



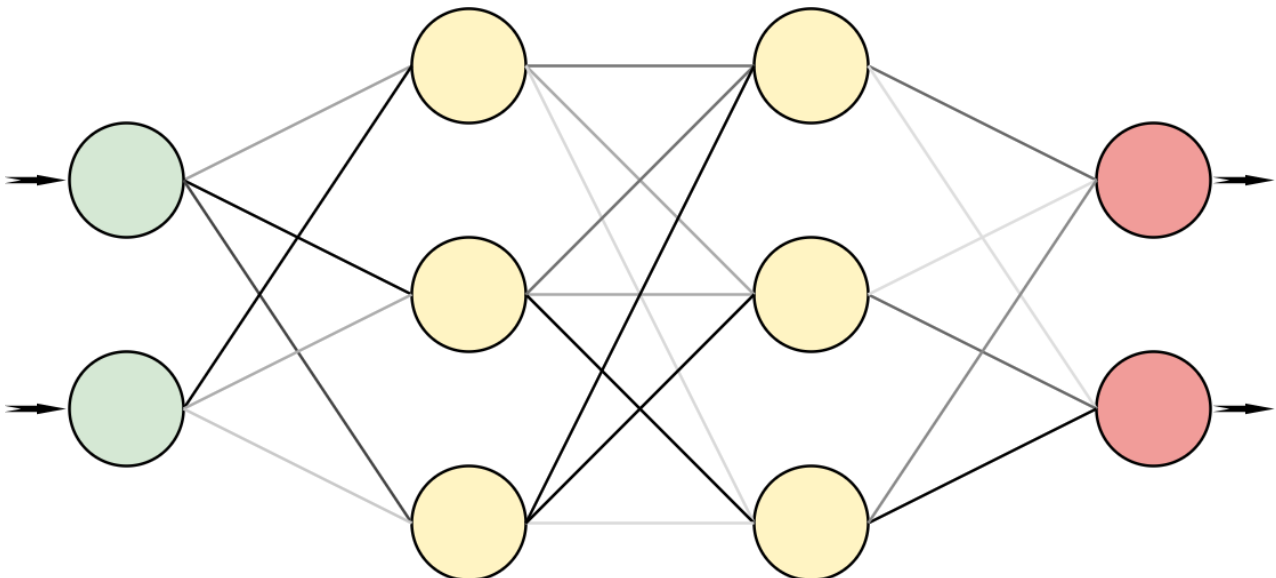
Examensarbete inom teknik och lärande

Avancerad nivå, 30 hp

Generativ AI i gymnasieskolan

Undersökning av en lektionsseries påverkan på
gymnasieelevernas färdigheter

BARTOSZ MICHAŁ PIÓRKOWSKI





2024-06-04

Kungliga Tekniska Högskolan

På uppdrag av Håkan Elderstig
Stockholm Science and Innovation School

FÖRFATTARE: Bartosz Michał Piórkowski

HANDLEDARE: Helena Isaksson Persson

EXAMINATOR: Ernest Ampadu

SAMMANFATTNING

Denna kvasiexperimentella studie syftade till att undersöka hur en lektionsserie kan struktureras och implementeras med mål att utveckla gymnasieelevers förmåga att använda sig av generativ artificiell intelligens som ett pedagogiskt verktyg.

För att möta detta syfte genomfördes tre lektioner om artificiell intelligens, maskin-inlärning, neurala nätverk och stora språkmodeller med fokus på utveckling av teknisk kunskap och praktiska färdigheter med inslag av etik och kritik. Valet av dessa teman grundades i ett tidigare etablerat ramverk för undervisning inom AI-läskunnighet. Vidare teman tas dessa teman upp som del av teknikprogrammet och den kommande AI-kursen enligt Skolverkets förslag. Lektionsseriens påverkan kvantifierades med hjälp av två enkäter – en innan och en efter genomförandet av lektionsserien.

Lektionsserien presenterades för två gymnasieklasser vilka bestod av totalt ungefär 50 elever. Urvalet av gymnasieklasserna grundades i deras anslutning till den uppdragsgivande läraren. Vidare valdes respondenterna till enkäten utifrån de elever som fysiskt deltog på den första och sista lektionen och frivilligt valde att svara på enkäten. Dessutom intervjuades fyra tekniklärare för att bättre anpassa lektionsinnehållet till målgruppen.

Analysen av svarsfrekvensen till enkätfrågorna visade att lektionsserien hade en statistiskt signifikant påverkan på elevernas tekniska kunskaper, men dess påverkan på elevernas praktiska färdigheter var i stort statistiskt insignifikant. Samtidigt påvisade frekvensanalysen att gymnasieeleverna i regel överskattade sin förmåga att kritiskt granska datorgenererad text och var i stort omedvetna om relevanta etiska frågeställningar.

Explorativa faktoranalysen visade att det existerar åtminstone två typer av elever. En elevgrupp av okänd storlek använder sig av stora språkmodeller för att accelerera sina studier genom att lösa problem de annars inte kunde lösa. I detta fall har artificiell intelligens en multiplicerande effekt på elevernas produktivitet. En annan elevgrupp av okänd storlek har i stället som mål att förbättra sina skolresultat genom att använda sig av stora språkmodeller för att lösa deras problem åt dem. Samtidigt överskattar dessa elever sin förmåga att granska datorgenererad text. I detta fall har artificiell intelligens en dämpande effekt på elevernas lärande.

Studiens slutsats är att det i dagsläget finns behov för undervisning av gymnasieelever på teknikprogrammet om artificiell intelligens. Detta utrymme kan i stort uppfyllas av en tre lektioner lång lektionsserie. Dock erkänner studien att det finns ytterligare utrymme för praktiska moment där läraren handleder eleverna i deras användning av verktyg såsom ChatGPT. Vidare finns det utrymme för kontinuerligt arbete med kritik och etik, möjligtvis som del av de tidigare nämnda praktiska momenten.

Nyckelord: utbildning, pedagogik, AI-läskunnighet, stora språkmodeller

ABSTRACT

This quasi-experimental study aimed to investigate how a series of lessons could be structured and implemented with the goal of developing secondary level students' ability to use generative artificial intelligence as an educational tool.

To meet this goal three lessons on artificial intelligence, machine learning, neural networks, and large language models were conducted, focusing on the development of technical knowledge and practical skills with the inclusion of ethics and critical thinking. The choice of these topics was based on a previously established framework for AI-literacy education. Further, these topics are brought up as part of the Swedish upper secondary school technology programme as well as the upcoming AI-course as per the proposal made by the Swedish Agency for Education. The impact of the lesson series was quantified using two form surveys – one before and one after the implementation of the lesson series.

The lesson series was presented to two student classes totalling roughly 50 students. The selection of student classes were based on their affiliation with the assigning teacher. Further, the survey respondents were sampled from the students who physically attended the first and last lesson and voluntarily elected to respond. Additionally, four technology teachers were interviewed to better adapt the teaching material to the student demographic.

Response analysis showed that the lesson series had a statistically significant impact on students' technical knowledge, but its impact on students' practical skills was largely statistically insignificant. At the same time, the frequency analysis indicated that students generally overestimated their ability to critically evaluate computer-generated text and were largely unaware of relevant ethical issues.

Exploratory factor analysis had shown that there exist at least two types of students. A student group of unknown size use large language models to accelerate their studies through solving problems they could not otherwise solve. In this case, artificial intelligence has a multiplying effect on the students' productivity. Another group of students of unknown size instead use large language models to solve their problems for them with the goal of improving their academic performance. At the same time, these students overestimate their ability to evaluate computer-generated text critically. In this case, artificial intelligence has a dampening effect on the students' learning.

The study concludes that there is a need for teaching secondary level students from the technology programme about artificial intelligence. This space can largely be fulfilled by a series of three lessons. However, the study acknowledges that there remains room for practical activities where the teacher guides students in their use of tools such as ChatGPT. Furthermore, there is room for ongoing work on critical thinking and ethics, possibly as part of the aforementioned practical activities.

Keywords: education, pedagogy, AI literacy, large language models

INNEHÅLL

Inledning	1
Syfte.	1
Frågeställning.	2
Avgränsning	2
Begreppsbeskrivning.	2
Artificiell intelligens	2
Maskininlärning	3
Artificiella neurala nätverk	3
Stora språkmodeller	3
Tidigare forskning.	4
AI-läskunnighet	4
Personliga assistenter	4
Metod	6
Intervju	7
Enkät	7
Analys	10
Lektionsserie	10
Etiska överväganden	13
Resultat.	13
Intervju	14
Lärarnas erfarenheter	14
Elevernas användning.	14
AI på skolan.	14
Enkät	15
Frekvensanalys.	15
Statistiskt signifikanta frågor	17
Statistisk analys	19
Faktoranalys.	19
Diskussion.	21
Svarsfrekvenser	21
Faktorer	23
Första faktorn	24
Andra faktorn	25
Lärarnas perspektiv	25
Framtida forskning	26
Slutsats	27
Referenser	28

INLEDNING

Inom den svenska gymnasieskolan har digitaliseringen sedan länge varit en central och pågående utmaning (Skolverket, 2018). Den snabba tekniska utvecklingen har inte bara förändrat hur samhället fungerar, utan även skapat nya möjligheter inom utbildningssektorn. I synnerhet har tillgängligheten till digitala verktyg med inbyggd *artificiell intelligens* (AI) ökat och verktyg som OpenAIs ChatGPT m.fl. har blivit alltmer populära.

En följd av den utbredda användningen av AI i samhället är att Skolverket (2024) planerar introducera AI som ett nytt ämne från och med hösten 2024. Enligt Skolverkets förslag kommer AI-ämnet ta del av programstrukturen för både teknik- och naturprogrammen samt vara tillgänglig som individuellt val för övriga program. Eftersom AI har stor samhällsnytta och används av stora delar av befolkningen finns det behov för ett bättre förståelse för hur sådana AI-system kan användas på ett pedagogiskt sätt, inte minst som del av den kommande AI-kursen, men även som del av gymnasieskolan i stort.

Detta examensarbete ämnar till att undersöka och bidra till förståelsen av hur digitala verktyg med AI kan integreras i gymnasieundervisningen. Examensarbetet strävar efter att erbjuda en holistisk och praktisk insikt i användningen av AI som en del av lärandet på gymnasieskolan med en fokus på *stora språkmodeller* (LLM). Detta uppnås genom att kombinera teoretisk kunskap med praktisk tillämpning i ett nära samarbete med en gymnasieskola i Stockholmsområdet samt ett samarbetsprojekt för praktisknära forskning.

Eftersom examensarbetet skrivs i kontext av programmet *Civilingenjör och Lärare på Kungliga Tekniska Högskolan* (KTH) kommer fokus ligga på två specifika delar. Den tekniska delen kommer att fokusera på processen för hur en lektionsserie kan utvecklas och den pedagogiska delen kommer i stället att fokusera på den praktiska implementationen av lektionsserien. Den efterföljande analysen och diskussionen kommer därför i grund och botten vara tvärvetenskaplig.

SYFTE

I stort fokuserar tidigare forskning inom generativ AI som del av utbildning på hur dessa system kan integreras i ett större sammanhang. Det övergripande syftet med detta examensarbete är därför att undersöka och analysera effektiviteten av en strukturerad pedagogisk intervention i formen av en lektionsserie på gymnasieelevers tekniska kunskaper och praktiska färdigheter inom generativ AI. Detta för att till skillnad från tidigare forskning identifiera konkreta interaktionsmönster och strategier som främjar lärande och förbättrar utbildningsresultat.

Målet är att genom skapande och implementering av en lektionsserie förstärka *gymnasieelevernas* förmåga att använda generativ AI som ett pedagogiskt verktyg för lärande. Fokuset ligger inte bara på att förbättra gymnasieelevernas tekniska

kunskaper och praktiska färdigheter, men även på att fördjupa deras förståelse för rollen av AI inom utbildning samt främja kritiskt tänkande och presentera etiska överväganden.

FRÅGESTÄLLNING

För att uppnå examensarbetets syfte kommer de två följande forskningsfrågor att besvaras:

1. Hur påverkas gymnasieelevers användningsvanor, praktiska färdigheter och teknisk kunskap efter genomförandet av en lektionsserie om generativ artificiell intelligens?
2. Hur kan en lektionsserie struktureras för att förbättra gymnasieelevernas förmåga att förstå och använda generativ artificiell intelligens som ett pedagogiskt verktyg?

Genom att svara på dessa forskningsfrågor strävar examensarbetet efter att inte bara analysera nuvarande användning av digitala verktyg med inbyggd AI utan även att identifiera konkreta metoder för att förstärka gymnasieelevernas förmåga att dra nytta av AI vid lärandesituationer.

AVGRÄNSNING

Studien genomfördes på elever som läser det andra året av teknikprogrammet på en gymnasieskola i Stockholmsområdet. Forskningsfrågorna kan därför endast besvaras inom ramen för elever som läser teknikprogrammet i den svenska gymnasieskolan. Huruvida resultaten kan extrapoleras till en bredare grupp – såsom elever som läser övriga gymnasieprogram eller elever på gymnasieskolor i andra länder – undersöks ej inom studien och lämnar utrymme för framtida forskning.

BEGREPPSBESKRIVNING

Inom studien förekommer ett antal tvetydiga begrepp. Denna del ämnar till att konkretisera dessa begrepp för att säkerställa att forskningsfrågorna besvaras på ett entydigt sätt.

ARTIFICIELL INTELLIGENS

Begreppet AI kan vara svår att definiera. Historiskt refererade AI till sofistikerade algoritmer vilka i grund och botten bestod av en stor mängd fördefinierade instruktioner av typen *gör X om Y* (Luckin & Holmes, 2016).

År 1990 definierade Kurzweil AI som ”*konsten av att skapa maskiner som genomför funktioner som kräver intelligens när de genomförs av människor*” (Kurzweil, 1990, författarens översättning).

I dagsläget är sådana algoritmer allmänt förekommande och refereras sällan till som AI. I stället används AI i större grad för att tala om *maskininlärning* (ML) (Luckin & Holmes, 2016).

MASKININLÄRNING

ML avser en specifik gren av AI som syftar till att hitta ("lära") mönster i data med hjälp av statistiska metoder (därför även känt som statistisk inlärning) (James et al., 2023).

ARTIFICIELLA NEURALA NÄTVERK

I dagsläget är *artificiella neurala nätverk* (ANN) en ofta använd metod för ML som strävar efter att replikera neuronstrukturen av en biologisk hjärna (Worden et al., 2023). Den matematiska modellen som ligger i grund till ANN togs fram så tidigt som år 1943 och är bland de äldsta ML metoderna som finns.

De ursprungliga ANN-modellerna visade sig vara ineffektiva då de krävde stora mängder processorkraft och sällan visade bättre resultat än alternativa metoder (Worden et al., 2023). Dock hade processorkraften ökat tillräckligt mycket runt sekelskiftet för att möjliggöra så kallade *djupa* ANN med betydligt fler neuronlager och neuroner per lager. Dessa djupa ANN visade sig vara betydligt mer effektiva och i dagsläget är en av de mest frekvent använda ML-metoderna.

STORA SPRÅKMODELLER

LLM är en klass av AI som består av modeller vilka används för bearbetning av naturligt språk (Polverini & Gregorcic, 2024). Dessa används ofta med prefixet *generativ* vilket kännetecknar en modell som genererar en respons som ofta består av naturligt språk i form av text eller talsyntes, men kan även användas för att generera bild eller film utifrån en textinmatning, förklarar Polverini och Gregorcic.

Moderna LLM använder sig av en transformer-arkitektur framtagen av Google (Topal et al., 2021) och i grund och botten består av en serie av ANN. Ett sådant exempel är modellen *generative pre-trained transformer* (GPT) i programvaran ChatGPT.

Ett välkänt fenomen inom fältet av LLM är så kallade hallucinationer – tillfällen där modellen returnerar ett svar som låter korrekt, men i verkligheten är inkorrekt (Alkaiissi & McFarlane, 2023). Förekomsten av hallucinationer kan förklaras av att LLM tränas för att generera nästa ord i en ordföljd – modellerna skapar relationer mellan ord, men förstår inte deras innebörd eller betydelse.

TIDIGARE FORSKNING

Inom denna del presenteras tidigare forskning inom relaterade fält kategoriserat efter forskningens syfte och tema.

Den forskning som presenteras här fokuserar på olika sätt som AI kan utnyttjas på inom utbildning. Dock är denna forskning internationell och det finns ingen tydlig koppling till den svenska gymnasieskolan. Vidare hade den överväldigande majoriteten av forskningen genomförts innan ChatGPT hade publicerats, vilket kan ses som en vändningspunkt för användandet av AI.

Inom denna kontext kommer examensarbetet att undersöka hur AI kan implementeras i den svenska gymnasieskolan. Vidare kommer denna studie att undersöka dessa användningsområden med den nya generationen av LLM i åtanke – till skillnad från tidigare forskning.

AI-LÄSKUNNIGHET

Forskning av Ng et al. (2022) visar att modern undervisning behöver fokusera mer på AI-läskunnighet (engelska: AI-literacy). Syftet med AI-läskunnighet i denna kontext är att maximera fördelarna och minimera nackdelarna som förekommer vid användande av AI som del av lärandesituationer.

Ng et al. (2021, 2022) beskriver att historiskt har AI-läskunnighet definierats som förmågan att förstå grundläggande tekniker och begrepp som ligger bakom AI-system. På senare tid har AI-läskunnighet i stället använts för att tala om den digitala kompetensen som krävs för att förstå hur AI transformerar det sätt vi kommunicerar och arbetar med datorer, menar Ng et al.

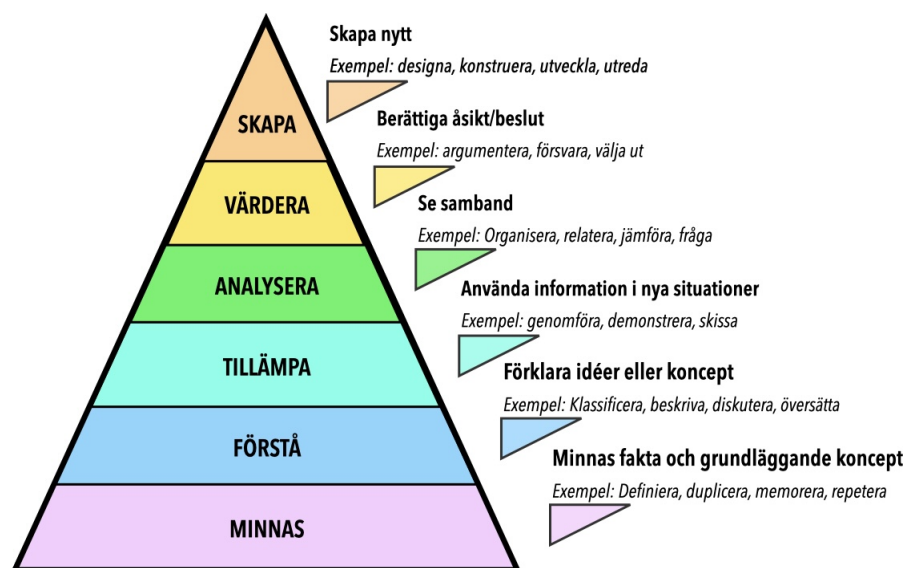
Vidare förklarar Ng et al. (2022) att AI har börjat användas i större utsträckning sedan 2016, vilket Ng et al. kopplar till att ett större utbud av digitala verktyg med inbyggd AI blivit tillgängliga. Ng et al. (2021) drar slutsatsen att elever behöver nå högre kognitiva nivåer av Blooms taxonomi (se Figur 1) för att effektivt kunna kommunicera, samarbeta och skapa med hjälp av AI. Där det tidigare var tillräcklig för en elev att förstå (nivå 2) och tillämpa (nivå 3) behöver eleven nu även analysera (nivå 4) och värdera (nivå 5).

PERSONLIGA ASSISTENTER

Jablonka et al. (2023) har visat att LLM-agenter kan anpassas till personliga assistenter som kan hjälpa eleverna med kursmaterial som lärarna förhandahåller. Modellerna kan generera instuderingsfrågor till kursmaterialet i text- eller videoformat. Vidare kan modellerna generera extra material vid de tillfällen då modellen upptäcker brister i individuella elevers kunskaper.

FIGUR 1

Blooms taxonomi med nivåförklaring (Filipsdotter, 2023).



Forskning av Zhai et al. (2021) visar att specialiserade AI-modeller kan användas av eleverna för att få omedelbar återkoppling vid lärandesituationer. Dock finns det brister med detta tillvägagångssätt, vilket leder Zhai et al. till slutsatsen att lärare måste vara en del av systemet. Lärarna behöver finnas tillgängliga för att hjälpa eleverna då programvaran inte kan, menar Zhai et al. Exempel på tillfällen där eleverna kan komma att behöva lärarstöd vid användning av LLM kan vara granskning av textutmatning - speciellt vid tillfällen då LLM-agenten hallucinerar.

Zhai et al. (2021) beskriver att inom utbildning kan en LLM-agent anpassas för att stimulera användarnas förmåga att själva komma fram till slutsatser i stället för att direkt returnera ett svar. Detta kan åstadkommas genom att modifiera självaste modellen, men även genom att användaren förser agenten med en kedja av argument i stället för en direkt fråga, förklarar Zhai et al. Även detta är ett exempel där eleverna kan komma att behöva lärarstöd för att formulera inmatningen på ett sätt som prioriterar lärande över att eleverna får ett svar utan att tänka själva.

Cain (2023) menar i stället att existerande LLM-agenter såsom OpenAIs ChatGPT eller Googles Gemini (tidigare Bard) är tillräckligt lättillgängliga att de med relativ enkelhet kan integreras i existerande skolplattformar. Denna teori håller på att testas i Norge där skolmyndigheten i Oslo har tecknat ett avtal för 110 000 GPT-3.5-Turbo licenser för användning av såväl elever som lärare (Bjørkeng, 2024).

Dock beskriver Zhai et al. (2021) risken att eleverna får hjälp med problem eller svar på frågor utan att själva förstå det underliggande resonemanget. Konsekvensen av denna risk är att eleverna inte lär sig lika mycket som om de hade använt klassiska metoder, vilket är varför läraren behöver fortfarande vara närvarande förklarar Zhai et al.

METOD

Denna studie hade i grund och botten följt en kvasi-experimentell design med en innan och efter jämförelse av en strukturerad pedagogisk intervention. Forskningsmetodiken bestod av totalt två enkäter, tre lektioner och fyra semistrukturerade intervjuer.

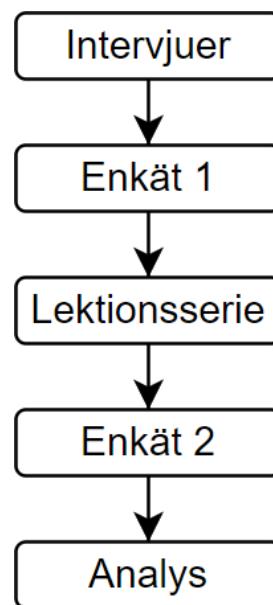
Samtliga enkäter och lektioner genomfördes på en gymnasieskola i Stockholmsområdet med höga socioekonomiska förhållanden. Enkäterna och lektionerna genomfördes för totalt ungefär 60 unika gymnasieelever från två klasser på teknikprogrammet. Det exakta antalet unika elever som tog del av lektionsserien är okänd eftersom det skedde en ej insignifikant mängd av rörelse in och ut ur klassrummet under lektionerna. Intervjurespondenterna bestod av teknicklärarna på den deltagande skolan. Enkäten besvarades av de närvarande eleverna, inget annat utskick eller urval gjordes. Urvalet av dessa gymnasieelever gjordes beroende på deras anslutning till den uppdragsgivande läraren.

Enkäterna och lektionerna genomfördes under skoltid schemalagd för övrig aktivitet under totalt tre veckor – en lektion per vecka samt en enkät innan den första lektionen och en enkät efter den sista lektionen. Tre av de fyra semistrukturerade intervjuerna genomfördes innan lektionsserien och den fjärde och sista semistrukturerade intervjun genomfördes efter lektionsserien då läraren var ej tillgänglig vid tidigare datum.

Notera att en minoritet av eleverna missade uppåt 30 minuter av lektionstiden vardera på grund av pågående utvecklingssamtal. Det exakta antalet elever och den exakta tiden per elev är okänd då ingen kontroll skedde vid in- och utgång. Dessutom finns det ingen garanti att alla som svarade på den första enkäten svarade även på den andra eller vice versa. Detta kan ha en negativ effekt på innan och efter jämförelsens validitet, men denna effekt hade inte kvantifierats inom studien.

Insamlade data analyserades med hjälp av *frekvensanalys*, *deskriptiv statistisk analys* och *explorativ faktoranalys* (EFA). Frekvensanalysen genomfördes genom att beräkna frekvensen av olika responser för att skapa en bättre förståelse av deras fördelning och relation. Den deskriptiva statistiska analysen i stället sammanfattar responserna med hjälp av medelvärde, standardavvikelse och förändring mellan den första och den andra enkäten. Den *explorativa* faktoranalysen valdes i stället för den *bekräftande* faktoranalysen. Detta med mål att generera en ny hypotes i stället för att bekräfta en existerande hypotes. Analysens övergripande syfte var att hitta samband samt utvärdera lektionsseriens effektivitet för att vidareutveckla dess struktur och innehåll.

FIGUR 2
Metodens flödesschema.



INTERVJU

De semistrukturerade intervjuerna var en sekundär del av forskningsmetodiken och syftade till att fördjupa förståelsen för lärarnas perspektiv, tidigare erfarenheter, förväntningar och attityder kring användningen av AI som ett pedagogiskt verktyg med mål att anpassa enkäternas och lektionernas innehåll efter skolan och klasserna.

Vidare syftade intervjuerna till att minimera risken att enkätfrågorna började skrivas för tidigt samt att innehållet och därmed resultatet visade sig vara irrelevant för studien (Persson, 2016a). I denna mån användes responserna som en grund för skapande av enkätfrågorna.

Intervjumodellen som användes är framtagen av Kallio et al. (2016) och delar upp frågor i två nivåer. Den första nivån består av frågor som ligger nära studiens ämne och är formulerade på ett sätt som uppmanar respondenten till att tala fritt. Den andra nivån består av uppföljningsfrågor som i sin tur kan vara förberedda eller spontana. Båda typerna av uppföljningsfrågor syftar till att låta respondenten fördjupa sig i ett ämne, beskriver Kallio et al.

Intervjufrågorna formulerades i enlighet med rekommendationer av Gideon (2012) och Persson (2016a) (för frågetext se Tabell 1). Kallio et al. (2016) fortsätter genom att rekommendera expertbedömning av frågorna med syfte att förbättra formuleringen och relevansen. Expertbedömningen gjordes av handledaren från KTH, vilket har varit till stor fördel vid förberedelse av intervjufrågorna.

Enligt rekommendation av Gideon och Moskos (2012) antecknades svaren på frågorna i realtid i stället för att spelas in och transkriberas. Gideon och Moskos menar att det i praktiken dels sparar tid och dels gör informationen lättillgänglig vid skapande av följdfrågor under självaste intervjun.

ENKÄT

De deltagande eleverna besvarade totalt två enkäter – en innan och en efter genomförandet av lektionsserien. Den andra enkäten inkluderade samtliga frågor från den första enkäten med mål att jämföra med den andra enkäten med den första för att kvantifiera lektionsseriens effektivitet.

Målgruppen för enkäterna var gymnasieeleverna från de två deltagande klasserna och endast de närvarande eleverna hade möjligheten att lämna in responser. Enkäterna implementerades med hjälp av programvaran Google Forms och distribuerades genom att en förkortad länk visades på projektorn. Av etiska skäl anonymiserades enkätsvaren och analyserades på gruppnivå i stället för individnivå. Vidare efterfrågade enkäterna inte elevernas personliga information och inställningarna ställdes in så att inget Google-konto behövdes för att lämna in.

TABELL 1

Lista av frågor som användes vid samtliga semistrukturerade intervjuer inklusive förberedda uppföljningsfrågor.

	#	Frågetext
Intro	1	Hur länge har du arbetat som lärare?
	2	Vilka ämnen arbetar du med för närvarande?
Läraren	3	Hur bekant är du med artificiell intelligens, maskininlärning och stora språkmodeller?
	4	Har du personligen använt AI på senaste?
	4.1	Vilken effekt hade ditt användande av AI på din produktivitet anser du?
	5	Från ditt perspektiv, vilka för- och nackdelar finns det för användning av AI inom gymnasieskolan?
	5.1	Anser du att AI kan användas med fördel inom någon av dina ämnen?
Eleverna	6	Vilken erfarenhet har du av att dina elever använder sig av AI?
	6.1	Från ditt perspektiv, hur har dina elevers användning av AI påverkat din arbetsprocess?
	7	Anser du att gymnasieelever ska undervisas om hur AI kan användas som ett verktyg?
	7.1	Från ditt perspektiv, finns det något specifikt för dina ämnen som sådan utbildning bör omfatta?
Skolan	8	Hur arbetar skolan med AI?
	9	Vilka anpassningar har skolan gjort med AI i åtanke?
	10	Vad är ditt perspektiv på att Skolverket ska introducera AI som en egen kurs?

Enligt Gideon (2012) är det viktigt att introducera respondenten till ämnet, vilket är anledningen till att enkätens syfte och centrala begrepp beskrevs för respondenterna innan enkäterna gjordes tillgängliga. Vidare rekommenderar Gideon att använda välformulerade och tydliga svarsalternativ, vilket är anledningen till att enkäterna utnyttjade SCB:s (Persson, 2016b) standardiserade likertskala samt möjligheten att inte svara på de frågor de inte vill:

Stämmer helt Stämmer bra Inget svar Stämmer dåligt Stämmer inte alls

Innan enkätfrågorna formulerades skapades en lista av centrala begrepp och undersökningsvariabler i enlighet med rekommendationer av Gideon (2012). Därefter ordnades listan manuellt på ett godtyckligt sätt och frågor skapades för varje element i listan med syfte att säkerställa en röd tråd och tydlig koppling till studien syfte.

Även enkätfrågorna formulerades i enlighet med rekommendationer av Gideon (2012) och Persson (2016a) inklusive valet av frågeord samt undvikande av ledande frågor, dubbelfrågor och frågor med dubbla negationer (för frågetext se Tabell 2).

TABELL 2

Lista av frågor som användes i både enkäten innan och enkäten efter genomförandet av lektionsserien samt typ av svarsalternativ.

	#	Frågetext	Svarstyp
Vanor	1	Hur många dagar har du använt språkmodeller...	0–14
	2	Vilka generativa språkmodeller har du använt...	Fritext
	3	I vilka syften har du använt generativa språkmodeller... ...de senaste två veckorna?	Fritext
Praktik		Jag tycker att generativa språkmodeller...	
	4	...ger mig korrekt information.	Likert
	5	...hjälp mig bli klar fortare.	Likert
	6	...förbättrar kvalitén av mina arbeten.	Likert
	7	...har hjälpt mig förbättra mina skolresultat.	Likert
	8	...hjälp mig lösa problem jag annars inte kunde lösa.	Likert
	9	...löser mina problem åt mig.	Likert
Teknik		Jag kan beskriva hur generativa språkmodeller...	
	10	...skiljer sig från andra typer av artificiell intelligens.	Likert
	11	...fungerar från användarperspektiv.	Likert
	12	...fungerar på teknisk nivå.	Likert
Kritik	13	Jag känner mig säker när jag granskar...	Likert
	14	Jag kan särskilja mellan människoskriven och... ...datorgenererad text.	Likert
Intresse	15	Jag är intresserad av att lära mig mer om...	Likert
	16	Jag tycker att gymnasieskolor bör satsa mer på... ...generativa språkmodeller.	Likert

De tre första frågorna syftade till att undersöka hur elevernas användningsvanor ser ut och om lektionsserien hade en påverkan på dessa vanor. Med användningsvanor menas i detta sammanhang hur ofta eleverna använder sig av generativ AI, vilka program de använder sig av samt med vilket syfte.

Därefter följde sex frågor vilka syftade till att i stället undersöka elevernas praktiska färdigheter gällande användning av generativ AI. Med praktiska färdigheter menas i detta sammanhang hur eleverna använder sig av generativ AI, vilka resultat de förväntar sig av deras användande samt vilken påverkan deras användande har på bland annat deras förmåga att lösa problem.

De följande tre frågorna syftade till att kvantifiera elevernas tekniska kunskaper. Med teknisk kunskap menas i detta sammanhang elevernas förmåga att beskriva och förklara hur LLM fungerar utifrån olika perspektiv och hur dessa skiljer sig från andra typer av AI.

Vidare syftade de två följande frågorna till att skapa en förståelse av elevernas kritiska tänkande. Slutligen syftade de två sista frågorna till att kvantifiera elevernas intresse för att lära sig mer om AI.

ANALYS

Eftersom majoriteten av enkätfrågorna följer likertskalan lämpar enkätsvaren sig till numerisk analys (Boone & Boone, 2012). Responserna analyserades därför med fokus på tendens och varians med hjälp av frekvensanalys, deskriptiv statistisk analys och EFA.

Först och främst analyserades enkätsvaren med hjälp av frekvensanalys. Denna analysmetod bygger på att undersöka frekvensen av svarsalternativen till individuella frågor med mål att kvantifiera tendensen (Clason & Dormody, 1994). Svarsfrekvensen visualiserades med hjälp av horisontella stapeldiagram, cirkeldiagram och tabeller (Batterton & Hale, 2017). För att avsluta frekvensanalysen beräknades även p -värden med hjälp av Mann–Whitney U-testet (MacFarland & Yates, 2016). Mann–Whitney U-testet genomfördes med hjälp av funktionen `mannwhitneyu` från modulen `stats` inom biblioteket `scipy` för programmeringsspråket Python 3. Valet av Mann–Whitney U-testet gjordes huvudsakligen eftersom den underliggande datan är ej normalfördelad och provstorleken skiljer sig i storlek.

Deskriptiv statistisk analys genomfördes med mål att kvantifiera variansen genom att beräkna medelvärdet och standardavvikelsen (Alkharusi, 2022). Därefter jämfördes svaren mellan enkäten innan och efter genomförandet av lektionsserien för att beräkna den genomsnittliga förändringen.

Inom EFA användes metoden *Principal Axis Factoring* (PAF) för att hitta samband mellan enkätfrågor genom att undersöka laddningar, likheter och egenvärden (Batterton & Hale, 2017; Goretzko et al., 2021). Analysen genomfördes med hjälp av biblioteket `factor_analyzer` för programmeringsspråket Python 3 enligt rekommendation av Batterton och Hale. Statistik från en metaanalys av EFA inom modern forskning var till stöd vid val av `promax` som rotationsmetod, Kaiser–Guttman som kriterium för behållning av faktorer och 0.5 som tröskel för behållning av likheterna (Goretzko et al., 2021).

LEKTIONSSERIE

Mellan de två enkäterna genomfördes totalt tre lektioner. Samtliga lektioner genomfördes av författaren och elevernas fokusnivåer vid interaktiva moment dokumenterades genom observationsanteckningar.

Lektionsseriens huvudsakliga syfte är att undervisa eleverna om AI-läskunnighet enligt definition av Ng et al. (2022). För att uppnå detta syfte hade lektionsserien strukturerats utefter ett ramverk för undervisning om AI-läskunnighet framtagen av Long och Magerko (2020). Long och Magerkos ramverk består huvudsakligen av en lista av kompetenser för AI-läskunnighet som bör återspegla lektionsseriens innehåll (se Tabell 3) och en lista av överväganden som bör tas i åtanke vid skapande av undervisningsmaterial (se Tabell 4).

TABELL 3

Kompetenser för AI-läskunnighet (Long & Magerko, 2020, författarens översättning). För längre beskrivning se källa.

#	Kompetens	#	Kompetens
1	Igenkänning av AI	10	Mänsklig roll i AI
2	Förståelse av AI	11	Data-läskunnighet
3	Tvärvetenskaplighet	12	Inläring från data
4	Generell och snäv AI	13	Kritisk tolkning
5	AI:s styrkor och svagheter	14	Handling och reaktion
6	Framtida AI	15	Sensorer
7	Kunskapsrepresentationer	16	Etik
8	Beslutsfattande	17	Programmerbarhet
9	Maskininlärningsprocess		

TABELL 4

Överväganden vid förberedelse av material för undervisning av AI-läskunnighet (Long & Magerko, 2020, författarens översättning). För längre beskrivning se källa.

#	Övervägande	#	Övervägande
1	Förklararbarhet	9	Identitet
2	Konkretisering	10	Föräldrastöd
3	Kontextualisering	11	Socialisering
4	Transparens	12	Elevernas intressen
5	Gradvis avslöjning	13	Erkännande av fördomar
6	Programmeringsmöjligheter	14	Främjande av nya perspektiv
7	Milstolpar	15	Sänkning av inträdesbarriären
8	Kritiskt tänkande		

Presentationsmodulerna och deras innehåll bestämdes på ett sådant sätt att samtliga 17 kompetenser omfattades. Vidare togs 12 av de 15 övervägandena i åtanke vid förberedelse av presentationsbilder. De tre kvarvarande överväganden uteslöts eftersom:

- Övervägande 7 syftar främst till undervisning av yngre barn och utelämnades då målgruppen var äldre barn.

- Övervägande 10 utelämnades då lektionsserien var relativt kort och det var opraktiskt att introducera föräldrarna i projektet.
- Övervägande 12 handlar om att utnyttja elevernas intressen för att motivera deras lärande och utelämnades eftersom författaren saknade den förutsatta sociala relationen.

Vidare anpassades lektionsseriens innehåll utefter förslag från en undersökning av Long och Magerkos ramverk (Zhou et al., 2020). Den viktigaste anpassningen var införandet av praktiska moment, vilket Zhou et al. rekommenderar för att stimulera elevernas engagemang och lärande.

Samtliga lektioner genomfördes i en större sal där båda klasserna kunde närvara samtidigt. De tre lektionerna omfattade fyra moduler vardera med en kort paus mellan de två första och två sista modulerna (se Tabell 5 för beskrivning av presentationsmoduler samt relaterade kompetenser ur Tabell 3 och överväganden ur Tabell 4). Om inget annat anges genomfördes presentationsmodulen i stil av en föreläsning med informativa presentationsbilder som visades med hjälp av två projektorer.

TABELL 5

Presentationsmoduler samt kompetenser de representerar och överväganden som hade tagits i åtanke vid skapade av presentationsbilder.

	#	Presentationsmodul	Kompetenser	Överväganden
Lektion 1	1	Artificiell intelligens	1, 3, 4	5, 15
	2	Maskininlärning	2, 8, 9	1, 4, 5
	3	Neurala nätverk	2, 8, 9	1, 4, 5
	4	Inlärning från data	7, 11, 12, 15	3, 4, 9, 13
Lektion 2	5	Generativ AI	1, 7, 12	1, 2, 4, 5
	6	Språkmodeller	2, 8, 17	1, 2, 4, 5
	7	Styrkor och svagheter	5, 15	14
	8	Praktiskt moment	14, 17	2, 6, 11, 15
Lektion 3	9	Datorgenererad och människoskriven text	1, 11, 13	8
	10	Tolkning och resonering	10, 13	8, 11
	11	Etik	10, 16	9, 11, 13
	12	AI i framtiden	3, 6	14

Den första lektionen var en renodlad föreläsning där eleverna introducerades till den ämnesspecifika teoretiska kunskapen. Detta med syfte att ge eleverna den teoretiska grund som behövs för att aktivt delta i de diskussioner som förekommer under de två följande lektionerna.

Som del av den andra lektionens sista presentationsmodul (praktiskt moment) fick samtliga elever möjligheten att i mindre grupper (1) använda flera olika språkmodeller och (2) skapa egna agenter med hjälp av de tillgängliga grundmodellerna.

De språkmodeller som användes under de två övningarna var hostade på den deltagande gymnasieskolans egna servrar och ingen användardata lämnade skolans system. Detta var ett initiativ från den uppdragsgivande läraren på den deltagande gymnasieskolan.

Under den tredje lektionens första presentationsmodul (datorgenererad och människoskriven text) fick eleverna möjligheten att i mindre grupper jämföra beskrivningar av (1) artificiell intelligens, (2) maskininlärning och (3) stora språkmodeller genererade med hjälp av ChatGPT mot det första stycket från respektive Wikipediaartikel. Den tredje presentationsmodulen (etik) bestod av gruppdiskussioner om (1) etik, (2) bias och (3) plagiering.

Vidare antecknades antalet elever som fokuserade på telefoner och datorer under de åtta tidigare nämnda övningarna för att kvantifiera andelen av elever som är ofokuserade. Detta med syfte att undersöka om eleverna är mer eller mindre fokuserade på något av delmomenten med mål att vidareutveckla de delar där eleverna är minst fokuserade.

ETISKA ÖVERVÄGANDEN

Stora delar av examensarbetet genomfördes på en gymnasieskola med stöd av ett samarbetsprojekt för praktikinära forskning. Därför behövdes åtgärder vidtas för att säkerställa lärarnas, elevernas, forskarnas och organisationernas integritet (Oldendick, 2012).

Då det förekommer stora skillnader mellan gymnasieskolor beroende på ort och socioekonomiskt välstånd är det dock viktigt att inkludera geografisk information, vilket är skälet till att gymnasieskolan anonymiserades, men inte dess ort. Lärare anonymiserades på individnivå och benämndes *Lärare 1*, *Lärare 2* och så vidare. Eleverna var fler i antal och anonymiserades i stället på gruppnivå.

Vidare informerades respondenterna att deras deltagande är frivilligt, att de inte behöver svara på de frågor de inte vill och att de kan dra tillbaka sitt samtycke när som helst (Oldendick, 2012).

Inom detta examensarbete har ChatGPT-3.5 använts för att översätta tekniska termer från engelska till svenska. AI har inte använts vid något annat tillfälle.

RESULTAT

Inom denna del presenteras data kategoriserat utefter den metod som användes vid insamling av datan. Data presenteras på ett transparent och neutrallt sätt – analys och tolkning av den data som presenteras återfinns inom diskussionsavsnittet.

INTERVJU

Inom denna avsnitt kommer resultatet från de fyra semistrukturerade intervjuerna att presenteras fördelade utefter om svaren handlar om *lärarna*, *eleverna* eller *skolan*.

LÄRARNAS ERFARENHETER

Generellt såg de fyra intervjuade lärarna utvecklingen inom AI som något positivt. Alla lärare hade tidigare använt sig av AI vid förberedelse av lektionsmaterial. Lärare 2 och 4 hade läst en respektive två universitetskurser om ML.

Lärare 1 hade sagt att *"AI är bra för att snabbt sätta sig in i något man själv inte är expert på."* Lärare 4 i stort höll med lärare 1 men lade till att *"man måste ställa sig kritisk"* och att *"det är svårt att veta vad som är rätt."* Vidare förklarade lärare 3 att *"i teknik får man användbara svar, men i matematik blir det helt fel."*

ELEVERNAS ANVÄNDNING

En stor fördel med verktyg som ChatGPT är att *"eleverna kan snabbt få ihop någonting"* förklarade lärare 1. Eleverna ofta får en färdig text av ChatGPT utan att behöva tänka själva, fortsatte lärare 1. Lärare 1 menar att detta inte alltid är onskvärt och att det vore föredraget om eleverna i stället fick konstruktiv återkoppling, synpunkter och rekommendationer i stället för lösningar.

Lärare 2 förklarade att det ofta händer att elever använder sig av ChatGPT och får svar som låter rätt, men i verkligheten är helt fel. Lärare 4 påstod att eleverna *"använder ChatGPT för att generera kod som de inte förstår,"* men detta påstående behöver inte nödvändigtvis vara applicerbar till endast kodgenerering.

Lärare 1 beskriver att det finns en stor skillnad i hur eleverna använder sig av AI. Många av eleverna undviker att använda sig av AI överhuvudtaget, förklarar lärare 1. Lärare 4 menar att *"lärare känner igen om en elev använder för mycket AI vid inlämningar."*

AI PÅ SKOLAN

Samtliga lärare bekräftar att skolan inte har infört några AI-anpassningar på organisationell nivå. Lärare 2 beskriver att denne inte längre använder sig av inlämningsuppgifter i form av längre texter och i stället använder sig av fler mindre uppgifter. Dock påpekar lärare 2 att denne inte genomfört någon analys och grundar sina Anpassningar i antaganden - *"det är möjligt att de [fortfarande] använder sig av ChatGPT."* I stället för att anpassa sina uppgifter fokuserar lärare 4 mer på att *"stämna av med eleverna att de vet vad de gjort."* Lärare 2 tycker att skolan borde skapa tydliga regler för vad som är och inte är tillåtet - *"eleverna kommer att använda AI oavsett och det är bäst att reglera [användningen]."*

Samtliga lärare tyckte att eleverna ska undervisas om hur de kan använda sig av AI som ett verktyg med fokus på etik och kritik och inte bara som ett tekniskt ämnesområde. Lärare 4 förklarar att ”AI är otroligt teoretiskt och komplext” och att denne är osäker hur en sådan utbildning kan struktureras för att hållas på en rimlig nivå för eleverna.

ENKÄT

Som del av studien genomfördes två enkäter besvarade av två klasser av gymnasieelever som alla läser det andra året på teknikprogrammet. Inom studien refereras de två enkäterna utefter om de genomfördes *innan* eller *efter* genomförandet av lektionsserien. De två enkäterna bestod av samma frågor (för frågetext se Tabell 2) och genomfördes både innan och efter lektionsserien med mål att jämföra resultaten mot varandra.

Vid enkättillfällerna antecknades antalet närvarande elever, vilket användes tillsammans med det totala antalet enkätsvar för att beräkna den effektiva responsfrekvensen (se Tabell 6).

TABELL 6

Antalet enkätsvar, antalet närvarande elever och den beräknade effektiva responsfrekvensen.

Enkät	Enkätsvar	Elever	Responsfrekvens
Innan	40	53	75.5%
Efter	28	49	57.1%

Notera att det minskade antalet enkätsvar vid det andra enkättillfället inte kan beskrivas med det minskade antalet närvarande elever. Effektivt hade responsfrekvensen minskat med 24.4% mellan den första och andra enkäten. Med andra ord hade ungefär en fjärdedel av de som svarat på den första enkäten valt att inte svara på den andra enkäten.

FREKVENSPANALYS

Målet med frekvensanalysen var att kvantifiera tendensen genom att jämföra svarsfrekvenserna mellan de två enkäterna. Notera att eleverna blivit informerade om studiens syfte och vad olika nyckelbegrepp såsom *språkmodeller* betyder innan de besvarade enkäten.

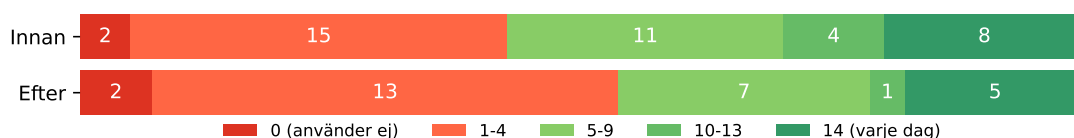
Mann–Whitney U-testet genomfördes för samtliga frågor för att beräkna p -värdet mot nollhypotesen H_0 att skillnaden mellan den första och andra enkäten inte är signifikant och den alternativa hypotesen H_1 att skillnaden är signifikant där $p \geq 0.05$ bekräftar H_0 och $p < 0.05$ förkastar H_0 .

Svarsalternativen till de tre första frågorna skiljer sig från de resterande 13 frågorna då de inte befinner sig på likertskalan. Dessa tre frågor syftar till att kvantifiera elevernas användningsvanor. De beräknade p -värden är 0.383, 0.599 och 0.742 för respektive fråga, vilket är ett tydligt tecken att lektionsserien inte hade en signifikant påverkan på elevernas användningsvanor.

Det är viktigt att påpeka hur det absoluta mängden av svar hade varierat stort mellan den första och den andra enkäten även fast förändringen är statistiskt insignifikant. Exempelvis hade 33 och 24 elever sagt att de hade använt sig av ChatGPT de senaste två veckorna vid första respektive andra enkäten. Minns dock att den andra enkäten fick betydligt färre responser än den första. I praktiken är det viktigare att fokusera på förändringen i andelen svar, vilket visualiseras av Figur 3, 4, 5.

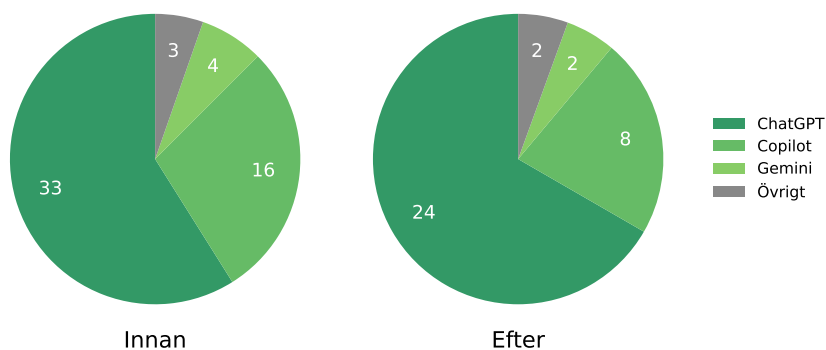
FIGUR 3

Svarsfrekvenserna till den första frågan: ”Hur många dagar har du använt språkmodeller de senaste två veckorna?” $p = 0.383$.



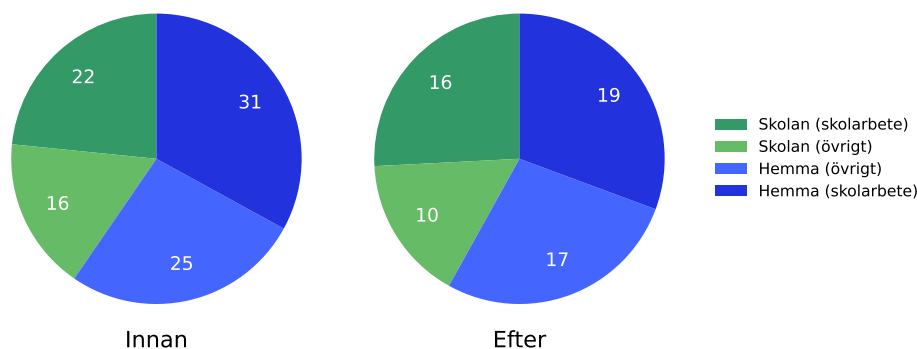
FIGUR 4

Svarsfrekvenserna till den andra frågan: ”Vilka generativa språkmodeller har du använt de senaste två veckorna?” Fler än ett svar var tillåtet per respondent. $p = 0.599$.



FIGUR 5

Svarsfrekvenserna till den tredje frågan: ”I vilka syften har du använt generativa språkmodeller de senaste två veckorna?” Fler än ett svar var tillåtet per respondent. $p = 0.742$.



Tabell 7 sammanfattar svarsfrekvenserna för både enkäten innan och enkäten efter genomförandet av lektionsserien. Data struktureras inom två kluster – en för vardera enkät – med fem kolumner vardera där den vänstra kolumnen motsvarar antalet elever som valde svarsalternativet *stämmer helt* och den högra kolumnen motsvarar antalet elever som valde svarsalternativet *stämmer inte alls*. Den enstaka kolumnen till höger om de två klustrerna representerar p -värdet beräknat med hjälp av Mann–Whitney U-testet.

TABELL 7

Svarsfrekvenserna för frågor med svar på likertskalan innan (vänster kluster) och efter (höger kluster) genomförandet av lektionsserien. Dessutom återfinns p -värden till höger om de båda klustrerna. För frågetext se Tabell 2.

#	Innan					Efter					p
	Helt	↔	Inte	Helt	↔	Inte					
4	1	34	1	3	1	0	20	1	7	0	0.084
5	23	16	1	0	0	11	11	1	4	1	0.042
6	6	20	1	12	1	5	16	3	3	1	0.336
7	5	22	1	8	4	2	17	3	3	3	0.900
8	17	15	1	4	3	7	15	2	3	1	0.316
9	6	15	1	15	3	2	13	3	9	1	0.899
10	5	12	1	16	6	5	19	1	2	1	0.002
11	6	21	0	9	4	7	19	1	1	0	0.019
12	2	14	0	14	10	6	14	1	5	2	0.002
13	15	20	0	5	0	2	21	2	3	0	0.020
14	8	17	0	13	2	0	21	2	5	2	0.939
15	17	21	0	2	0	9	14	2	3	0	0.192
16	16	18	0	7	0	6	15	2	4	1	0.192

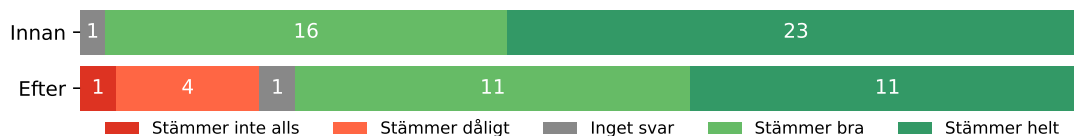
STATISTISKT SIGNIFIKANTA FRÅGOR

Utifrån p -värdena (se Tabell 7) är det tydligt att lektionsserien hade en statistiskt signifikant påverkan endast för gymnasieelevernas svar på frågorna 5, 10, 11, 12 och 13.

Endast fråga 5 (se Figur 6) – en av de två frågorna som syftade till att kvantifiera elevernas praktiska färdigheter – visade på en statistiskt signifikant förändring. Observera att innan lektionsserien ansåg *ingen* elev att generativa LLM ger dem inkorrekt information. Efter lektionsserien svarade några av eleverna att LLM ger dem inkorrekt information och en mindre andel ansåg att LLM ger dem korrekt information. Detta kan indikera att eleverna överskattade sina praktiska färdigheter.

FIGUR 6

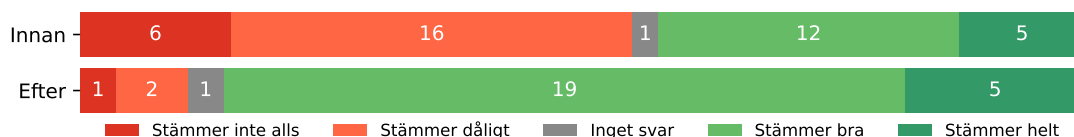
Svarsfrekvenserna till den femte frågan: ”*Jag tycker att generativa språkmodeller ger mig korrekt information.*” $p = 0.042$.



Samtliga av de tre tekniskt inriktade frågorna visade en positiv utveckling i elevernas självuppfattade kunskapsnivåer (se Figur 7, 8 och 9). Därför finns det ingen tvekan om att lektionsserien lyckades utveckla gymnasieelevernas tekniska kunskaper om hur bland annat LLM fungerar.

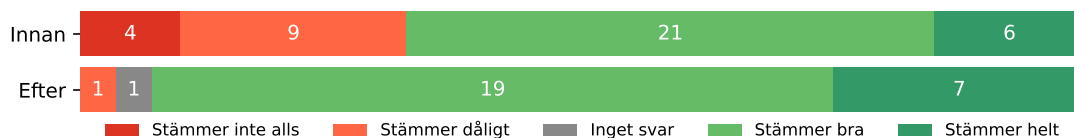
FIGUR 7

Svarsfrekvenserna till den tionde frågan: ”*Jag kan beskriva hur generativa språkmodeller skiljer sig från andra typer av artificiell intelligens.*” $p = 0.002$.



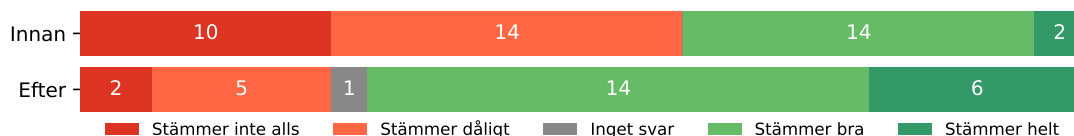
FIGUR 8

Svarsfrekvenserna till den elfte frågan: ”*Jag kan beskriva hur generativa språkmodeller fungerar från användarperspektiv.*” $p = 0.019$.



FIGUR 9

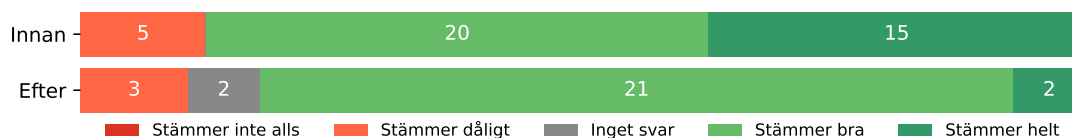
Svarsfrekvenserna till den tolfte frågan: ”*Jag kan beskriva hur generativa språkmodeller fungerar på teknisk nivå.*” $p = 0.002$.



Enkäten inkluderade två frågor som syftade till att undersöka gymnasieelevernas kritiska ställning till AI. Av dessa visade endast fråga 13 (se Figur 10) en statistiskt signifikant förändring. Både innan och efter genomförandet av lektionsserien ansåg en överväldigande majoritet av eleverna sig vara säkra när de granskar datorgenererad text. Efter lektionsserien hade en majoritet av de som svarat *stämmer helt* i stället valt att svara *stämmer bra*. Detta visar att eleverna fortfarande känner sig säkra på att granska, men att de inte längre överskattar sin förmåga.

FIGUR 10

Svarsfrekvenserna till den trettonde frågan: ”Jag känner mig säker när jag granskar datorgenererad text.” $p = 0.020$.



STATISTISK ANALYS

Målet med den deskriptiva statistiska analysen (se Tabell 8) var att kvantifiera tendensen och variansen genom att jämföra medelvärden och standardavvikelsen mellan svaren till frågorna ur de två enkäterna.

För att genomföra den deskriptiva statistiska analysen behövde likertskalan översättas till en ekvidistant numerisk skala där *stämmer inte alls* motsvarar det minsta värdet 1 och *stämmer helt* motsvarar det största värdet 5.

Den genomsnittliga förändringen beräknades för varje fråga genom att subtrahera medelvärdet ur enkäten innan lektionsserien från medelvärdet ur enkäten efter lektionsserien. Positiv förändring innebär att gymnasieeleverna har valt positiva svarsalternativ mer frekvent vid det andra enkättillfället än vid det första och vice versa för negativ förändring.

Dessutom har förändringen färgkodats beroende på om förändringen är positiv eller negativ. Om frekvensanalysen visade att lektionsserien inte hade en statistiskt signifikant påverkan på svaren till frågan (se Tabell 2) har färgen i stället tonats ned.

FAKTORANALYS

EFA är en explorativ analysmetod som syftar till att identifiera en okänd faktorstruktur och generera hypoteser.

EFA producerade en likhetsvektor för samtliga frågor, en laddningsvektor per faktor och en egenvärdesvektor för faktorerna. Enligt Kaiser–Guttman-kriteriet har faktorer med egenvärden $\lambda < 1$ förkastats. På samma sätt har även frågor med likheter < 0.5 förkastats. De kvarstående frågorna och faktorerna användes vid en sekundär EFA för att identifiera den slutgiltiga faktorstrukturen.

EFA för den första enkäten (se Tabell 9) visar två faktorer vilka tydliggör en stark relation mellan frågorna 5, 8, 13, 15 och 16 samt en svag relation mellan frågorna 4, 7, 9, 10, 11, 12, och 14.

Även EFA för den andra enkäten (se Tabell 10) visar två faktorer vilka i stället tydliggör en stark relation mellan frågorna 4, 6, 7, 9, 10, 13 och 14 samt en svag relation mellan frågorna 5, 8, 11, och 15.

TABELL 8

Medelvärde och standardavvikelse innan (vänster) och efter (center) genomförande av lektionsserien samt den resulterande förändringen i medelvärde (höger). För frågetext se Tabell 2.

#	Innan		Efter		Förändring
	Medelvärde	Std.avv.	Medelvärde	Std.avv.	
4	3.775	0.733	3.464	0.811	-0.311
5	4.550	0.552	3.964	1.170	-0.586
6	3.450	1.154	3.750	1.005	+0.300
7	3.400	1.236	3.429	1.136	+0.029
8	3.975	1.250	3.857	1.044	-0.118
9	3.150	1.292	3.214	1.101	+0.064
10	2.850	1.350	3.893	0.916	+1.043
11	3.400	1.277	4.143	0.651	+0.743
12	2.600	1.336	3.607	1.227	+1.007
13	4.125	0.939	3.786	0.738	-0.339
14	3.400	1.277	3.571	0.790	+0.171
15	4.325	0.730	4.036	0.922	-0.289
16	4.025	1.050	3.750	1.076	-0.275

TABELL 9

Primär EFA (vänster) och sekundär EFA (höger) för frågor med svar på likertskalan för enkäten *innan* genomförandet av lektionsserien. För frågetext se Tabell 2.

#	Laddningar					#	Laddningar		
	1	2	3	Likheter			1	2	Likheter
4	0.044	0.176	1.146	1.346		4	0.393	0.547	0.544
5	1.136	0.223	0.145	1.363		5	1.119	0.383	1.401
6	0.186	0.547	0.389	0.486		6			
7	0.062	0.179	0.822	0.712		7	0.359	0.710	0.634
8	1.049	0.021	0.093	1.110		8	1.032	0.135	1.083
9	0.116	0.989	0.069	0.997		9	0.020	0.944	0.891
10	0.097	1.217	0.288	1.575	→	10	0.278	1.021	1.121
11	0.148	0.321	0.643	0.539		11	0.367	0.744	0.689
12	0.407	0.838	0.264	0.938		12	0.397	1.078	1.320
13	0.830	0.066	0.197	0.732		13	0.918	0.150	0.866
14	0.346	0.785	0.023	0.737		14	0.300	0.795	0.722
15	0.874	0.132	0.265	0.853		15	1.006	0.009	1.012
16	0.886	0.245	0.014	0.845		16	0.877	0.191	0.806
λ	9.249	2.888	0.646			λ	8.347	2.824	

TABELL 10

Primär EFA (vänster) och sekundär EFA (höger) för frågor med svar på likertskalan för enkäten *efter* genomförandet av lektionsserien. För frågetext se Tabell 2.

Laddningar					Laddningar				
#	1	2	3	Likheter	#	1	2	Likheter	
4	0.442	0.047	0.644	0.613	4	1.063	0.091	1.139	
5	0.262	1.199	0.026	1.507	5	0.289	1.139	1.381	
6	0.710	0.250	0.100	0.577	6	0.622	0.440	0.582	
7	1.012	0.168	0.132	1.070	7	0.921	0.071	0.854	
8	0.495	0.538	0.036	0.537	8	0.355	0.708	0.628	
9	0.099	0.043	1.049	1.114	9	0.985	0.141	0.990	
10	0.747	0.255	0.050	0.627	10	0.596	0.465	0.572	
11	0.679	0.433	0.067	0.653	11	0.392	0.668	0.599	
12	0.269	0.508	0.326	0.437	12				
13	0.786	0.016	0.238	0.675	13	0.862	0.173	0.773	
14	0.679	0.104	0.450	0.676	14	1.032	0.039	1.068	
15	0.234	0.811	0.011	0.712	15	0.069	0.948	0.904	
16	0.434	0.455	0.209	0.440	16				
λ	11.614	1.055	0.291		λ	9.663	1.039		

DISKUSSION

Inom denna avsnitt kommer resultatet att analyseras och diskuteras utifrån relevanta pedagogiska teorier och tidigare forskning. Detta med mål att sätta resultatet i ett bredare sammanhang för att besvara forskningsfrågorna och möta studiens syfte.

Diskussionen kommer först att presentera, resonera och analysera svarsfrekvensen för de sex frågorna med statistiskt signifikant förändring mellan den första och andra enkäten. Detta eftersom de ligger till grund för den följande diskussionen och analysen av hypotesen om de två faktorerna som genererades av EFA. Därefter kommer diskussionen även omfatta lärarnas perspektiv på generativ AI i gymnasieskolan utifrån det kvalitativa resultatet från intervjuerna. Slutligen kommer rekommendationer för framtida forskning att omdiskuteras.

SVARSFREKVENSER

Inom detta avsnitt kommer svarsfrekvensen till de frågor vars svar hade en statistiskt signifikant förändring från enkäten innan till enkäten efter genomförandet av lektionsserien att tolkas och diskuteras.

Den femte frågan (se Figur 6) syftade till att kvantifiera gymnasieelevernas uppfattning av reliabiliteten av svar genererade av en generativa LLM. I denna mening syftar reliabilitet till huruvida modellens utmatning är sanningsenlig eller ej. Då nästan alla elever från början ansåg att generativa LLM ger dem korrekt information kan en tolkning vara att eleverna var ej medvetna om möjliga fallgropar. Det är tydligt att eleverna var ej kritiska till reliabiliteten av generativa LLM sedan innan lektionsserien och då i stället valde att lita på AI-modellen utan tvekan. Jämfört med svarsfrekvensen från enkäten efter lektionsserien är det tydligt att lektionsserien hade fått en statistiskt signifikant andel av eleverna att vara mer kritiska till reliabiliteten av en generativa LLM.

Den tionde frågan (se Figur 7) syftade till att kvantifiera gymnasieelevernas tekniska kunskap om hur generativ AI av typen LLM skiljer sig från andra typer av AI såsom exempelvis klassificerare eller regressorer. Från början ansåg ungefär hälften av eleverna att påståendet stämde och ungefär hälften av eleverna ansåg att påståendet stämde inte. Efter lektionsserien kan en intressant utveckling observeras då en överväldigande majoritet av eleverna i stället påstod att påståendet stämde in. Denna förändring i svarsfrekvensen är ett tydligt tecken på att lektionsserien hade en verklig och observerbar påverkan på utvecklingen av elevernas tekniska kunskap om hur generativa LLM skiljer sig från andra typer av AI.

Den elfte frågan (se Figur 8) syftade till att kvantifiera elevernas förståelse för hur de som användare interagerar med en LLM. Detta inkluderar hur inmatning kan formuleras för att strukturera modellens utmatning. Innan lektionsserien ansåg en stor minoritet av eleverna att de inte kunde beskriva hur generativa LLM fungerar från användarens perspektiv. Lektionsserien hade en tydlig påverkan på elevernas teoretiska kunskaper om hur användare interagerar med LLM eftersom nästan alla elever ansåg att påståendet stämde efter lektionsseriens genomförande. Utifrån resultatet kan en teoretisera att en stor andel av eleverna via privat användande av LLM (såsom ChatGPT) genom försök och misstag hade lärt sig vad som fungerar sämre och vad som fungerar bättre. Då andelen av elever som ansåg påståendet ej stämma in hade minskat kan en tolkning vara att den återstående minoriteten – som ej hade förkunskaper – såg lektionen som tillräcklig för att möta nivån av eleverna med förkunskaper. Trots att en majoritet av eleverna hade förkunskap om hur LLM kan användas fanns det fortfarande utrymme för utveckling som lektionsserien tillgodosåg.

Den tolfte frågan (se Figur 9) syftade till att kvantifiera elevernas kunskap om hur LLM fungerar på en teknisk nivå. Detta inkluderar hur en LLM är strukturerad, hur den underliggande ML-modellen tränas och hur agenten genererar en utmatning utifrån en inmatning. Innan lektionsserien ansåg mer än hälften av eleverna att detta påstående inte stämde. Lektionsserien lyckades informera ungefär hälften av dessa elever tillräckligt för att de skulle anse att påståendet stämde. Dock ansåg ungefär en fjärdedel av eleverna sig fortfarande vara osäkra. Resultatet från den första enkäten kan återigen tolkas som att en minoritet av eleverna är personligen

intresserade i teknisk utveckling och har lärt sig hur dessa system fungerar på egen hand. Att en minoritet av eleverna fortfarande ansåg påståendet ej stämma in kan vara ett tecken på att LLM (och AI mer generellt) är ett tröskelbegrepp. En observation är att att lektionsserien hade tagit en stor andel av eleverna över tröskeln, men i slutändan var den inte tillräcklig för att ta alla elever över tröskeln.

Den trettonde frågan (se Figur 10) syftade till att kvantifiera hur väl eleverna kan granska datorgenererad text. I denna mening syftar granskning till verifiering av textens validitet, sannolikhet och användbarhet. Det mest intressanta är att eleverna i snitt ansåg att påståendet stämde in *sämre* efter lektionsserien än innan. Väldigt många elever ansåg sig vara helt säkra på att granska datorgenererad text, men efter lektionsserien ansåg dessa elever sig vara säkra, men inte helt säkra. En slutats som kan dras är att eleverna överskattade sin granskningsförmåga. Vidare kan detta tolkas som att en stor minoritet av eleverna kände sig självsäkra i sin granskning av datorgenererad text, men efter lektionsserien insåg de att granskning är en mer komplicerad process än vad de tidigare hade trott.

En statistiskt signifikant påverkan på utvecklingen av elevernas tekniska kunskap observerades vid analys av svarsfrekvensen för samtliga tekniska frågorna (10, 11 och 12). Utifrån detta är det tydligt att lektionsserien hade fungerat väl som en introducerande moment till AI-fältet och därmed mött studiens syfte. Även studiens syfte att främja kritiskt tänkande hos eleverna kan observeras vid analys av svarsfrekvensen för fråga 13.

De frågor som inte berördes inom detta avsnitt (jämför med Tabell 2) visade en statistiskt insignifikant förändring. Detta betyder att skillnaden mellan svarsfrekvensen och därmed förändringen i elevernas åsikter mellan de två enkäterna är minimal. Detta tolkas som att lektionsserien inte hade lyckats påverka elevernas kunskapsnivåer inom dessa områden.

FAKTORER

Inom studien genomfördes en primär och sekundär EFA för vardera enkät. Syftet med analysen var att hitta en serie av faktorer innan och efter genomförandet av lektionsserien med mål att beskriva hur eleverna påverkades av interventionen.

De sekundära EFA för båda enkäterna resulterade i två faktorer vardera där den ena kan anses vara starkt relaterad och den andra svagt relaterad. En observation är att den första enkätens starka faktor i stort överlappar med den andra enkätens svaga faktor (härefter refererad till som den *första* faktorn). På samma sätt kan den första enkätens svaga faktor observeras överlappa med den andra enkätens starka faktor (härefter refererad till som den *andra* faktorn).

Notera att styrkan och svagheten av en faktor är ej relaterad till andelen elever som faktorn representerar. I stället motsvarar styrkan och svagheten hur väl hypotesen kan styrkas med datan. Då båda faktorerna vid olika tillfällen observerades vara både starkt och svagt relaterade kan deras validitet anses vara jämförbar.

FÖRSTA FAKTORN

Den första faktorn representerar de elever som använder sig av LLM med mål att (1) bli klara med sina arbetsuppgifter fortare, (2) lösa problem som de annars inte kunde lösa utan hjälp av LLM och (3) är intresserade av att lära sig mer om LLM.

Dessa elever använder sig av LLM som ett personlig assistent med ett tydligt mål att effektivisera sin arbetsprocess och därmed även sitt lärande (Zhai et al., 2021). En tolkning är att denna faktor i praktiken representerar de elever som sedan tidigare är personligen intresserade av LLM och hur tekniken kan användas för att öka deras produktivitet.

Innan lektionsserien ansåg dessa elever sig vara säkra på att granska, tolka och resonera kring datorgenererad text. Notera dock att frekvensanalysen visade att eleverna i regel överskattade sin förmåga att granska datorgenererad text innan genomförandet av lektionsserien (se Figur 10). Även här kan denna trend observeras då dessa elever inte längre ansåg sig vara säkra på att granska datorgenererad text efter genomförandet av lektionsserien. Detta tolkas dock inte som en regression i elevernas förmågor utan i stället som en utveckling i elevernas kunskap om vad som behövs för att effektivt granska, tolka och resonera kring datorgenererad text.

En intressant observation är att eleverna från början ansåg att skolan bör satsa mer på LLM, men att denna åsikt inte längre var värderat lika mycket efter genomförandet av lektionsserien. Det finns två tolkningar, antingen att lektionsserien var tillräckligt informativ för eleverna och att de inte längre ansåg sig behöva någon vidare undervisning, eller att lektionsserien tydliggjorde för eleverna att de inte hade behövt någon undervisning från första början. Av de två alternativen verkar den första vara mer verklighetsenlig då svarsfrekvensen visar att eleverna i snitt svarade med större säkerhet på samtliga av de tekniska frågorna 10, 11, och 12 (se Tabell 8).

Samtidigt anser dessa elever sig fått ett större förståelse för hur LLM fungerar från användarperspektiv, vilket bland annat omfattar hur användarens inmatning kan struktureras för att förbättra modellens utmatning. Detta kan anses ha en förstärkande effekt på elevernas möjlighet att använda sig av LLM med hänsyn till de tre tidigare nämnda målen. Ett sådant exempel vore det första målet - att de kan bli klara ännu fortare än om de inte hade deltagit i lektionsserien.

Utvecklingen i elevernas förståelse för relationen mellan användarens inmatning och modellens utmatning är en identifierare för den fjärde nivån av Blooms taxonomi (se Figur 1). Då denna nivå av Blooms taxonomi inte hade observerats innan lektionsserien är det även ett tecken att elever som representeras av denna faktor hade förbättrat sin AI-läskunnighet (Ng et al., 2021).

ANDRA FAKTORN

Den andra faktorn representerar de elever som anser att (1) generativa språkmodeller ger dem korrekt information och att de (2) har förmågan att särskilja mellan datorgenererad och människoskriven text, vilket de använder för att (3) lösa deras problem åt dem med mål att (4) förbättra sina skolresultat.

Vidare hade dessa elever börjat anse sig vara säkra på att granska, tolka och resonera kring datorgenererad text efter genomförandet av lektionsserien. Notera dock att svarsfrekvensen visade att eleverna överskattar sin förmåga att granska, tolka och resonera kring datorgenererad text.

Observera dock att datan i grund och botten är självrapporterad och att det är eleverna själva som beskriver sina förmågor. Inom studien finns det inget objektivi mått för hur väl eleverna verkligen kan granska datorgenererad text eller särskilja mellan datorgenererad och människoskriven text. En tolkning är att denna faktor i praktiken representerar de elever som är mindre kritiska och använder sig av verktyg såsom ChatGPT som en genväg för att förbättra sina skolresultat.

En intressant observation som stödjer denna hypotes är att dessa elever i regel anser att påståendet under fråga 9 – Jag tycker att generativa språkmodeller löser mina problem åt mig – stämmer in bättre än påståendet under fråga 8 – Jag tycker att generativa språkmodeller hjälper mig lösa problem jag annars inte kunde lösa. Detta i samband med målet att förbättra sina skolresultat kan återigen tolkas som att dessa elever använder sig av LLM som en genväg. En parallell kan dras till forskningen av Zhai et al. (2021) som beskrev risken att eleverna med hjälp av LLM kan få svar på frågor utan att själva förstå det underliggande resonemanget. Utifrån resultatet kan existensen av denna risk inte bara bekräftas, men även indirekt observeras förekomma hos en andel av eleverna.

En sekundär tolkning är att dessa elever använder sig av LLM för att påskynda genomförandet av repetitiva uppdrag som de annars hade kunnat lösa med hjälp av traditionella metoder. Ett godartat exempel kan teoretiseras vara verifiering av resultat i matematik. Dock kan ett elakartat exempel teoretiseras vara generering av text eller programkod som lösning till diverse inlämningsuppgifter för att uppnå målet att förbättra sina skolresultat.

LÄRARNAS PERSPEKTIV

Lärarna hade beskrivit ett behov av undervisning inom AI som fokuserar inte bara på den underliggande tekniken, utan även på etiska problem och kritiskt tänkande. Detta är något som stöds av undervisningsramverket framtaget av Long och Magerko (2020) och berkäftas genom forskning av Zhou et al. (2020). Viktigt att notera är att lärarna även beskrev situationer där elever hade använt sig av LLM på ett otillåtet sätt för att artificiellt förbättra sina betyg. Samtidigt har lärarna även förklarat att vissa elever helt avstår från LLM då de inte vet vad som

är tillåtet. Av relaterade skäl menar Ng et al. (2022) att etik och kritik är centrala delar av AI-läskunnighet och bör vara en central del av undervisningen. I denna mån hade lektionsserien lyckats förbättra elevernas förmåga att kritiskt granska utmatningen och användningen av LLM.

Samtidigt hade lärarna erkänt att elever kan använda sig av verktyg såsom ChatGPT för att snabbt få ett svar utan att behöva tänka själva. Forskning av Jablonka et al. (2023) visar att ett sådant användningsområde kan uppfyllas av en så kallad personlig assistent. Cain (2023) menar i stället att denna funktion kan fyllas av en generell LLM tränad på en tillräckligt stor datamängd, vilket bäst representerar den observerade situationen där de mest frekvent använda LLM är ChatGPT, Copilot eller Gemini (se Figur 4). Dock menar Jablonka et al. att en denna funktion ska fyllas av en LLM tränad på relevant litteratur och undervisningsmaterial, vilket kan vara ett tekniskt utvecklingsområde de kommande åren. Zhai et al. (2021) har i sin forskning visat att eleverna kan använda sig av personliga assistenter för detta ändamål, men att en lärare bör vara närvarande för att handleda eleverna i deras användande av verktygen.

Lärarna hade beskrivit att eleverna får svar av LLM som ser korrekta ut men i praktiken ofta är inkorrekta eller missvisande. En tolkning är att eleverna saknar den grundläggande kunskapsnivån som behövs för att kunna avgöra om utmatningen verkligen är korrekta eller ej. Ng et al. (2021) har beskrivit detta fenomen som att eleverna saknar förmågan att förstå och tillämpa LLM (nivå 2 och 3 i Blooms taxonomi, se Figur 1). Frekvensanalysen för den tionde, elfte och tolfte frågan visade att lektionsserien lyckades utveckla elevernas förmåga att analysera (nivå 4). På samma sätt visade frekvensanalysen för den trettonde frågan att lektionsserien även lyckades utveckla elevernas förmåga att värdera (nivå 5) och därmed uppfyllt det mål som satts av Ng et al.

FRAMTIDA FORSKNING

Denna studie har tydliggjort att det finns utrymme för vidare forskning om hur generativ AI kan praktiskt implementeras i gymnasieskolor och vilka konsekvenser användande av AI har på elevernas lärande. Framtida forskning inom området bör fokusera på flera viktiga aspekter för att vidare fördjupa förståelsen och amplifiera effekten av pedagogiska interventioner relaterade till generativ AI.

En av de viktigaste faktorerna som bör tas i åtanke vid framtida forskning är längden av de praktiska momenten i relation till de teoretiska momenten. Den lektionsserie som tagits fram som del av studien använde en relativt liten andel av tiden till praktiska moment. Då varken användningsvanorna eller elevernas praktiska färdigheter blivit statistiskt signifikant påverkade av lektionsserien kan en teori vara att denna andel av praktiska moment varit för liten.

En möjlig väg framåt är att inkludera fler praktiska moment i undervisningen. Detta inte bara som en separat lektionsserie utan som en integrerad del av den ordinarie undervisningen. Genom att ge eleverna möjlighet att tillämpa sina teoretiska kunskaper i flera praktiska övningar kan man förvänta sig att förbättra elevernas förmåga att använda sig av generativ AI.

Ett annat viktigt område för framtida forskning är att undersöka elevernas uppfattningar och förståelse av LLM. Eftersom detta koncept visades vara en tröskel för många elever kan en ökad fokus på att förklara och tydliggöra dess relevans och tillämpningar vara av vikt.

En annan intressant riktning för framtida forskning är att undersöka hypotesen om två distinkta elevgrupper med olika syften och mål för att deras användning av generativ AI. Det skulle vara värdefullt att fastställa hur stor andelen av varje grupp är och vilka faktorer som kan påverka elevernas tillhörighet till en specifik grupp.

SLUTSATS

Sammanfattningsvis visade denna studie att en strukturerad pedagogisk intervention i form av en lektionsserie om generativ artificiell intelligens hade en effekt på gymnasieelevers kunskaper och färdigheter. Trots att elevernas användningsvanor förblev relativt oförändrade efter lektionsserien har viktiga insikter identifierats.

Det fanns inga signifikanta bevis för att elevernas praktiska färdigheter har förbättrats, vilket indikerar behovet av en ökad betoning på praktiska moment i undervisningen. Däremot kunde en tydlig framgång ses när det gällde utvecklingen av elevernas tekniska färdigheter, vilket visar på potentialen för undervisning om artificiell intelligens på teknikprogrammet i den svenska gymnasieskolan.

Vidare visade studien att elever kan delas in i två huvudsakliga grupper baserat på deras syften med användningen av generativ artificiell intelligens. En grupp använder tekniken för att utmana sig själva och lösa problem som de tidigare inte kunde hantera, medan en annan grupp använder den för att förstärka sina befintliga kunskaper och förbättra sina skolresultat. Denna observation understryker att sättet lärare undervisar på i dagsläget behöver anpassas.

Slutligen visar resultatet att elevernas användningsvanor och praktiska färdigheter i stort blivit oförändrade och därför teoretiserades att en ökad fokus på praktiska laborationer skulle kunna förbättra effekterna av en lektionsserie. Framtida forskning bör därför utforska inklusionen av fler praktiska moment i undervisningen för att vidare främja elevernas lärande genom användning av generativ artificiell intelligens.

REFERENSER

- Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, *15*(2). <https://doi.org/10.7759/cureus.35179>
- Alkharusi, H. (2022). A descriptive Analysis and Interpretation of Data from Likert Scales in Educational and Psychological Research. *Indian Journal of Psychology and Education*, *12*(2), 13–16.
- Batterton, K. A., & Hale, K. N. (2017). The Likert Scale What It Is and How To Use It. *Phalanx*, *50*(2), 32–39.
- Bjørkeng, P. K. (2024). *Oslo-skolen kaster seg på ChatGPT: – Vi hadde ikke annet valg*. Hämtad 19 februari 2024, från <https://www.aftenposten.no/kultur/i/gEoXG1/oslo-skolen-kaster-seg-paa-chatgpt-vi-hadde-ikke-annet-valg>
- Boone, H. N., & Boone, D. A. (2012). Analyzing Likert Data. *The Journal of Extensions*, *50*(2). <https://doi.org/10.34068/joe.50.02.48>
- Cain, W. (2023). Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends*, *68*(1), 47–57. <https://doi.org/10.1007/s11528-023-00896-0>
- Clason, D. L., & Dormody, T. J. (1994). Analyzing Data Measured By Individual Likert-Type Items. *The American Association for Agricultural Education*, *35*(4), 31–35. <https://doi.org/10.5032/jae.1994.04031>
- Filipsdotter. (2023). *Blooms reviderade taxonomi (2001)* [Figur]. Wikimedia Commons. Hämtad 14 mars 2024, från [https://commons.wikimedia.org/wiki/File:Blooms_reviderade_taxonomi_\(2001\).jpg](https://commons.wikimedia.org/wiki/File:Blooms_reviderade_taxonomi_(2001).jpg)
- Gideon, L. (2012). The Art of Question Phrasing. I L. Gideon (Red.), *Handbook of Survey Methodology for the Social Sciences* (s. 91–107). Springer. https://doi.org/10.1007/978-1-4614-3876-2_7
- Gideon, L., & Moskos, P. (2012). Interviewing. I L. Gideon (Red.), *Handbook of Survey Methodology for the Social Sciences* (s. 109–118). Springer. https://doi.org/10.1007/978-1-4614-3876-2_8
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510–3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Jablonka, K. M., Ai, Q., Al-Feghali, A., Badhwar, S., Bocarsly, J. D., Bran, A. M., Bringuier, S., Brinson, L. C., Choudhary, K., Circi, D., Cox, S., de Jong, W. A., Evans, M. L., Gastellu, N., Genzling, J., Gil, M. V., Gupta, A. K., Hong, Z., Imran, A., . . . Blaiszik, B. (2023). 14 examples of how LLMs can

- transform materials science and chemistry: a reflection on a large language model hackathon. *Digital Discovery*, 2(5), 1233–1250. <https://doi.org/10.1039/d3dd00113j>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning*. Springer. https://doi.org/978-3-031-38747-0_1
- Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing*, 72(12), 2954–2965. <https://doi.org/10.1111/jan.13031>
- Kurzweil, R. (1990). *The Age of Intelligent Machines*. The MIT Press.
- 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*. <https://doi.org/10.1145/3313831.3376727>
- Luckin, R., & Holmes, W. *Intelligence Unleashed – An argument for AI in Education*. London, 2016.
- MacFarland, T. W., & Yates, J. M. (2016). Mann–Whitney U-test. I *Introduction to Nonparametric Statistics for the Biological Sciences using R* (s. 103–132). Springer. https://doi.org/10.1007/978-3-319-30634-6_4
- Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2022). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, 28(1), 8445–8501. <https://doi.org/10.1007/s10639-022-11491-w>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & shen Qiao, M. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2(1). <https://doi.org/10.1016/j.caeai.2021.100041>
- Oldendick, R. W. (2012). Survey Research Ethics. I L. Gideon (Red.), *Handbook of Survey Methodology for the Social Sciences* (s. 23–35). Springer. https://doi.org/10.1007/978-1-4614-3876-2_3
- Persson, A. (2016a). Frågeteknik. I A. Persson (Red.), *Frågor och svar – om frågekonstruktion i enkät- och intervjuundersökningar* (s. 33–80). SCB-Tryck.
- Persson, A. (2016b). Svartalernativ. I A. Persson (Red.), *Frågor och svar – om frågekonstruktion i enkät- och intervjuundersökningar* (s. 81–133). SCB-Tryck.
- Polverini, G., & Gregorcic, B. (2024). How understanding large language models can inform the use of ChatGPT in physics education. *European Journal of Physics*, 45(2). <https://doi.org/10.1088/1361-6404/ad1420>

- Skolverket. *Digitaliseringen i skolan – möjligheter och utmaningar*. Stockholm, 2018.
- Skolverket. (2024). *Artificiell intelligens - nytt ämne i gymnasieskolan och komvux*. Hämtad 2 juni 2024, från <https://www.skolverket.se/regler-och-ansvar/forandringar-inom-skolområdet/artificiell-intelligens---nytt-amne-i-gymnasieskolan-och-komvux>
- Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet. *Proceedings of the 2021 International Conference on Interdisciplinary Applications of Artificial Intelligence*. <https://doi.org/10.48550/arXiv.2102.08036>
- Worden, K., Tsialiamanis, G., Cross, E. J., & Rogers, T. J. (2023). Artificial Neural Networks. I T. Rabczuk & K.-J. Bathe (Red.), *Machine Learning in Modeling and Simulation* (s. 85–120). Springer. https://doi.org/10.1007/978-3-031-36644-4_2
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., LIU, J.-B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020 (N. Cai, Red.). *Complexity*, 2021(1). <https://doi.org/10.1155/2021/8812542>
- Zhou, X., van Brummelen, J., & Lin, P. (2020). Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework. <https://doi.org/10.48550/arXiv.2009.10228>

