

Seminarserie Kunstig intelligens og meg = sant?

Språkmodeller

- Hva er LLM og hvordan virker de?
- Hvordan kan vi bruke LLM til å lære?
- Hvordan kan vi bruke LLM til å forske?
- Hvordan bruke LLM korrekt og etisk?



Gerald Torgersen

gerald.torgersen@odont.uio.no

Avdeling for kjeve og ansiktsradiologi



Ordliste

- LLM - large language model (eks. ChatGPT GPT-4)
- Kontekstvindu – hvor mye tekst modellen kan forholde seg til på en gang dvs. samtalekonteksten
- Tokens – hele eller deler av ord som LLM jobber med
- Prompting: nynorsk for å spørre en LLM om noe
 - Zero shot: Spør modellen direkte
 - One shot: Gi et eksempel på hva du er ute etter sammen med spørsmålet
 - Few shot: Gi flere eksempler ...
- System prompt eller instruksjon – en tekst du legger inn i LLM som beskriver kontekst og måten du ønsker at LLM skal kommunisere med deg.

Store språkmodeller

Large language models (LLM)

- Hvordan fungerer de?
- Språk kodet som tall (tokens)

Tannleger er lykkelige mennesker

[51, 1036, 273, 1414, 2781, 14869, 74, 18126, 7404, 96218, 7197] 11 tokens

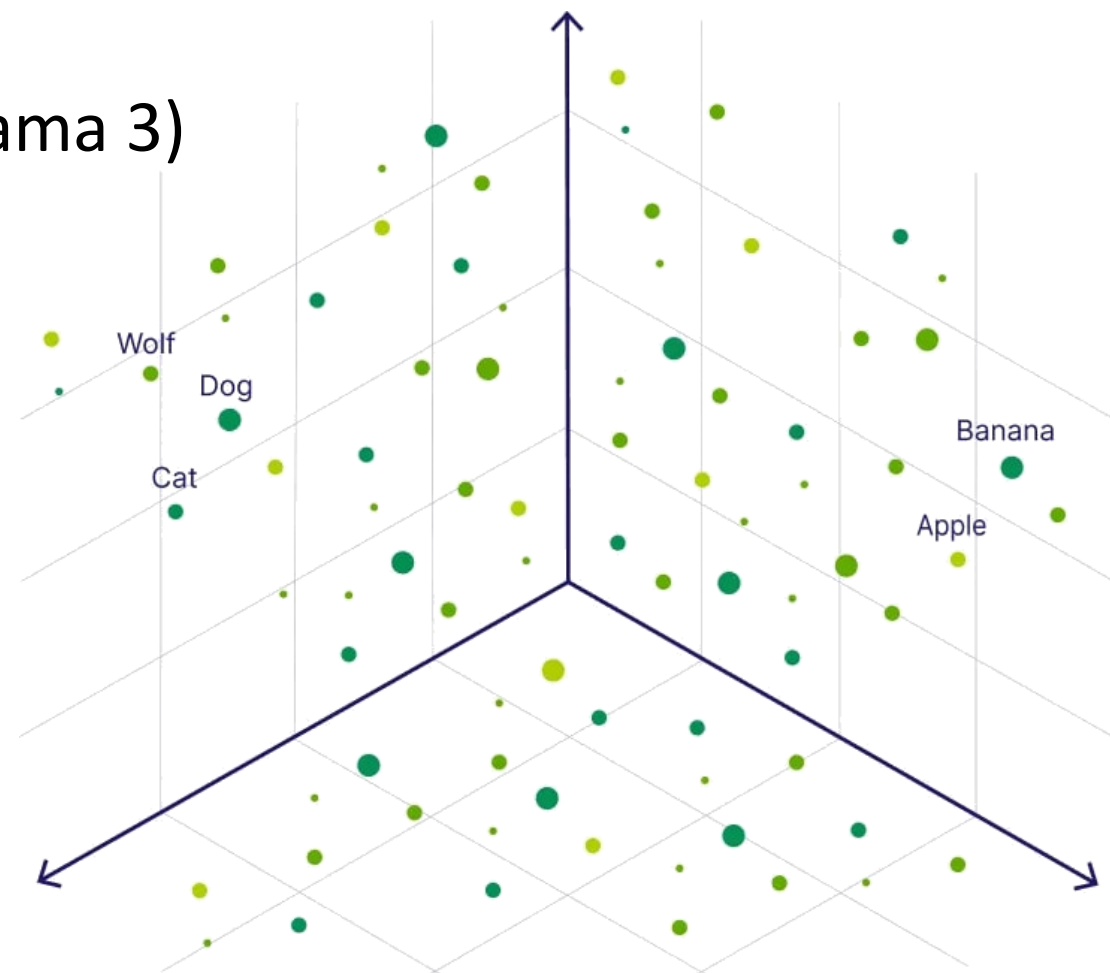
Dentists are happy people

[35, 306, 1705, 527, 6380, 1274] 6 tokens

- Veldig dype nevrone nett som gjetter neste ord
- Ikke egentlig kunnskap i systemet
- Justert med menneskelige tilbakemeldinger
(Reinforcement learning with human feedback (RLHF))

Semantikk kan uttrykkes matematisk

- Vektorer i et 4096 dimensjonalt rom (Llama 3)
- Vektorene er lært i treningen
- Vanlig vektorregning brukes for å finne relasjon mellom begreper
- Trenger vi å vite det for å bruke LLM?



Illustrasjon: <https://dev.to/codemaker2015/goodbye-databases-its-time-to-embrace-vector-databases-190l>

Bærekraft



- Et svar fra ChatGPT krever ca 2,9 Wt $\rightarrow$ 345 svar per kWt¹
 - Av dette: 40 % til utvikling og 60 % til bruk $\rightarrow$ 575 netto-svar per kWt
- Et Google søke krever ca. 0,3 Wt¹
- Løsninger
 - Spesielle language processing units (LPU) (ex. Groq²)
 - Mindre språkmodeller spesialisert
 - Optimalisere modeller og algoritmer
- Personvern, datasikkerhet
 - Du kan ikke være sikker på hvor dine data havner eller hva de brukes til

¹[Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption](#), EPRI report, 2024

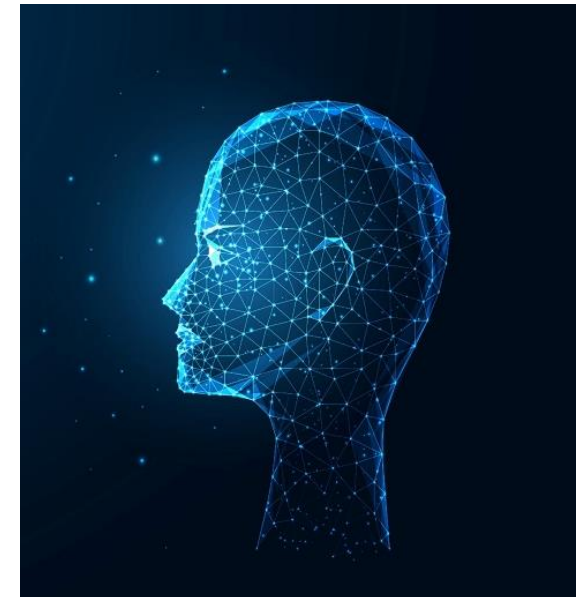
²[The Groq[®] LPU[™] Inference Engine 10x More Energy Efficient Than GPUs. Here's Why.](#), Groq, 2024

Bruk av LLM

- Kreativiteten ligger i hva man ber den om å gjøre
 - Gode instruksjoner (prompts) tilstrekkelig detaljert
 - Gi konteksten for samtalen
 - Finjustere svarene ved å gi tilleggsopplysninger
 - Gjør den oppmerksom på eventuelle feil
 - Jobb videre med tekstforslaget
- Oppgi at du har brukt en LLM
 - hva har du brukt den til
 - ev. hvor mye av teksten er laget av LLM
- GPT-UiO: Gratis og kompatibel med GDPR
 - <https://gpt.uio.no/>

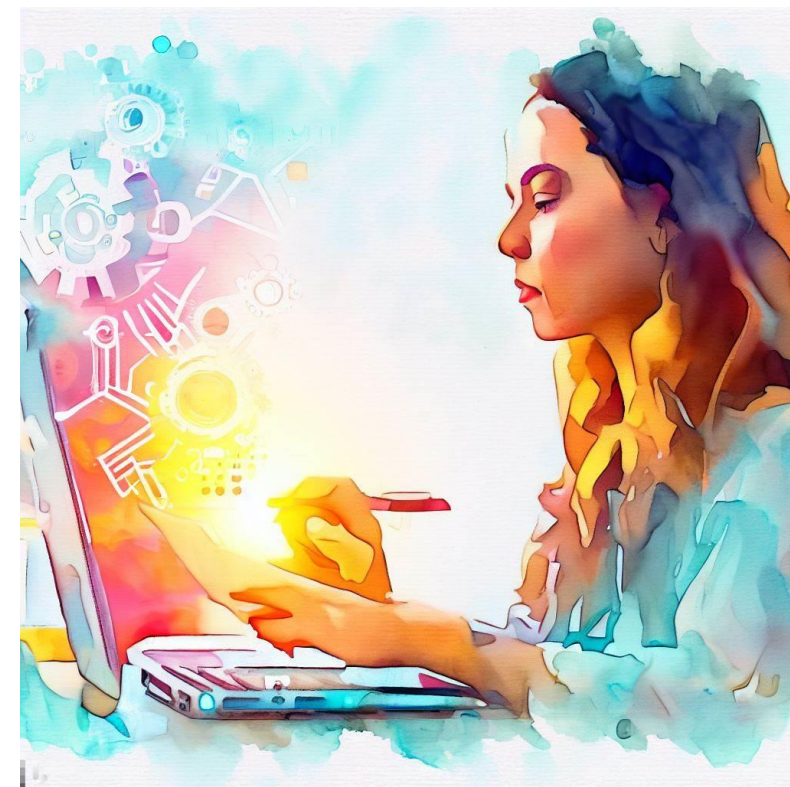


Bing IC: AI is your good friend and assistant helping you in your daily work.



Noen bruksområder til jobb og studier

- LLM er fantastiske verktøy til bl. a.
 - Personlig assistanse i arbeidsoppgaver
 - Hvordan gjør jeg ...
 - Forklar meg ...
 - Gi meg noen ideer til ...
 - Hvorfor funker ikke denne programkoden ...
 - Les dette arbeidet og gi meg forslag til forbedringer
 - Individuelt tilpasset undervisning og læring
 - Kollokviepartner som aldri blir lei
 - Skrivehjelp
 - Interaktive eksamensoppgaver
 - Oversettelse av pensumtekst
 - Lage illustrasjoner som fritt kan brukes



Bing IC: Make a watercolor painting about the bright side of using AI in work

Hva LLM er gode på

- Tekst/språk (de har lest alt som er å lese i verden, og husker det)
- Programmering
 - Python, R, Stata, Excel
- Forklare begreper

- Forklar hva effektiv stråledose er sånn at en 3-åring forstår det:

Effektiv stråledose er som å tenke på hvor mye solsikke du får fra sola. Hvis du spiller ute i solen for lenge, kan du bli solbrent, og det er ikke bra. Effektiv stråledose sier oss hvor mye av den "solsikkesaften" kroppen vår får, men fra spesielle ting som røntgenbilder hos legen. Vi bruker det for å passe på at vi ikke får for mye, slik at vi ikke blir "solbrente" inni kroppen vår. Det hjelper legene å holde oss trygge.

GPT-4o i GPT-UiO, august 2024

OK, ville kanskje snakket om andre ting med en 3-åring 🤪🤪🤪



kan du forklare det samme så en 10-åring forstår det?



Selvfølgelig! Effektiv stråledose er en måte å måle hvor mye stråling kroppen din får, og hvor farlig det kan være for deg. Stråling kan komme fra forskjellige kilder, som røntgenbilder hos legen eller fra radioaktive materialer.



Når du tar et røntgenbilde, bruker legene stråling for å se innsiden av kroppen din. De bruker begrepet "effektiv stråledose" for å finne ut hvor mye energi kroppen din får fra strålingen, og hvordan dette kan påvirke forskjellige deler av kroppen din. Dette hjelper dem å passe på at du får minst mulig stråling, men nok til å få et godt bilde for å hjelpe deg med å bli frisk. Tenk på det som en måte å måle og balansere strålingen på, slik at den er trygg for deg.

Konkrete prosjekter

- Can generative AI assistants enhanced learning?

- I samarbeid med Monica Jeanette Wicksell Torp og Anna Mavroudi (IP, UV)

- SimSambot: En virtuell pasient

- Samarbeid mellom oss og UV og MED
- Hvem er vi?
 - Katrine Gahre Fjeld, Lene Hystad Hove, Ann Catrin Høyvik, Blagica Milanovic, Sada Maroufi og meg
 - Spesielt flott at vi har fått med oss engasjerte masterstudenter

Tildeling av IT-midler til utdanningsprosjekter ved OD

To utdanningsprosjekter har fått økonomisk støtte fra UiO på til sammen 1,5 millioner kroner. Kunstig intelligens er vesentlig i prosjektene, og skal bidra til læring av pensum og trening i pasientkommunikasjon.



Gerald Torgersen på Avdeling for kjeve- og ansiktsradiologi bidrar i begge prosjektene. Det første prosjektet «En læringsassistent basert på kunstig intelligens» har fått innvilget 300 000 kr til å utvikle og teste muligheten for å bruke spesielt tilpassede språkmodeller for å hjelpe studentene med å lære og forstå pensumet vi underviser i radiologi. Gerald Torgersen og Monica Wicksell Torp på OD samarbeider med Anna Mavroudi på UV. De hadde en pilot i fjor høst og jobber videre med dette i høstsemesteret.

«SimSambot» er et samarbeidsprosjekt med medisin (gynekologi) og UV (spes. ped.) der målet er å utvikle en språkmodell som simulerer kommunikasjon mellom en pasient og en kliniker. Gerald Torgersen og Katrine Gahre Fjeld (gerodontologi) samarbeider om å lage en robot som representerer en eldre pasient. Disse botene skal gi studentene trening i å kommunisere. Prosjektet har fått 1 200 000 kr.

<https://www.odont.uio.no/om/aktuelt/aktuelle-saker/2024/tildeling-av-it-midler-til-utdanningsprosjekter-ve.html>

Hva er en prompt engineer?

- "Prompt engineering" refererer til prosessen med å **utvikle effektive spørsmål** og svar som brukes til å trene modeller for naturlig språkprosessering, for å **hjelpe chatboter og andre NLP-baserte systemer til å forstå og svare på brukerforespørsler og -spørsmål på en effektiv og naturlig måte.** (GPT-4)
- Måten man kommuniserer med AI påvirker resultater
 - Søppel inn 💩 – søppel ut 💩 💩 prinsippet
- Lære mer: [Prompt Engineering Guide](#)



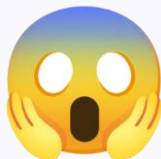
Bing: Make a picture of a prompt engineer

Er GPT-4 en universalgeni?

G Jeg har en 12-liters kanne og en 6-liters kanne. Jeg vil måle 6 liter. Hvordan kan jeg gjøre det med disse? 

G Du kan måle nøyaktig 6 liter ved å bruke følgende trinn:  

1. Fyll 12-literskannen helt opp (nå har du 12 liter i 12-literskannen).
2. Bruk 12-literskannen til å fylle opp 6-literskannen (nå har du 6 liter i 6-literskannen og 6 liter igjen i 12-literskannen).
3. Tøm 6-literskannen helt ut (nå har du en tom 6-literskanne).
4. Hell de 6 literne som er igjen i 12-literskannen over i 6-literskannen (nå har du 6 liter i 6-literskannen og en tom 12-literskanne).

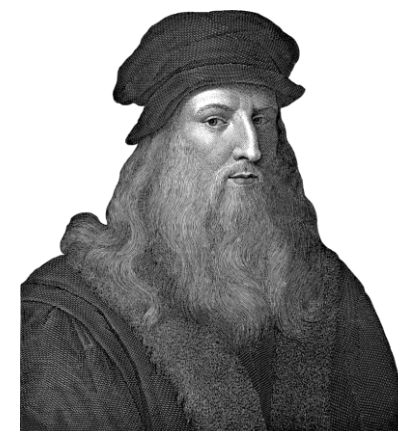
Nå har du nøyaktig 6 liter i 6-literskannen. 

G kan det gjøres enklere? 

G Ja, du kan gjøre det på en enklere måte. Følg disse trinnene:  

1. Fyll 6-literskannen helt opp (nå har du 6 liter i 6-literskannen).
2. Hell innholdet fra 6-literskannen over i 12-literskannen (nå er 12-literskannen halvfull med 6 liter, og 6-literskannen er tom).

Nå har du nøyaktig 6 liter i 12-literskannen, og prosessen er avsluttet. 



Leonardo da Vinci

Hallusinasjoner

- Når LLM bare finner på svaret
- Mindre hallusinasjoner nå
- Men det kan være unøyaktigheter i svarene
- Ditt ansvar å sjekke for feil



Hvor mange r er det i ordet strawberry?



I ordet "strawberry" er det to r-er.



kan du stave ordet og telle en gang til?



Ja, selvfølgelig! Ordet "strawberry" staves slik:

s-t-r-a-w-b-e-r-r-y

La oss telle antall r-er:

1. r (etter t)

2. r (etter e)

Forutinntatt? skjevheter - bias

- Eksempel fra Google Translate^{*}
 - Min kjæreste er ingeniør → My boyfriend is ...
 - Min kjæreste er sykepleier → My girlfriend is ...^{*}
- Kjønn, etnisitet, religion, alder, ...
- Hvor ligger egentlig problemet?



 Kan du oversette til engelsk: "Min kjæreste er ingeniør"

 "My boyfriend/girlfriend is an engineer"

GPT-4o august 2024

Certainly! Here is a portrait of a Founding Father of America:



Sure, here is an image of a pope:



Sure, here is an image of a Viking:

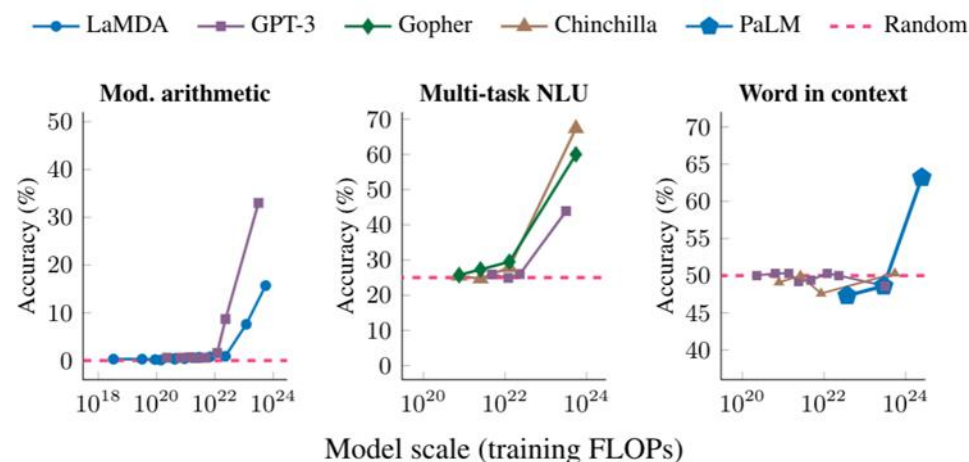


Google Gemini 2024

Emergens

Emergens brukes i vitenskapsteori om fenomener der ulike elementer, ved å påvirke hverandre, skaper nye egenskaper som en ikke kunne forutsett utfra kjennskap til elementene isolert. [Store norske leksikon](#)

“when quantitative changes in a system result in qualitative changes in behavior” Jacob Steinhardt, Professor at UC Berkeley



The ability to perform multi-step arithmetic (left), succeed on college-level exams (middle), and identify the intended meaning of a word in context (right) all emerge only for models of sufficiently large scale. The models shown include [LaMDA](#), [GPT-3](#), [Gopher](#), [Chinchilla](#), and [PaLM](#). Kilde: Wei et al., [Emergent Abilities of Large Language Models](#), 2022

Typer LLM med noen eksempler

- Kommersielle lukkede systemer
 - OpenAI ChatGPT
 - Anthropic Claude
 - Google Gemini
- Open source modeller
 - Meta Llama
 - Mistral
 - Alibaba Qwen
- Open source modeller kan
 - Kjøres på egen maskinvare (datasikkerhet)
 - Modifiseres (fine-tunes til formål)

Sammenlikn forskjellige LLM



LMSYS Org

The Large Model Systems Organization develops large models and systems that are open, accessible, and scalable. It is currently run by students and faculty members from UC Berkeley Sky Lab.

<https://lmarena.ai>

Teste LLM - benchmarks

- Databaser med spørsmål og svar
- Fokuserer på forskjellige kompetanser
- Hva om LLM trenes spesielt mot de?
 - Forurensing (contamination)
 - Nye tester må stadig utvikles
- Demokrati?
 - Tester basert på stemmer
 - [LMSYS Chatbot Arena Leaderboard](#)
 - [HuggingFace Open LLM Leaderboard](#)

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro	Llama-400b (early snapshot)
Graduate level reasoning <i>GPQA, Diamond</i>	59.4%* 0-shot CoT	50.4% 0-shot CoT	53.6% 0-shot CoT	—	—
Undergraduate level knowledge <i>MMLU</i>	88.7%** 5-shot	86.8% 5-shot	—	85.9% 5-shot	86.1% 5-shot
	88.3% 0-shot CoT	85.7% 0-shot CoT	88.7% 0-shot CoT	—	—
Code <i>HumanEval</i>	92.0% 0-shot	84.9% 0-shot	90.2% 0-shot	84.1% 0-shot	84.1% 0-shot
Multilingual math <i>MGSM</i>	91.6% 0-shot CoT	90.7% 0-shot CoT	90.5% 0-shot CoT	87.5% 8-shot	—
Reasoning over text <i>DROP, F1 score</i>	87.1 3-shot	83.1 3-shot	83.4 3-shot	74.9 Variable shots	83.5 3-shot Pre-trained model
Mixed evaluations <i>BIG-Bench-Hard</i>	93.1% 3-shot CoT	86.8% 3-shot CoT	—	89.2% 3-shot CoT	85.3% 3-shot CoT Pre-trained model
Math problem-solving <i>MATH</i>	71.1% 0-shot CoT	60.1% 0-shot CoT	76.6% 0-shot CoT	67.7% 4-shot	57.8% 4-shot CoT
Grade school math <i>GSM8K</i>	96.4% 0-shot CoT	95.0% 0-shot CoT	—	90.8% 11-shot	94.1% 8-shot CoT

* Claude 3.5 Sonnet scores 67.2% on 5-shot CoT GPQA with maj@32

** Claude 3.5 Sonnet scores 90.4% on MMLU with 5-shot CoT prompting

Eksempel på noen benchmarks for forskjellige LLM

Kilde: [Anthropic](#)

Eksempel: Massive Multitask Language Understanding (MMLU)

A 67-year-old woman comes to the physician for a follow-up examination. She had a pulmonary embolism and required...	professional_medicine	["Cerebral infarction during the hospitalization", "Complication of the IVC filter placement", "Compression o...
A 25-year-old gravida 3 para 2 female is admitted to the	obstetrics_gynecology	["Braxton Hicks contractions", "lower uterine retraction ring", "hypotonic uterine dysfunction", "primary...
A 67-year-old woman comes to the physician for a follow-up examination. She had a pulmonary embolism and required treatment in the hospital for 3 weeks. She had a retroperitoneal hemorrhage; anticoagulant therapy was temporarily discontinued, and she underwent placement of an inferior vena cava (IVC) filter. She had a hematoma that was resolving on discharge from the hospital 2 weeks ago. Today, she says she has had a persistent sensation of tingling and numbness of her left thigh that she did not report in the hospital because she thought it would go away; the sensation has improved somewhat during the past week. Her only medication is warfarin. Vital signs are within normal limits. Examination of the skin shows no abnormalities. Muscle strength is normal. Sensation to light touch is decreased over a 5 x 5-cm area on the lateral aspect of the left anterior thigh. Which of the following is the most likely cause of this patient's decreased sensation?	obstetrics_gynecology	["The findings are clinically and statistically significant", "The findings are clinically insignificant...
	cardiology	["Atrial fibrillation", "Cor pulmonale", "Systemic hypertension", "Tricuspid valve regurgitation"]
	neurology	["Median nerve at the wrist", "Musculocutaneous nerve at the forearm", "Radial nerve at the forearm", "Ulnar nerve...
	biochemistry	["Fatty acid oxidation", "Fatty acid synthesis", "Gluconeogenesis", "Glycogenolysis"]
		["Cerebral infarction during the hospitalization", "Complication of the IVC filter placement", "Compression of the lateral femoral cutaneous nerve", "Hematoma of the left thigh"]

Meanwhile in dentistry ...

New

1. 15+ years of Synthetic Dental Data for your LLM Training/fine tuning

New

2. First ready to use Specialized Sikka.ai Dental LLM – uses 15+ years of data

sikka.ai

1-800-94-SIKKA

2870 Zanker Rd Suite 120, San Jose, CA 95134

info@sikka.ai

Facebook

Twitter

LinkedIn

Terms of Service

Privacy Policy

HIPAA Policy

© 2024 Copyright Sikka Software Corporation. All Rights Reserved.

Log in

or create an account

Er

P

R

Silver

\$35

/ month per location**

Request a Quote

**Read-only. Request a quote for writeback.

Ref

✓ Sikka Marketplace

✓ Trends overview

✓ Practices, Providers, Employers

Gold

\$45

/ month per location**

Request a Quote

**Read-only. Request a quote for writeback.

✓ Sikka Marketplace

✓ Trends overview

✓ Practices, Providers, Employers

Payments

\$125

/ month per location

Request a Quote

✓ Sikka Marketplace

✓ Trends overview

✓ Practices, Providers, Employees SPA

Platinum

\$175

/ month per location**

Request a Quote

**Read-only. Request a quote for writeback.

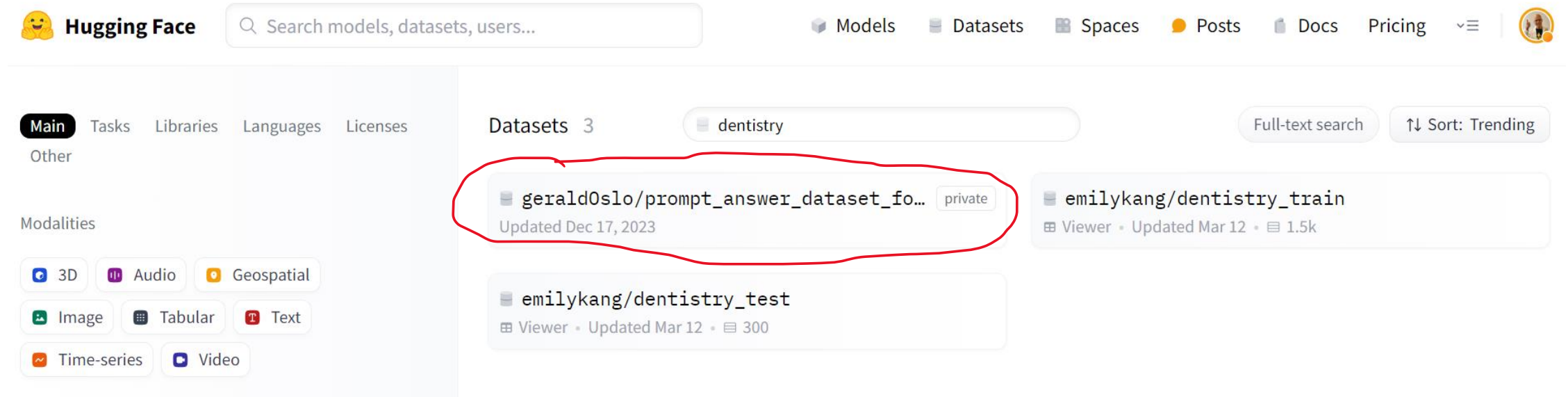
✓ Sikka Marketplace

✓ Trends overview

✓ Practices, Providers, Employers

Men litt finnes ...

- https://huggingface.co/datasets/emilykang/dentistry_train



Det finnes mere på [Kaggle](#), mye bildemateriale (røntgen, foto)

Noen forskningsverktøy

- [Elicit](#): Automate time-consuming research tasks like summarizing papers, extracting data, and synthesizing your findings.
- [LENS.ORG](#): The Lens serves integrated scholarly and patent knowledge as a public good to inform science and technology enabled problem solving.
- [ResearchRabbit](#): Visualize networks of papers and co-authorships. Use graphs as new “jumping off points” to dive even deeper!
- Har selv mest brukt generelle modeller som GPT-4 og Claude 3.5 til å gi meg sammendrag og diskutere innholder i artikler

LLM innen odontologi mulige anvendelser

- Automatisering av journalskriving
- Automatisering av klinikkflyt
- Behandlingsassistent
- Diagnostikk?



Neste episode ...

Live demo med et medisinsk dataset. Kan KI forutsi diabetes?

Seminarserie Kunstig intelligens og meg = sant?

Nevrale nettverk (NN) og maskinlæring

- Perceptronet - nevronet
- Hvordan lærer NN?
- Superenkel anvendelse på medisinske data



Gerald Torgersen
gerald.torgersen@odont.uio.no
Avdeling for kjeve og ansiktsradiologi

Hello AI

- Same place ...
- Same time ...
- Same procedure ...
- Same old presenter 😂