevelse av at teknologien er nyttig, muligens også kan lede til en oppfattelse av høy *kredibilitet*. Analysen av spørreskjemaet viser at nytteverdi har en positiv korrelasjon med *kredibilitet* ($r = 0.72$, $p < .01$). På den andre siden, uttrykte også elevene som var mest kritiske at teknologien var nyttig, men til andre formål enn som søkemotor og «jukse-maskin». Disse elevene viste størst forståelse for hvordan teknologien virket, noe som ga dem mulighet til å dra nytte av den på tross av manglende *kredibilitet*. Pia brukte språkmodellene til å få tilbakemeldinger på egenprodusert tekst, istedenfor å få den til å generere tekst for henne. Hun viste bekymringer for skjevheter i modellens responser, men kunne allikevel bruke de til å utforske begge sidene av en sak, for så selv å gjøre seg opp en mening. Analysen av spørreskjemaet viser at synet på teknologiens *kredibilitet* har sterkere korrelasjoner med *oppfattet nytteverdi*, enn *intensjon om å bruke*. Dette kan indikere at synet på *kredibilitet* i større grad

påvirker hvordan elevene bruker språkmodeller, enn om de bruker språkmodeller. Imidlertid målte ikke spørreskjemaet ulike måter å bruke språkmodellene, så hvorvidt denne sammenhengen eksisterer i hele utvalget, utenfor fokusgruppene, er uvisst.

I en studie gjort av Ding et al. (2023) er det en signifikant forskjell i *oppfattet brukervennlighet* basert på nivåer av tillit, hvor de som hadde høyest tillit oppfattet ChatGPT som mest brukervennlig. Mine funn kan indikere en lignende sammenheng, da elevene som var minst kritiske i størst grad drev med klipp og lim, og de mest kritiske elevene brukte mer avanserte teknikker som innebar mer arbeid. Pia er et godt eksempel på det å bruke mer avanserte teknikker. Det kan imidlertid være andre årsaker bak forskjellen i *oppfattet brukervennlighet* i studien til Ding et al. (2023).

Kognitiv absorpsjon har i tidligere studier vist seg å være av stor betydning for bruk av teknologi (Agarwal & Karahanna, 2000; Saadé & Bahli, 2005). Agarwal og Karahanna (2000) finner at *kognitiv absorpsjon* har en direkte effekt på både *oppfattet nytteverdi* og *oppfattet brukervennlighet*, i tillegg til en liten effekt på *intensjon om å bruke*. Analysen av spørreskjema viser at *kognitiv absorpsjon* korrelerer relativt sterkt med alle delene av rammeverket. Det er også det konstruert i rammeverket som korrelerer sterkest med *intensjon om å bruke* ($r = 0,67$, $p < .01$). For *oppfattet nytteverdi* ($r = 0,73$, $p < .01$) og *oppfattet brukervennlighet* ($r = .51$, $p < .01$) er korrelasjonene med *kognitiv absorpsjon* av samme styrke som de er med *kredibilitet*. Dette viser at opplevelser av *kognitiv absorpsjon*, ment til å måle hvorvidt en blir oppslukt av språkmodellene, henger sammen med andre positive oppfattelser av teknologien. Å forklare hvorfor dette er tilfellet er dog mer usikkert. En mulig forklaring er at den indre motivasjonen som kjennetegner kognitiv absorpsjon fører til en opplevelse av at teknologien er mindre kognitivt krevende (Agarwal & Karahanna, 2000, s. 675-676), noe som kan gjøre at teknologien oppfattes mer brukervennlig og nyttig. Et annet perspektiv er at det faktisk blir mindre kognitivt krevende når en er oppslukt, fordi en har all oppmerksomheten rettet mot aktiviteten (Saadé & Bahli, 2005).

Det kvalitative perspektivet viser at språkmodellene oppfattes som mindre kognitivt krevende enn Google. Tidligere forskning viser at språkmodellene evner å tilby informasjon og tilpasse tekster til elevenes forståelsesnivå (Jauhiainen & Guerra, 2023; Jeon & Lee, 2023; Lo, 2023). Dette vil gjøre det mindre kognitivt krevende å lese og forstå disse, i motsetning til Google hvor elevene

må anstrenge seg for å vurdere kilder, forstå vanskelige tekster og lete etter svaret i tekstene. PISA undersøkelsene viser at mange elever kun leser om de må, og stadig flere norske elever har dårlige leseferdigheter (Jensen et al., 2023). Dette kan tenkes å forsterke motivasjonen til å bruke språkmodeller, kanskje spesielt de som tilbyr tale-til-tale. Gitt at elevene oppfatter å ha kontroll over interaksjonen på en slik måte at de nesten garantert får akkurat det svaret de er ute etter. Elevenes preferanse for ChatGPT og MyAI på bakgrunn av en oppfatning av å være i kontroll, underbygger at oppfatning om kontroll påvirker deres intensjon om å bruke. Det er viktig å presisere at dette er basert på en oppfatning av å være i kontroll, ikke objektive mål på hvorvidt man faktisk er i kontroll. Derfor kan oppfatningene også variere, for eksempel prefererte Lars Copilot, og uttrykte ikke at han manglet kontroll.

Kognitiv absorpsjon blir også ytterligere forsterket av opplevelsen av *glede* og tilfredsstillelse (Agarwal & Karahanna, 2000). I elevenes beskrivelser kommer det blant annet frem at interaksjonene med språkmodellene kan være morsomme, både i seg selv og i måten de presenterer fagstoff. Naturfaglige tekster fra internett kan oppleves som skrevet av en «*vitenskapsmann på 70 [år]*», mens språkmodellen kan presentere dette på en mer engasjerende og forståelig måte. Dette kan de selv kontrollere ved å styre språkmodellens «personlighet». Tidligere forskning viser at opplevelsen av interaktivitet bidrar til å fremme en *intensjon om å bruke* (Camilleri, 2024), men dette behøver ikke fremme læring. Tidligere studier viser imidlertid at slike åpne samtaleroboter stiller høyere krav til selvreguleringsferdigheter, da elevene kan bli fristet til å prate om andre ting, noe som igjen fører til lavere læringsutbytte (Alemdag, 2023). En annen måte å tolke korrelasjonen mellom *oppfattet nytteverdi* og *kognitiv absorpsjon* er gjennom behovet for å legitimere egen atferd. Det å oppleve teknologien som morsom fører til at en bruker mye tid på den, noe en legitimerer ovenfor seg selv ved å oppfatte den som nyttig (Agarwal & Karahanna, 2000).

5.3. KI-literacy og elevenes forståelse av teknologien

Majoriteten av elevene jeg har undersøkt ser ut til å ha lav KI-literacy. De manglet forståelse av hva KI er, hvordan det virker, samt menneskers påvirkning på KI. Ding et al. (2023) fant også at elevene misforsto språkmodellene som søkemotorer eller som identiske til Apples Siri, men dette der ut til å være vesentlig mer prevalent blant elevene jeg har intervjuet. Elevene i fokusgruppe fire uttrykte imidlertid vesentlig høyere KI-literacy.

Funnet ser ut til å støtte tidligere forskning som viser at barn oppfatter en KI sin intelligens ut fra synlige faktorer som hvor effektivt det løser en oppgave, heller enn de ikke-observerbare delene som den underliggende teknologien (Long & Magerko, 2020). Språkmodellene elevene har brukt evner å prestere svært godt, og de fremstår menneskelignende (Fraser, 2023). Det er dermed forståelig at mange av elevene har høy tillit til, og oppfatter språkmodellene som intelligente agenter. Selv innenfor høyere utdanning ser man menneskeliggjøring av språkmodeller og at det henger sammen med overdreven tillit (Ding et al., 2023). Det er grunn til å tro at problemet blir større i tiden fremover når modellene stadig presterer bedre (OpenAI, 2023, s. 59-60).

For at skolen skal kunne forberede elevene på det samfunnet og arbeidslivet de skal ta del i, må skolen kunne tilpasse seg den raske teknologiutviklingen, da denne har stor innflytelse på samfunnet og våre liv (NOU 2014: 7, 2014, s. 113, 129; NOU 2015: 8, 2015, s. 46-47). KI er en slik teknologi som stadig får mer betydning for samfunnet og livene våre, noe som viser at KI-literacy vil være viktig for elevene (Bueie, 2024). Alle fagene på skolen må på sine premisser bidra til denne utviklingen (Bueie, 2024). Naturfaget har en viktig rolle gjennom kjerneelementet teknologi: «Elevene skal forstå, skape og bruke teknologi, [...] i arbeid med naturfag. [...]. De skal vurdere hvordan teknologi kan bidra til løsninger, men også skape nye utfordringer.» (Utdanningsdirektoratet, 2020). For å forstå KI best mulig og bidra til dybdelæring, vil det i naturfag være naturlig å koble dette til scientific literacy og algoritmisk tenking, som er relaterte kompetanser til KI-literacy (Long & Magerko, 2020). Weintrop et al. (2016) argumenterer for at algoritmisk tenking er viktig i naturfag, blant annet fordi omtrent alle naturvitenskapelige fagområder nå har grener som avhenger av modelleringer, simuleringer og prosessering av store mengder data. KI, ofte maskinlæring (ML), blir brukt og utforsket som et verktøy innenfor naturvitenskap (Krenn et al., 2022), noe som er med på aktualisere ML som en del av algoritmisk tenking. Fokuset på de overordnede delene av et fag, som ML er innenfor KI, er også i tråd med Ludviksen-utvalget sin anbefaling om å fokusere på det som vil være relevant over tid i en verden som endres raskt (NOU 2015: 8, 2015).

Mine funn viser at elevene sliter med å forstå hva det er som gjør at denne teknologien fremstår intelligent. Utviklingen av denne forståelsen hører naturlig hjemme i naturfag, gjennom å utvikle en forståelse av de grunnleggende mekanismene bak KI, som ML. Arbeidet med kjerneelementet teknologi i naturfag, vil kunne gi flere ulike måter å utvikle en forståelse for ML-prosesser,

gjennom å skape og bruke. Det finnes i dag flere ulike ressurser hvor elevene selv får mulighet til å skape og bruke ML-algoritmer, både gjennom micro:bit og på PC (Bilstrup et al., 2022; Touretzky et al., 2023). Selv om disse er ugjenomsiktige modeller, vil elevene kunne få erfaring med hvordan den tilsynelatende intelligensen utvikles, og de vil kunne få erfaringer med at disse modellene gjør feilaktige slutninger. Modellene kan deretter videreutvikles, slik at de ikke gjør liknende feil i fremtiden. Hensikten med dette er å få frem at det hverken er bevissthet eller magi, det er heller systemer for å ta beslutninger, som utvikles av mennesker ved hjelp av datamaskiner (Touretzky et al., 2023).

Funnene mine indikerer imidlertid at det er behov for å også øke kunnskapen om de språkmodellene elever bruker i dag. Fraser (2023) argumenterer for at samtaleroboter bør forstås som et delt dokument, hvor brukeren og språkmodellen skriver hver sin del. Dokumentet starter med en forklaring som «Dokumentet er en transkripsjon av en samtale mellom ...», deretter arbeides dette dokumentet med ved at brukeren skriver sin del, og språkmodellen sin i tur og orden (Skjerm bilde 1). Tjenester som ChatGPT får dette til å fremstå som et samtalevindu noe som skaper en illusjon av at en prater med noen (Fraser, 2023). Long og Magerko (2020, s. 8) beskriver at barn først personifiserer samtaleroboter, for deretter å innse at de er programmerbare, en innsikt som er svært viktig. Videre argumenterer de for at alle barn bør kunne få erfaringer med å programmere KI på et nivå tilpasset dem. Denne måten lære om teknologien bak språkmodeller vil kunne benytte seg av Use-Modify-Create progresjonsmodellen beskrevet av (Lee et al., 2011). Et slikt åpent dokument, med en Enter-knapp for å bespråkmodellen forsete, tilbyr et stort rom for utforskning. Da vil elevene kunne få erfaringer med språkmodeller som noe programmerbart, som forhåpentligvis bidrar til å endre ideen om at det kun er «*Google som snakker*». Det er god grunn til å tro at en slik måte å introdusere språkmodeller til elevene er mer i tråd med anbefalingene fram Passi og Vorvoreanu (2022) om å tilpasse brukeropplevelsen til brukerens KI-literacy.

Playground

Your presets Save View code Share ...

Dette dokumentet inneholder en transkribert samtale mellom en masterstudent og språkmodellen 'Naturfagsekspert-GPT'. Naturfagsekspert-GPT gir konsise og presise svar, samtidig som den opprettholder en hjelpsom og oppmuntrende tone.

Masterstudent: Hei.

Naturfagsekspert-GPT: Hei, hva kan jeg hjelpe deg med?

Masterstudent: Hvordan kan elever på ungdomsskolen lære om teknologien bak språkmodeller?

Naturfagsekspert-GPT: Det er viktig for elever på ungdomsskolen å forstå hvordan språkmodeller fungerer, ettersom denne teknologien er stadig mer relevant og viktig i dagens verden. Det beste stedet å starte er å introdusere dem for grunnleggende konsepter som maskinlæring og kunstig intelligens, og deretter gå inn på mer spesifikke temaer som naturlig språkbehandling og datainnsamling. Det er også nyttig å la dem få prøve ut noen enkle programmeringsoppgaver, slik at de kan få en praktisk forståelse for hvordan språkmodeller utvikles og brukes. Gjennom anvendelse og eksperimentering vil elevene kunne få en dypere forståelse for teknologien bak språkmodeller og dens potensial.

Masterstudent: Er det virkelig relevant for elevene å lære om dette når de går på ungdomsskolen?

Naturfagsekspert-GPT: Ja, det er definitivt relevant for elevene å lære om teknologien bak språkmodeller allerede på ungdomsskolen. Dette skyldes at denne teknologien er stadig mer integrert i vårt daglige liv, og elevene vil sannsynligvis komme til å møte den på ulike måter både som forbrukere og i arbeidslivet i fremtiden. Det er derfor viktig å begynne å lære om den tidlig, slik at elevene kan utvikle en forståelse for hvordan den fungerer og hvordan den kan brukes på en ansvarlig og effektiv måte.

Masterstudent: Hvordan skal elevene kunne lære om all denne teknologien når det er så mye annet som står i læreplanen? Vil ikke dette gå utover muligheten for dybdelæring?

Naturfagsekspert-GPT: Det er selvfølgelig viktig å balansere læreplanen og ikke overbelaste elevene med for mye innhold. Men ved å innlemme relevant teknologi i undervisningen, vil elevene få en bredere forståelse for fagets anvendelse og kontekst. Dette kan også bidra til å øke motivasjonen og engasjementet til elevene, ettersom de får se et mer praktisk formål med det de lærer. Det er derfor viktig at lærere tar hensyn til både grunnleggende ferdigheter og moderne teknologier når de utformer undervisningsopplegg for elevene sine.

Skjerm bilde 1: Et eksempel på hvordan en kan bruke en språkmodell ment til å fullføre tekst til å lage noe som fremstår som en dialog. Den første replikken til «Naturfagsekspert-GPT» er skrevet av meg, etter det blir interaksjonen til en dialog. Skjerm bildet er hentet fra OpenAIs playground (OpenAI, 2024).

Jeg finner også at mange av elevene anser språkmodellene til å ha høy kredibilitet, spesielt i konteksten av naturfag, da noen gir uttrykk for at dette er entydige fakta. Det er ønskelig at elevene utvikler et bredere syn på kunnskap, som noe som utvikler seg, og er bygget opp av argumenter med bakgrunn i empiriske data, slik at de blir klare over at kunnskap ikke er noe som kun er gitt av en autoritet (Ferguson & Krangle, 2020). I undervisningsopplegget til Bitzenbauer (2023), fikk elevene i oppgave å kritisk vurdere en naturfaglig tekst generert av ChatGPT,

deretter skulle de endre teksten slik at den var fri for feil og hadde kildehenvisninger. Elevene sammenliknet også de ulike ledetekstene de ga til ChatGPT og de tilhørende svarene. I etterkant av dette, var elevene mer kritiske til troverdigheten til ChatGPT og mente det var viktig å ha kunnskap om teknologien om man skulle bruke den. Slike undervisningsopplegg fremstår dermed relevante for å imøtekomme oppfatninger som «*Det skal jo være riktig da, for det er jo et program*». Imidlertid uttrykker noen av elevene at de prefererer å bruke MyAI, på tross av at den har lavere kredibilitet, trolig på grunn av tilgjengeligheten. Dette tilsier at det kanskje ikke er nok å vite at teknologien ikke bør stoles blindt på, når umiddelbare svar er så viktig for noen av elevene.

6. Konklusjon

Gjennom denne studien har jeg undersøkt elevers erfaringer med bruk av språkmodeller i naturfag. De fleste elevene bruker teknologien som en søkemotor, og i litt varierende grad til å jukse. Elevene foretrekker språkmodeller fremfor Google fordi de opplever dem som enklere og mer engasjerende. De fleste anerkjenner imidlertid at teknologien har manglende kredibilitet, men det ser ut til at kun de som også har høyere KI-literacy handler deretter.

På bakgrunn av dette argumenterer jeg for at KI-literacy bør inkluderes i elevers naturfagundervisning, og at det hører hjemme i naturfaget gjennom kjerneelementet teknologi, så vel som naturfagets bidrag til algoritmisk tenking og scientific literacy.

Sammenhengen mellom fornuftig bruk av språkmodellene og KI-literacy er imidlertid basert på få elever, og bør undersøkes nærmere. Gitt at bruken av språkmodeller oppleves vesentlig mindre kognitivt krevende er det nærliggende å tro at disse fremdeles vil brukes relativt ukritisk som søkemotorer.

Jeg tror fremdeles språkmodellene har potensiale i skolen. I teoridelen har jeg beskrevet flere interessante måter språkmodellene kan brukes i læringsarbeidet, og det faktum at elevene jeg intervjuet opplevde språkmodellene som mindre kognitivt krevende måter å finne informasjon på, trenger ikke utelukkende negativt. Fremtidig forskning bør undersøke hvordan bruken av teknologien kan endres til mer hensiktsmessige måter for å faktisk støtte læringsprosessen.

Referanser

- Abbas, M., Jam, F. A. & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Agarwal, R. & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, 24(4), 665-694. <https://doi.org/10.2307/3250951>
- Albayati, H. (2024). Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study. *Computers and Education: Artificial Intelligence*, 6, 100203. <https://doi.org/10.1016/j.caeai.2024.100203>
- Alemdag, E. (2023). The effect of chatbots on learning: a meta-analysis of empirical research. *Journal of Research on Technology in Education*, 1-23. <https://doi.org/10.1080/15391523.2023.2255698>
- Assaker, G. (2020). Age and gender differences in online travel reviews and user-generated-content (UGC) adoption: extending the technology acceptance model (TAM) with credibility theory. *Journal of Hospitality Marketing & Management*, 29(4), 428-449. <https://doi.org/10.1080/19368623.2019.1653807>
- Bilstrup, K.-E. K., Kaspersen, M. H., Sørensen, M.-L. S. K. & Petersen, M. G. (2022). *Opportunities and Challenges of Teaching Machine Learning as a Design Material with the micro:bit*. Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference, Aarhus, Denmark. <https://doi.org/10.1145/3547522.3547689>
- Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *CONTEMPORARY EDUCATIONAL TECHNOLOGY* 15(3). <https://doi.org/10.30935/cedtech/13176>
- Blikstad-Balas, M. (2023, 22.06.23). literacy. I *Store norske leksikon*. <https://snl.no/literacy>
- Bran, A. M., Cox, S., White, A. D. & Schwaller, P. (2023). ChemCrow: Augmenting large-language models with chemistry tools. <https://doi.org/10.48550/arXiv.2304.05376>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G. & Askell, A. (2020). Language models are few-shot learners. *Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.2005.14165>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J. A., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y.-F., Lundberg, S. M., Nori, H., Palangi, H., Ribeiro, M. T. & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *ArXiv, abs/2303.12712*. <https://doi.org/10.48550/arXiv.2303.12712>
- Bueie, A. A. (2024). Elevperspektiver på å bruke samtaleroboter i skriving. "Den skriver bra, men den skriver ikke deg". I T. Stenseth & M. Oddvik (Red.), *Læring og undervisning i en digital verden* (1. utgave. utg.). Gyldendal.
- Børne- og Undervisningsministeriet. (2024). *Ekspertgruppen om ChatGPT og andre digitale hjelpemidler*. B.-o. Undervisningsministeriet. <https://www.uvm.dk/aktuelt/nyheder/uvm/2024/april/240424-ekspertgruppe-klar-med-anbefalinger-for-brug-af-chatgpt-ved-proever>
- Camilleri, M. A. (2024). Factors affecting performance expectancy and intentions to use ChatGPT: Using SmartPLS to advance an information technology acceptance framework. *Technological Forecasting and Social Change*, 201, 123247. <https://doi.org/10.1016/j.techfore.2024.123247>
- Chiu, T. K. F., Moorhouse, B. L., Chai, C. S. & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, 1-17. <https://doi.org/10.1080/10494820.2023.2172044>

- Clark, T. M. (2023). Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions. *Journal of Chemical Education*, 100(5), 1905-1916. <https://doi.org/10.1021/acs.jchemed.3c00027>
- Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education* (Eighth edition. utg.). Routledge.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- De nasjonale forskningsetiske komiteene. (2014). *Generelle forskningsetiske retningslinjer*. <https://www.forskningsetikk.no/retningslinjer/generelle/>
- Ding, L., Li, T., Jiang, S. & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, 20(1), 63. <https://doi.org/10.1186/s41239-023-00434-1>
- Duenger Bøhn, E. (2024). Kan generativ kunstig intelligens opplyse og danne oss? *Nordisk Tidsskrift for Pedagogikk og Kritikk*, 10(1). <https://doi.org/10.23865/ntpk.v10.6175>
- Duong, C. D., Vu, T. N. & Ngo, T. V. N. (2023). Applying a modified technology acceptance model to explain higher education students' usage of ChatGPT: A serial multiple mediation model with knowledge sharing as a moderator. *The International Journal of Management Education*, 21(3), 100883. <https://doi.org/https://doi.org/10.1016/j.ijme.2023.100883>
- Elstad, E. (2023). *Læreren møter ChatGPT*. Universitetsforlaget.
- Fangen, K. (2010). *Deltagende observasjon* (2. utg. utg.). Fagbokforl.
- Farrokhnia, M., Banihashem, S. K., Noroozi, O. & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1-15. <https://doi.org/10.1080/14703297.2023.2195846>
- Fauskanger, J. & Mosvold, R. (2014). Innholdsanalysens muligheter i utdanningsforskning. *Norsk pedagogisk tidsskrift*, 98(2), 127-139. <https://doi.org/10.18261/ISSN1504-2987-2014-02-07>
- Ferguson, L. E. & Krange, I. (2020). Hvordan fremme kritisk tenkning i grunnskolen? *Norsk pedagogisk tidsskrift*, 104(2), 194-205. <https://doi.org/10.18261/issn.1504-2987-2020-02-09>
- Fijačko, N., Gosak, L., Štiglic, G., Picard, C. T. & John Douma, M. (2023). Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation*, 185, 109732. <https://doi.org/https://doi.org/10.1016/j.resuscitation.2023.109732>
- Fraser, C. (2023, 27.02.2023). *Who are we talking to when we talk to these bots?* Medium. <https://medium.com/@colin.fraser/who-are-we-talking-to-when-we-talk-to-these-bots-9a7e673f8525>
- Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A. & Berner, J. J. (2023). Mathematical Capabilities of ChatGPT. *ArXiv, abs/2301.13867*.
- Halkier, B. (2010). *Fokusgrupper* (K. Gjerpe, Red.). Gyldendal akademisk. https://urn.nb.no/URN:NBN:no-nb_digibok_2013082808155
- Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16(1), 45-74. <https://doi.org/10.1007/s12559-023-10179-8>
- Hsieh, H.-F. & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277-1288. <https://doi.org/10.1177/1049732305276687>
- Jauhiainen, J. S. & Guerra, A. G. (2023). Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability*, 15(18), 14025. <https://www.mdpi.com/2071-1050/15/18/14025>
- Jensen, F., Pettersen, A., Frønes, T. S., Eriksen, A., Løvgren, M. & Narvhus, E. K. (2023). PISA 2022: Norske elevers kompetanse i matematikk, naturfag og lesing. I. Cappelen Damm Akademisk/NOASP (Nordic Open Access Scholarly Publishing). <https://doi.org/10.23865/noasp.205>

- Jeon, J. & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873-15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274>
- Kommunal- og moderniseringsdepartementet. (2020). *Nasjonal strategi for kunstig intelligens*. Regjeringen. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19(1), 010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., dos Passos Gomes, G., Häse, F., Jinich, A., Nigam, A., Yao, Z. & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12), 761-769. <https://doi.org/10.1038/s42254-022-00518-3>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J. & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Kunnskapsdepartementet. (2017). *Overordnet del – verdier og prinsipper for grunnopplæringen*. <https://www.regjeringen.no/no/dokumenter/verdier-og-prinsipper-for-grunnopplaringen/id2570003/>
- Kunnskapsdepartementet & KS. (2023). *Strategi for digital kompetanse og infrastruktur i barnehage og skole*. Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/strategi-for-digital-kompetanse-og-infrastruktur-i-barnehage-og-skole/id2972254/>
- Kvale, S. & Brinkmann, S. (2015). *Det kvalitative forskningsintervju* (3. utg., 2. oppl. utg.). Gyldendal akademisk.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smith, J. & Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads*, 2(1), 32–37. <https://doi.org/10.1145/1929887.1929902>
- Li, C. & Xing, W. (2021). Natural Language Generation Using Deep Learning to Support MOOC Learners. *International Journal of Artificial Intelligence in Education*, 31(2), 186-214. <https://doi.org/10.1007/s40593-020-00235-x>
- Liu, G. & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125-138. <https://doi.org/10.1080/17501229.2023.2240316>
- LNU. (2023, 19.12.23). *Brev til Stortingets forsknings- og utdanningskomité*. landslaget for norskundervisning. <https://www.norskundervisning.no/nyheter-og-artikler/brev-til-stortingets-forsknings-og-utdanningskomit>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://www.mdpi.com/2227-7102/13/4/410>
- Long, D. & Magerko, B. (2020). *What is AI Literacy? Competencies and Design Considerations*. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. <https://doi.org/10.1145/3313831.3376727>

- Lowry, P. B., Gaskin, J. E., Twyman, N. W., Hammer, B. & Roberts, T. L. (2013). Taking “Fun and Games” Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM). *Journal of the Association for Information Systems*, 14(11). <https://doi.org/10.17705/1jais.00347>
- Marangunić, N. & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, 14(1), 81-95. <https://doi.org/10.1007/s10209-014-0348-1>
- Menon, D. & Shilpa, K. (2023). “Chatting with ChatGPT”: Analyzing the factors influencing users' intention to Use the Open AI's ChatGPT using the UTAUT model. *Heliyon*, 9(11), e20962. <https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e20962>
- Meseguer-Artola, A., Aibar, E., Lladós, J., Minguillón, J. & Lerga, M. (2016). Factors that influence the teaching use of Wikipedia in higher education. *Journal of the Association for Information Science and Technology*, 67(5), 1224-1232. <https://doi.org/https://doi.org/10.1002/asi.23488>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A. & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/https://doi.org/10.1016/j.caeai.2023.100199>
- Nardi, P. M. (2018). *Doing survey research : a guide to quantitative methods* (Fourth edition. utg.). Routledge.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/https://doi.org/10.1016/j.caeai.2021.100041>
- NOU 2014: 7. (2014). *Elevenes læring i fremtidens skole — Et kunnskapsgrunnlag* Kunnskapsdepartementet.
- NOU 2015: 8. (2015). *Fremtidens skole -*
- Fornyelse av fag og kompetanser.* Kunnskapsdepartementet.
- OpenAI. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*. <https://doi.org/https://doi.org/10.48550/arXiv.2303.08774>
- OpenAI. (2024). *Playground*. OpenAI. <https://platform.openai.com/playground/chat>
- Palos-Sanchez, P. R., Saura, J. R. & Velicia-Martin, F. (2022). A Case Study on a Hedonic-Motivation System Adoption Model in a Game-Based Student Response System. *International Journal of Human-Computer Interaction*, 40(3), 701-718. <https://doi.org/10.1080/10447318.2022.2121801>
- Passi, S. & Vorvoreanu, M. (2022). Overreliance on AI Literature Review. *Microsoft Research*. <https://www.microsoft.com/en-us/research/uploads/prod/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf>
- Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. Open Ai. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Rokhayani, A., Rukmini, D., Hartono, R. & Mujiyanto, J. (2022). Integrating Technology in Online Learning Based on Computer-Mediated Communication Artificial Intelligence to Improve Students' Achievement. *Journal of Higher Education Theory & Practice*, 22(15), 234-244. <https://doi.org/10.33423/jhetp.v22i15.5575>
- Sarikaya, R. (2017). The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, 34(1), 67-81. <https://doi.org/10.1109/MSP.2016.2617341>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M. & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2024.101894>

- Saadé, R. & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model. *Information & Management*, 42(2), 317-327. <https://doi.org/https://doi.org/10.1016/j.im.2003.12.013>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273-1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Thagaard, T. (2018). *Systematikk og innlevelse : en innføring i kvalitative metoder* (5. utg. utg.). Fagbokforl.
- Touretzky, D., Gardner-McCune, C. & Seehorn, D. (2023). Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education*, 33(2), 233-266. <https://doi.org/10.1007/s40593-022-00314-1>
- Utdanningsdirektoratet. (2020). *Læreplan i naturfag 1.–10. trinn (NAT01-04)*. <https://data.udir.no/kl06/v201906/laereplaner-1k20/NAT01-04.pdf?lang=no>
- Vedersø, B., Andersen, H. L., Andreasen, M., Aslak, M., Augustinus, I. B., Damsgaard, J., Jensen, T. W., Lauridsen, P. S. & Sørensen, B. M. (2024, 25.04.2024). Medlemmer af ekspertgruppe: Her er de anbefalinger, vi ikke blev bedt om. *Altinget*. <https://www.alinget.dk/uddannelse/artikel/medlemmer-af-ekspertgruppe-her-er-de-anbefalinger-vi-ikke-blev-bedt-om>
- Venkatesh, V. & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315. <https://doi.org/DOI:10.1111/J.1540-5915.2008.00192.X>
- Vogelsang, K., Steinhüser, M. & Hoppe, U. (2013). *A Qualitative Approach to Examine Technology Acceptance* (Bd. 1).
- Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L. & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of science education and technology*, 25(1), 127-147. <https://doi.org/10.1007/s10956-015-9581-5>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.
- Wibeck, V., Dahlgren, M. A. & Öberg, G. (2007). Learning in focus groups: an analytical dimension for enhancing focus group research. *Qualitative Research*, 7(2), 249-267. <https://doi.org/10.1177/1468794107076023>
- Wu, K., Zhao, Y., Zhu, Q., Tan, X. & Zheng, H. (2011). A meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. *International Journal of Information Management*, 31(6), 572-581. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2011.03.004>
- Wu, R. & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10-33. <https://doi.org/https://doi.org/10.1111/bjet.13334>
- Xuan-Quy, D., Ngoc-Bich, L., Bac-Bien, N. & Xuan-Dung, P. (2023). LLMs' Capabilities at the High School Level in Chemistry: Cases of ChatGPT and Microsoft Bing Chat. *ChemRxiv*. <https://doi.org/doi.org/10.26434/chemrxiv-2023-kxxpd>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/https://doi.org/10.1111/bjet.13370>
- Yin, J., Goh, T.-T., Yang, B. & Xiaobin, Y. (2021). Conversation Technology With Micro-Learning: The Impact of Chatbot-Based Learning on Students' Learning Motivation and Performance. *Journal of Educational Computing Research*, 59(1), 154-177. <https://doi.org/10.1177/0735633120952067>
- Zulic, A., Reikerås, M. & Granbo, K. (2023, 16.06.23). 1 av 5 elever bruker ChatGPT til skolearbeid. *NRK*. <https://www.nrk.no/kultur/na-sier-1-av-5-norske-elever-at-de-bruker-chatgpt-til-skolearbeid-1.16442993>

Vedlegg 1: Intervjuguide

Før selve intervjuet

- Kort repetisjon om hensikten til forskningsprosjektet og intervjuet,
- Informere på nytt om samtykke og retten til å trekke seg.
- Ønske at dere skal diskutere dere imellom

Oppvarming, tidligere erfaring

- Har dere brukt Copilot/ChatGPT/My AI tidligere?
 - Hva her dere brukt det til?
- Hvis dere er nysgjerrige og lurer på noe relatert til naturfag hvordan går dere for å finne svar?
 - Hvorfor bruker dere språkmodeller?
 - Hvordan bruker dere språkmodeller?

Nytteverdi, strategier

- Er det noe dere føler språkmodeller har hjulpet dere å forstå?
 - Hva spurte dere om?
- Har dere opplevd at språkmodeller fungerer dårlig til noe?
 - Har dere prøvd å forklare dette til språkmodellene?

Brukervennlighet

- Hvordan oppleves språkmodellene å bruke?
 - Svarer de slik dere ønsker?

Kognitiv absorpsjon

- Er det noen forskjeller mellom xxx og yyyy i form an glede/tilfredstillelse
- Er det forskjeller i hvor stor grad dere fokuserer på oppgaven når dere bruker språkmodeller vs andre verktøy, for eksempel Wikipedia?
- Opplevelse av å forsvinne inn i et tema? Tidsdissosiasjon og nysgjerrighet.
- Hva er det mest frustrerende med å bruke språkmodeller?

Kredibilitet

- Har dere opplevd at språkmodeller gir feilaktig informasjon?
 - Hvordan kan dette skje?
 - Erfaringer med at språkmodellene sier hvem som vinner neste fotball VM?

- Hva påvirker språkmodellenes svar?

Avslutning

- Oppsummerer intervjuet og min tolkning.
 - Stemmer det?
 - Er det noe dere føler jeg har gått glipp av eller ønsker å legge til?
- Hvordan opplevde dere intervjuet?
- Fikk dere sagt det dere ønsket å si, eller styrte jeg samtalen i for stor grad?
- Har dere noen spørsmål til meg?

Vedlegg 2: Spørreskjema

Intensjon om å bruke	Jeg kommer nok ikke til å bruke Copilot i fremtiden.
	Om jeg får muligheten vil jeg bruke Copilot for å lære naturfag i fremtiden.
Oppfattet nytteverdi	Jeg mener der er mer hjelpsomt å spørre læreren enn Copilot.
	Jeg mener det er mer nyttig å søke på Google for å finne svar enn å spørre Copilot.
	Jeg opplevde at Copilot hjalp meg å forstå naturfag bedre.
Oppfattet brukervennlighet	Jeg forsto raskt hvordan jeg kunne bruke Copilot.
	Det var vanskelig å forstå hvordan jeg kunne bruke Copilot.
Troverdighet	Jeg stoler på at svarene fra Copilot er korrekte.
	Jeg har tillit til at Copilot gir objektive og nøytrale svar.
Ekspertise	Copilot er ikke en ekspert i naturfag.
	Min erfaring er at Copilot er kompetent til å besvare spørsmål i naturfag.
Kontroll	Jeg opplevde at Copilot svarte slik jeg ønsket.
	Copilot svarte på andre ting enn det jeg ba den om.
Nysgjerrighet	Samtalene med Copilot vekket min nysgjerrighet.
	Copilot inspirerte meg ikke til å undersøke nye og flere ting.
	Samtalene med Copilot gjorde at jeg ønsket å lære mer.
Fokusert innlevelse	Jeg synes det var enkelt å holde konsentrasjonen når jeg brukte Copilot.

	Jeg ble lett distrahert av andre ting enn Copilot.
Glede	Det var gøy å bruke Copilot til å lære naturfag.
	Det var frustrerende å bruke Copilot til å lære naturfag.
Tidsdissosiasjon	Tiden fløy når jeg brukte Copilot.
	Jeg ble oppslukt i samtalene med Copilot.

Vedlegg 4: Godkjenning fra Sikt



Vurdering av behandling av personopplysninger

Referansenummer
105528

Vurderingstype
Standard

Dato
19.02.2024

Tittel

ChatGPT i klasserommet: Undersøkelse av elevenes strategier og oppfatninger i prøveforberedelser

Behandlingsansvarlig institusjon

OsloMet – storbyuniversitetet / Fakultet for lærerutdanning og internasjonale studier / Institutt for grunnskole- og faglærerutdanning

Prosjektansvarlig

Aase Marit T. Sørum Ramton

Student

Emil Mikalsen

Prosjektperiode

15.01.2024 - 31.12.2024

Kategorier personopplysninger

Alminnelige

Lovlig grunnlag

Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 31.12.2024.

Meldeskjema [↗](#)

Kommentar

OM VURDERINGEN

Sikt har en avtale med institusjonen du forsker eller studerer ved. Denne avtalen innebærer at vi skal gi deg råd slik at behandlingen av personopplysninger i prosjektet ditt er lovlig etter personvernregelverket. Vi har nå vurdert at du har lovlig grunnlag til å behandle personopplysningene.

BARN SAMTYKKER SELV

Prosjektet vil innhente samtykke fra mindreårige til behandling av personopplysninger. Vår vurdering er at barn over 15 år kan samtykke selv til behandling av alminnelige personopplysninger, og at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Vi vil anbefale at foreldre får informasjon om at prosjektet gjennomføres i deres barn sin klasse.

FØLG DIN INSTITUSJONS RETNINGSLINJER

Det er institusjonen du er ansatt/student ved som avgjør hvordan du må lagre og sikre data i ditt prosjekt og hvilke databehandlere du kan bruke. Husk å bruke leverandører som din institusjon har avtale med (f.eks. ved skylagring, nettspørreskjema, videosamtale el.).

Personverntjenester legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til oss ved å oppdatere meldeskjemaet. Se våre nettsider om hvilke endringer du må melde: <https://sikt.no/melde-endringer-i-meldeskjema>

OPPFØLGING AV PROSJEKTET

Vi vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Vedlegg 5: Samtykkeerklæring

Vil du delta i forskningsprosjektet

Copilot i klasserommet: Undersøkelse av elevenes strategier og erfaringer med bruk i naturfag?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å lære om hvordan ungdomsskoleelever kan bruke Copilot som et verktøy i naturfagundervisningen og hvordan det oppleves. I dette skrivet vil jeg gi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Som avslutning på lærerutdanningen skal jeg skrive en masteroppgave, det innebærer at vi skal gjennomføre et lite forskningsprosjekt. Jeg har valgt å forske på ungdomsskoleelevers bruk av Copilot. I mitt forskningsprosjekt ønsker jeg å se på hvordan ungdomsskoleelever kan bruke Copilot som et verktøy for å tilegne seg kunnskap i naturfag. Jeg håper å lære om hvilke ulike strategier som blir brukt, og hvor nyttig det oppleves å bruke Copilot.

Hvem er ansvarlig for forskningsprosjektet?

OsloMet, universitetet jeg går på, er ansvarlig for prosjektet. Min veileder Aase Marit T. Sørum Ramton er prosjektansvarlig.

Hvorfor får du spørsmål om å delta?

Du blir bedt om å delta fordi du er elev på 10. trinn, og det er det jeg ønsker å undersøke.

Hva innebærer det for deg å delta?

Forskningsprosjektet vil foregå som en del av naturfagundervisningen. Om du ønsker å være en del av forskningsprosjektet, innebærer det at jeg vil observere deg mens du jobber. Jeg vil gå rundt i klasserommet og se på hva alle de som ønsker å delta jobber med og kanskje stille noen spørsmål om det er noe jeg lurer på. Etter timen vil jeg gi deg et kort spørreskjema med spørsmål om hvordan du opplevde å bruke Copilot og hvor hjelpsomt og troverdig du følte det var å bruke Copilot som et hjelpemiddel. Jeg ønsker også å gjennomføre gruppeintervjuer, men det er ikke sikkert du blir valgt ut til dette. Der vil jeg spørre dere om hvilke strategier dere har brukt, hva dere opplevde at fungerte bra og dårlig osv. Jeg vil ta opp lyd under intervjuet, kort tid etter intervjuet vil jeg skrive ned hva som blir sagt og slette lydfilen. I masteroppgaven min vil jeg anonymisere dere, det vil si at jeg gir dere fiktive navn, så ingenting dere sier vil kunne knyttes til dere. Intervjuet kommer til å ta rundt en skoletime.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake (endre mening) uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Dersom du velger å ikke delta vil du fremdeles jobbe med øving til naturfagsprøven-

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger

Jeg behandler alle opplysningene konfidensielt og i samsvar med personvernregelverket. Jeg vil bare samle inn notater fra det jeg observerer i timene, svarene på de anonyme spørreskjemaene og intervjuene.

- Notatene mine fra observasjonen vil være anonyme og kun lagres på OsloMet sin skylagringstjeneste, hvor det bare er jeg som har tilgang.
- Spørreundersøkelsen vil gjennomføres ved hjelp av tjenesten Nettskjema. Nettskjema er utviklet av Universitetet i Oslo til bruk i forskning, og sørger for at svarene dine forblir anonyme.
- Intervjuene vil bli tatt opp, ved hjelp av Nettskjema. Kort tid etter intervjuet vil jeg bruke Nettskjema til å transkribere (skrive ned) det som ble sagt, og slette lydfilene. I de nedskrivende intervjuene vil du få et fiktivt navn, slik at det du sier ikke kan spores tilbake til deg. Intervjuene vil bli transkribert ved hjelp av Nettskjema, noe som sørger for at dataen ikke forlater Universitetet i Oslo sine servere. De nedskrevne og anonymiserte intervjuene lagres på OsloMet sin skylagringstjeneste hvor det bare er jeg som har tilgang.

For å få tilgang til Copilot vil dere bruke skolens egen tilgang, som er i tråd med personvernsregelverket og i tråd med kommunens retningslinjer. Det som blir skrevet til Copilot vil ikke lagres eller kunne knyttes til deg. Jeg har heller ikke tilgang til å se det du skriver.

Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?

Prosjektet vil avsluttes innen desember 2024. I god tid før dette er intervjuene blitt transkribert og anonymisert, det vil dermed ikke være noen personopplysninger som blir lagret. Når prosjektet er avsluttet er alle dataene anonymisert.

Hva gir oss rett til å behandle personopplysninger om deg?

Jeg behandler opplysninger om deg basert på ditt samtykke.

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

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger jeg behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Min veileder fra OsloMet, Aase Marit T. Sørum Ramton på epost: aasemarit.ramton@oslomet.no
- Eller meg, Emil Mikalsen på epost s344828@oslomet.no
- Vårt personvernombud: Ingrid S. Jacobsen på epost: personvernombud@oslomet.no

Hvis du har spørsmål knyttet til vurderingen som er gjort av personverntjenestene fra Sikt, kan du ta kontakt via:

- Epost: personverntjenester@sikt.no eller telefon: 73 98 40 40.

Med vennlig hilsen

Førstelektor

Aase Marit T. Sørum Ramton

Masterstudent

Emil Mikalsen

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Copilot i klasserommet: Undersøkelse av elevenes strategier og erfaringer med bruk i naturfag?*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å bli observert og svare på spørreskjemaet
- å delta i intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Ditt navn, dato)