

Lab 3: Generativ AI og Store Språkmodeller

ELMED219-2026

ELMED219

Vår 2026

- 1 Generativ AI
- 2 LLM Transformers
- 3 AI i Medisin
- 4 Etikk og XAI
- 5 Notebooks

Hva er Generativ AI?

Tradisjonell AI (Diskriminativ) skiller mellom klasser (Katt vs Hund). **Generativ AI** skaper nye data som ligner på treningsdataene.

- **Tekst:** ChatGPT, Claude, Gemini (LLM).
- **Bilder:** DALL-E 3, Midjourney, Stable Diffusion.
- **Lyd/Video:** Sora, Suno.

Foundation Models: Modeller trent på enorme mengder data som kan tilpasses mange nedstrøms oppgaver (Few-shot learning).

Kjernen i en LLM er enkel: **Prediker neste ord.**

$$P(w_t | w_{t-1}, w_{t-2}, \dots, w_1)$$

- Ved å lære å predikere neste ord på hele internett, lærer modellen grammatikk, fakta, resonnering og til og med koding.
- **Tokenisering:** Tekst brytes ned til tokens" (delord). "Hospital" \rightarrow ["Hos", pital"].

Transformer-arkitekturen (2017)

Det store gjennombruddet: **"Attention is All You Need"** (Vaswani et al.).

Tidligere modeller (RNN/LSTM) leste tekst sekvensielt (glemsk). Transformere leser hele setningen samtidig og bruker **Self-Attention**.

Prompt

En abstrakt, kunstnerisk representasjon av Transformer-arkitekturs 'Attention'-mekanisme. Glødende strømmer av tekst og kode flyter inn i et sentralt prisme og brytes om til strukturert kunnskap. Ordet 'Attention' (Oppmerksomhet) dannes av lysstrålene. Futuristisk, eterisk, nyanser av lilla og blått. Bredt format."

Transformer Self-Attention



Attention Mekanismen (Matematisk)

Hvert ord (Query Q) sjekker relevansen mot alle andre ord (Key K) for å hente informasjon (Value V).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

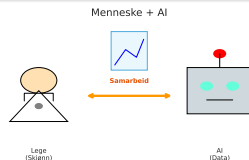
- **Query:** "Hva leter jeg etter?"
- **Key:** "Hva inneholder jeg?"
- **Score (QK^T):** Hvor godt matcher vi?
- **Softmax:** Normaliser til sannsynligheter (vekter).
- **Value:** Informasjonen som sendes videre.

Bruksområder for LLM i Helse

- 1 **Automatisering av dokumentasjon:** Lytte til pasientsamtale og generere journalnotat automatisk (Sparer tid!).
- 2 **Oppsummering:** Trekke ut nøkkelinfo fra lange, komplekse journaler ("Har pasienten historikk med hjertesvikt?").
- 3 **Pasientkommunikasjon:** Oversette medisinsk sjargong til forståelig språk ("Hva betyr 'myokardinfarkt'?").
- 4 **Beslutningsstøtte:** Foreslå differensialdiagnoser basert på symptomer.

Prompt

En lege i hvit frakk som holder et nettbrett, stående ved siden av en vennlig, holografisk AI-assistent avatar. De ser på en pasients holografiske medisinske journal sammen. Scenen symboliserer samarbeid. Myk, betryggende lyssetting. Profesjonell fotostil blandet med 3D-render elementer."



Hvordan snakker vi med modellen for å få best resultat?

- **Zero-shot:** Spør direkte. Klassifiser denne teksten.”
- **Few-shot:** Gi eksempler. ”Her er to eksempler på sarkasme... Klassifiser denne:”
- **Chain-of-Thought (CoT):** Be modellen tenke høyt. Tenk steg for steg.” (Dramatisk bedre på resonnering!).
- **Rollespill:** Du er en erfaren kardiolog...”

Farene: Hallusinasjoner og Bias

Hallusinasjoner: LLM-er er stokastiske papegøyer". De kan finne opp fakta som ser overbevisende ut.

- *Eksempel:* Finne opp referanser til forskningsartikler som ikke finnes.

Bias (Skjevhet): Modellene er trent på internett. De arver våre fordommer.

- Kan diskriminere basert på kjønn, etnisitet eller sosioøkonomisk status i medisinske råd.

Forklarbar AI (XAI)

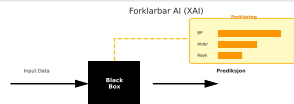
For å stole på AI i klinikken, må vi forstå den.

Metoder:

- **SHAP (Shapley Additive exPlanations):** Hvor mye bidro hver feature til resultatet? (Spillteori).
- **LIME:** Tilpasser en enkel modell lokalt rundt prediksjonen.
- **Saliency Maps:** For bilder (som vi så i Lab 2).

Prompt

En 'Black Box' (Sort Boks) kube som er sprukket opp, og avslører glødende tannhjul og logiske kretser på innsiden. Lysstråler skinner ut fra sprekke og lyser opp et dokument merket 'Forklaring'. Metafor for Forklarbar AI (XAI). Mørk bakgrunn med dramatisk lyssetting."



- ❶ **01-introduksjon-genai.ipynb**: API-oppsett og enkle kall.
- ❷ **02-transformer-arkitektur.ipynb**: Se på attention-vektorer i praksis.
- ❸ **04-prompt-engineering.ipynb**: Teste ulike strategier for journalsammendrag.
- ❹ **06-ai-etikk-medisin.ipynb**: Prøve å lure modellen til å gi skadelige råd (Jailbreaking) og diskutere sikring.
- ❺ **08-nevrosymbolsk-ai.ipynb**: Kombinere LLM med strukturerte data (Knowledge Graphs).

Avslutning av Kurset

Vi håper ELMED219 har gitt dere en verktøykasse for fremtiden.

- AI kommer ikke til å erstatte leger...
- ...men leger som bruker AI vil erstatte leger som ikke gjør det.

Takk for følget!