

Teknologi og digitalisering

I denne spalten vil Lars Erlend Leganger og noen av hans kollegaer skrive om aktuelle temaer innen teknologi/digitalisering som direkte eller indirekte også vil påvirke revisors hverdag. Lars Erlend er AI-ekspert og direktør i PwC. Han har en PhD i teoretisk fysikk fra NTNU.

Kunstig intelligens i klasserommet

Denne artikkelen ser på hvordan ChatGPT og annen GenAI blir brukt i lærings- og eksamenssituasjoner, og hva dette kan bety for hvordan vi lærer – og eksaminerer – i fremtiden. Neste gang det gjennomføres fagfornyelse og kunnskapsløft i den norske skolen kan det fort være AI-drevne digitale læringsprodukter, og ikke fysiske lærebøker og statisk digitalt innhold, som er de sentrale læremidlene.



PhD
Lars Erlend Leganger
Direktør i PwC

Våren 2023 tok den norske skolen i bruk de siste nye læreplanene i en lenge planlagt fagfornyelse. Omtrent samtidig ble norske elever, lærere, og studenter truffet av en mer kaotisk og uventet fornyelse, da generativ kunstig intelligens (GenAI), med tekstgeneratoren ChatGPT i spissen, ble allemannseie. Oppfinnsomme studenter og lærere oppdaget med skrekkblandet fryd at AI-løsninger kunne være kraftige verktøy ikke bare i læringsprosessen, men også på eksamen. Går tiden med hjemmeeksamen og “alle hjelpemidler tillatt” mot slutten? Er egentlig bruken av GenAI på eksamen et problem?

ChatGPT er god på det den har sett før – og den har sett det meste

Generativ kunstig intelligens er maskinlæringsmodeller som kan produsere nytt innhold – som tekst, bilder, kode, lyd – basert på mønstre de har (maskin)lært fra store datasett. Den nyeste kommersielt tilgjengelige versjonen av ChatGPT bygger på fjerde versjon av språkmodellen GPT, trent opp på enorme mengder innsamlet tekstdata. Nøyaktig hva data-

grunnlaget for GPT versjon 4 er, holdes hemmelig av eierne, de en gang så idealistiske og åpne OpenAI, men tidligere versjoner av GPT-modellen var trent opp på enorme mengder automatisk innsamlede nettsider, som Reddit og Wikipedia, samt en stor mengde bøker.¹

Det er fremdeles langt igjen fra ChatGPT til generell menneskelignende intelligens, men på enkelte oppgaver presterer ChatGPT helt på høyden med mennesker. Dette inkluderer blant annet idémyldring, disposisjoner og utkast, sammendrag, omskrivning, og til en viss grad anvendelse av kjente metoder på nye problemstillinger. Med treningsdata som omfatter mye av det som kan oppdrives av verdens bøker og nettsider, inkludert hele Wikipedia, gjør dette ChatGPT godt skodd til å utføre mange store og små tekstbaserte aktiviteter knyttet til læring, undervisning, og vurdering av læringsutbytte.

AI på eksamen

I Norge har mediene rapportert om eksamensoppgaver besvart med hjelp av AI,² og eksamensoppgaver rettet ved hjelp av AI.³ Bruken varierer fra kvali-

tetsheving, med AI som et av flere støtteverktøy i eksamensarbeidet (f.eks. til research og språkvask), til studenter som gjør kort prosess og mater ChatGPT direkte med eksamensspørsmål og bruker svarene blindt uten videre egeninnsats.⁴ I systematiske tester av GPT-4 presterer modellen i 90-persentilen på standardiserte eksamener i et bredt spekter av kategorier, fra jus, til kunsthistorie, til sommelier/vinkunnskap.⁵

Certified Public Accountant AI

På eksamener i revisjon og regnskap så det først ut som ChatGPT kom til kort. I januar gjorde ChatGPT-3.⁵ det merkbart dårligere enn mennesker på et bredt utvalg av oppgaver,⁶ og i begynnelsen av mai kunne Accounting Today rapportere at ChatGPT-3.5 var sjanseløs på Certified Public Accountant (CPA)-eksamen.⁷ Men ut-av-boksen bruk av ChatGPT er én ting – mye av disruptjonspotensialet i denne typen modeller er at avanserte brukere kan, med relativt liten innsats, gjøre model-

1 Se <https://arxiv.org/pdf/2005.14165.pdf>, tabell 2.2.

2 <https://www.vg.no/nyheter/i/Kn1O7M/student-skrev-eksamen-med-ai-fikk-a-paa-hovedoppgave>

3 <https://www.vg.no/nyheter/i/P4X2qJ/brukte-ai-til-aasensurere-50-bacheloroppgaver>

4 <https://www.studvest.no/spurte-chatgpt-fikk-c-pa-eksamen/>

5 Se <https://cdn.openai.com/papers/gpt-4.pdf>, tabell 1.

6 https://www.researchgate.net/publication/370211135_The_ChatGPT_Artificial_Intelligence_Chatbot_How_Well_Does_It_Answer_Accounting_Assessment_Questions

7 <https://www.accountingtoday.com/news/we-ran-the-cpa-exam-through-chatgpt-and-it-failed-miserably>

lene signifikant bedre på spesifikke problemstillinger. I mai rapporterte Eulerich et. al. at en forsterket utgave av ChatGPT-4, som hadde fått se noen øvingsoppgaver og var utvidet med tilgang på kalkulator (språkmodeller er notorisk dårlige i matematikk), hadde en treffrate på multiple choice-oppgaver som tilsa at den ville bestått ikke bare en



På lengre sikt er det sannsynlig at effektiv og trygg bruk av GenAI-verktøy går fra å være potensielt uønsket, i alle fall for hjemmeleksker og eksamener, til å bli en sentral del av undervisningen i mange fag.

CPA-eksamen, men også Certified Management Accountant (CMA), Certified Internal Auditor (CIA), og den amerikanske skatteetatens enrolled agent (EA)-eksamen.⁸

Hva gjorde matematikklærerne da kalkulatoren kom?

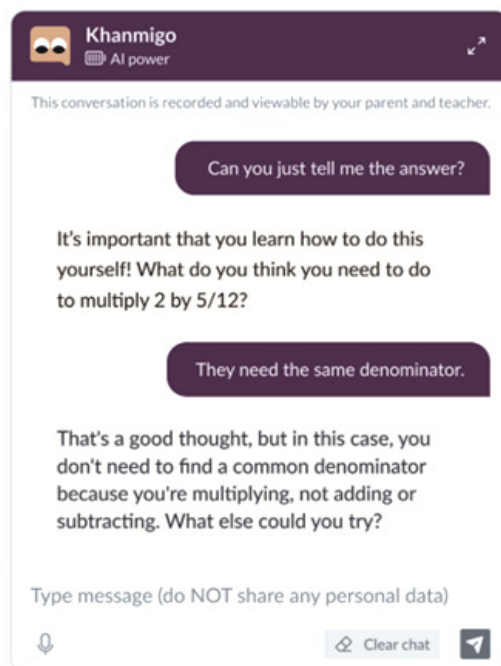
At en AI kan bestå eksamen, betyr ikke nødvendigvis at den faktisk mestrer faget som blir eksaminert. Det er ikke alt som er like enkelt å eksaminere på en skalerbar måte, og menneskelige CPA-kandidater vil (forhåpentligvis) ha fått med seg flere ferdigheter enn å løse multiple choice-oppgaver. Samtidig kan det være svært ressurskrevende og lite skalerbart å gå i dybden på hver enkelt kandidats forståelse. Grep som gjør eksamineringen mer standardisert og effektiv, har lett for også å gjøre den mer overfladisk og håndterbar for AI. Dette er ingen ny problemstilling – det er en evig debatt i skolen hvorvidt fokus på resultatene av standardiserte tester forleder lærere til å ned-

prioritere vanskelig målbar dybdeinnsikt til fordel for enklere målbar overflatekunnskap (“teaching for the test”).

Som eksaminator er det forskjellige måter å møte denne utviklingen på. På den ene siden kan en gjøre grep som gjør det vanskeligere for eksamenskandidater å bruke bredt tilgjengelige GenAI-verktøy. Et enkelt og brutalt effektivt tiltak er å gå helt bort fra hjemmeeksamener og alle-hjelpemidler-tillatt-eksamener, og i stedet gjøre all vurdering av læringsutbytte muntlig eller skriftlig i eksamenslokaler, med vaktthold som sikrer at GenAI og andre uønskede hjelpemidler ikke brukes. Det er slik det gjøres i dag for delvis å vurdere matematikk-elevens regneferdigheter uten kalkulator, eller norsk-elevens sidemålsferdigheter uten tilgang på automatiske oversettelsesverktøy.

Samtidig er hjemmeeksamener kanskje den eksamensformen som best etterligner hvordan kunnskap brukes i den virkelige verden – det er sjelden en ingeniør avkreves muntlig svar på stående fot på om brua vil holde, eller at en viktig juridisk betenkning skal utarbeides på et par timer, uten tilgang på eksterne hjelpemidler. Som alternativ til å avskaffe hjemmeeksamener helt foreslår NTNU at bekymrede faglærere f.eks. benytter bilder, lyd, og figurer i oppgavene og/eller besvarelsene – formater som standard ChatGPT takler dårlig.⁹ På samme vis har Universitetet i Agder identifisert 15 områder der ChatGPT (slik den forelå i februar 2023) kommer til kort, som spørsmål om samtiden, korrekte sitater og kildehenvisninger, og å skille mellom relevant og irrelevant informasjon.¹⁰ Håpet er at kløktig utformede eksamensoppgaver skal gjøre det så tungvint å benytte GenAI-verktøy at det er enklere for kandidatene å gjøre som før og lære seg tingene selv. I praksis risikerer en å havne i en situasjon der en tester at kandidaten enten 1) har opp-

nådd læringsmålene og svarer på oppgaven med egen kunnskap, eller 2) har tilgang på og mestrer å bruke mer avanserte GenAI-verktøy som håndterer oppgavens snubletråder, det være seg egenutviklede, som ovennevnte Eulerich et. al. på CPA-eksamen, eller skreddersydde betalte løsninger, som Harvey AI for jus.¹¹ Så er spørsmålet om det er så farlig om det er 1) eller 2) (eller begge), så lenge kandidaten får jobben gjort?



Khanmigo gir en-til-en undervisning i matematikk (utklipp fra <https://openai.com/customer-stories/khan-academy>)

Etablerte aktører må integrere AI i produktene og tjenestene sine

Generativ AI påvirker læring og undervisning også når verktøyene ikke er tilgjengelige på eksamen. Studenter og lærere kan mate læreplaner og temaer rett inn i ChatGPT og få et bredt spekter av AI-genererte læringsressurser tilbake, fra leseplaner, kontrollspørsmål for å teste forståelse, repetisjonsoppgaver, og oppsummeringer, til kreative “energizers” og forslag til hvordan en kan tenke utenfor boksen for å gjøre en undervisnings-

8 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452175

9 <https://i.ntnu.no/wiki/-/wiki/Norsk/Eksamen+og+kunstig+intelligens+-+for+fag%C3%A6rar>

10 <https://www.uia.no/nyheter/15-ting-chatgpt-ikke-klarar>

11 Se Revisjon og Regnskap Nr. 4 2023: Harvey AI bringer teknologien bak ChatGPT til advokatbransjen https://www.revregn.no/journal/2023/4/m-569/Harvey_AI_bringer_%C2%ADteknologien_bak_ChatGPT_til_advokatbransjen



Curipod-generert undervisningsmateriale (generert fra forfatteren som bruker).

økt mer engasjerende for elevene. Utfordrere i læringsmiddelbransjen som er fremme i skoa på teknologi, har allerede lansert de første digitale produktene med generativ AI integrert: Khan Academys Khanmigo gir AI-drevet entil-en undervisning, med oppmuntrende råd og dialogbasert hjelp steg for steg på veien til riktig svar¹² (figur 1). Norske Curipod kan generere forslag til fulle undervisningsopplegg ut fra stikkord,

med elevaktiviteter ferdig integrert¹⁵ (figur 2).

På lengre sikt er det sannsynlig at effektiv og trygg bruk av GenAI-verktøy går fra å være potensielt uønsket, i alle fall for hjemmelekser og -eksamener, til å bli en sentral del av undervisningen i mange fag, på samme måte som riktig bruk av kalkulator har blitt en integrert del av matematikkundervisningen. En sentral

12 <https://www.khanacademy.org/khan-labs>

15 <https://curipod.com/ai>

nøkkelferdighet her er evnen til å kvalitetssikre leveransene fra GenAI-løsnin-gene, uten at dette tar like lang tid som å gjøre alt selv.¹⁴ Hvordan undervisning i bruk av AI løses i praksis, gjenstår å se. Teknologen Ben Thompson ser for seg en fremtid der elever bruker spesialutviklede skole-AI-verktøy overvåket og kontrollert av læreren, der læreren kan løpende vurdere og gi råd til bruk og mestring av verktøyet, og til og med få AI-en til å levere bevisst gale svar, for å teste om elevene fanger det opp i sine kvalitetssikringsprosesser.¹⁵ Det er all grunn til å tro at evnen til å benytte AI på en effektiv og trygg måte vil være ettertraktet i fremtidens arbeidsmar-ked.

Neste gang det gjennomføres fagforny-else og kunnskapsløft i den norske skolen kan det fort være AI-drevne digitale læringsprodukter, og ikke fysiske lærebøker og statisk digitalt innhold, som er de sentrale læremid-lene. Etablerte aktører som ikke er rigget for å integrere generativ AI i produktene og tjenestene sine, er i ferd med å få det travelt – i læremiddelbran-sjen og i andre bransjer.

Løsning til kryssord på side 15

PROS- ESSER FOR KON- TROLL		AV- LEGGES I RETTEN	DET NYE KON- SEPTET	→ S	EN- ZYM	ITALI- ENSK ELV RØNNE GUD	→ P	ELV I ALASKA	U	STØPE, FORNE	FØLGE- LIG	H	ASTER- OIDE	O	B	I
REGN- SKAPS- STAT- UTGEN	D	INDISK DELSTAT VERBAL- PARTNER	G	O	A	EHIRAT	A	J	M	A	N	SINT	A	R	G	
B	JAPANSK BY	A	N	A	N	JANK- SON AKS	B	O	I	K	O	T	SELSKAPS- FORN- UT- PRØVE	D	A	
A	SINNET, VESEN- ET	N	A	T	U	R	E	N	ØYE- KAST- ET	T	I	T	T	E	N	
L	HIMMEL- RET- NING	ØKE, VOKSE	FLY- SEL- SKAP	S	A	S	DEKA- LEGENS- DEL	T	I	FLATE- MÅL	MÅLE- LENNE MANG STENNER	E	R	G		
A	→ N	S	E	N	ELTIE FOR- TURLET	K	N	→ A	FISK- MÅNE	L	A	K	S	MORT FOR FUNN- SONS- HEMME	↓	
SAR- BEID- NTALE ROMER- TALL 50	E	Ø	S	DRILK- SEKSTEN- FLY	Ø	L	SITTE TIL HEST	R	I	SKADE- DYR	R	O	T	T	E	
L	I	R	E	F	R	E	M	M	O	S	←	R	E	T	T	

14 Et eksempel på hvor galt det kan gå når en ikke har kritisk tilnærming til ChatGPT-resultater: <https://www.forbes.com/sites/mattnovak/2023/05/27/lawyer-uses-chatgpt-in-federal-court-and-it-goes-horribly-wrong/>

15 <https://stratichery.com/2022/ai-homework/>