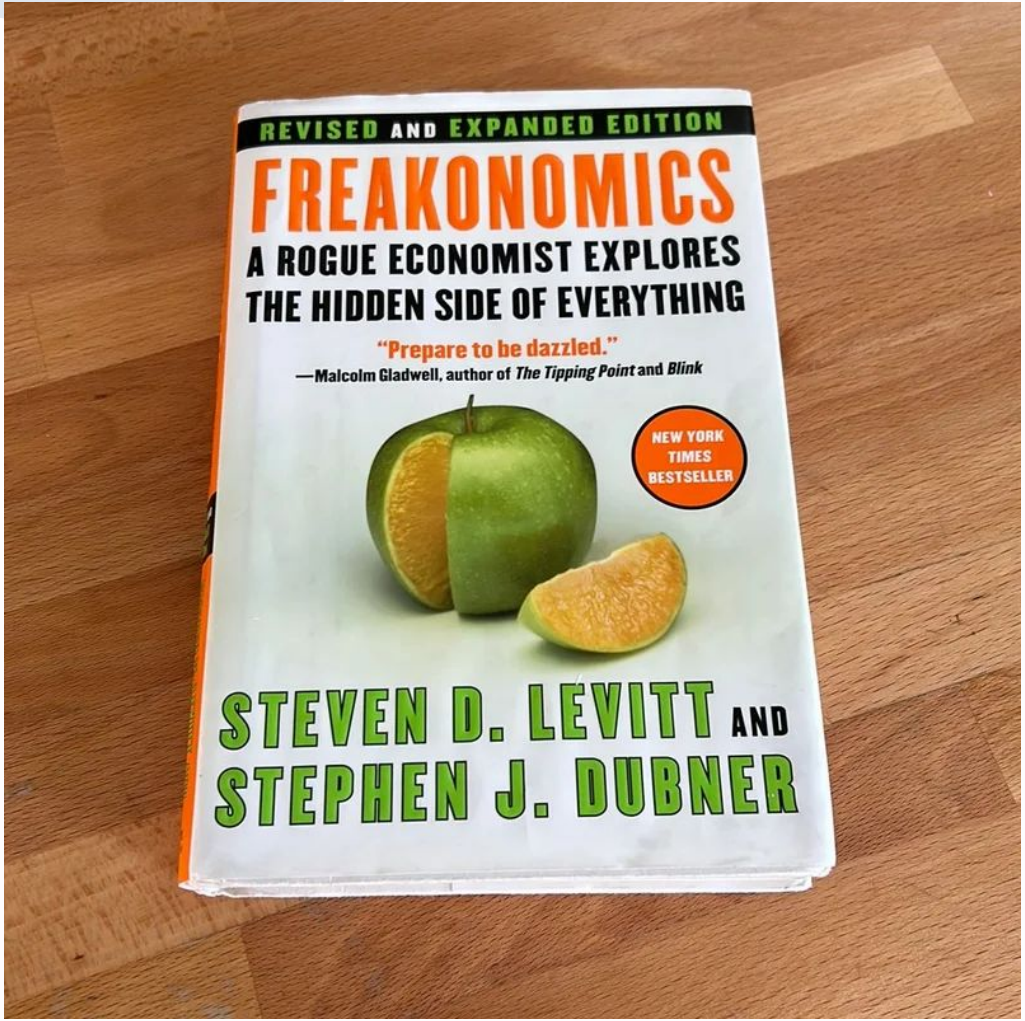

REVISED AND EXPANDED EDITION
FREAKONOMICS
A ROGUE ECONOMIST EXPLORES THE HIDDEN SIDE OF EVERYTHING
"Prepare to be dazzled."
—Malcolm Gladwell, author of The Tipping Point and Blink
NEW YORK TIMES BESTSELLER
STEVEN D. LEVITT AND STEPHEN J. DUBNER


Telefon
Navigation
Radio

Anarkistisk software udvikling

- Intet staging/dev.
- Valgfri branching, du må gerne pushe til main.
- Ingen Azure, GCP, AWS - vi selfhoster til 1/50 af prisen.
- Valgfrit om man vil skrive tests.
- Fucker vi op? Ja, men vi kan også rette hurtigt op.

Dansk GPT

01
Hvad er DanskGPT?

DanskGPT

DanskGPT er bygget på en open source sprogmodel fra Meta. Specifikt er modellen en Large Language Model. LLM'er udskiller sig ved at forstå kontekst og semantik.

Foundation Modeller

En foundation model er en model trænet på billioner (milliarder * 1000) af ord.

Disse modeller koster typisk mere end 10 millioner kr. at træne.

En foundation model er trænet til at færdiggøre sætninger.

Meta har udgivet to generationer af disse modeller; LLaMA 1 og LLaMA2. DanskGPT er trænet på LLaMA2, LLaMA3, Mistral og en masse andre modeller.

Hvordan virker den?

Computere forstår ikke ord, så for at træne en sprogmodel kræver det, at man omdanner ord til tal. Denne process kaldes *tokenization.*

Et token er en repræsentation af et sub-ord som et tal. Dette kan f.eks. være: HEJ ville blive til [44, 302, 1502]

Øvelse 1

a 0	b 1	c 2	d 3	e 4	f 5	g 6
h 7	i 8	j 9	k 10	l 11	m 12	n 13
o 14	p 15	q 16	r 17	s 18	t 19	u 20
v 21	w 22	x 23	y 24	z 25	æ 26	ø 27
å 28	UPPERCASE 29	MELLEMRUM 30	”-” 31			

7 4 9 30 12 4 3 30 9 4 17

Øvelse 2

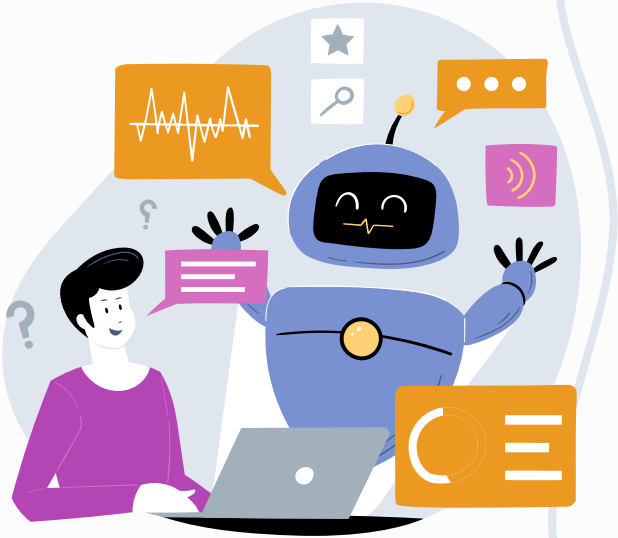
a 0	b 1	c 2	d 3	e 4	f 5	g 6
h 7	i 8	j 9	k 10	l 11	m 12	n 13
o 14	p 15	q 16	r 17	s 18	t 19	u 20
v 21	w 22	x 23	y 24	z 25	æ 26	ø 27
å 28	UPPERCASE 29	MELLEMRUM 30	”-” 31			

29 7 0 11 11 14 30 12 4 3 30 9 4 6

02
Historien bag

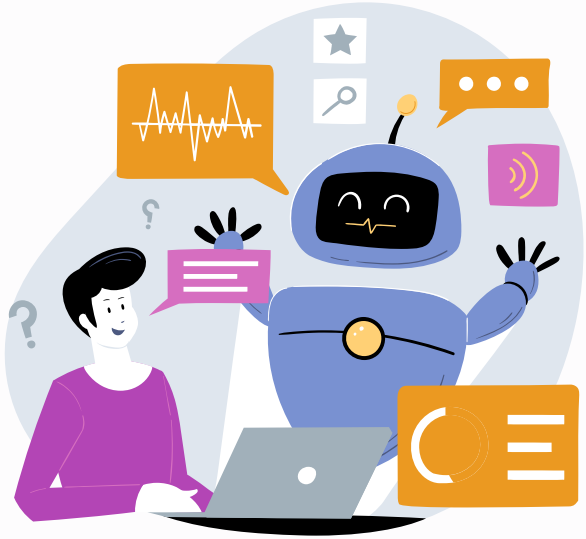
Forløb

- Julefrokost 2022
- Domsdatabasen // AI Dommer - December 2022
- NanoGPT - December 2022
- LLaMA 1 - Januar 2023
- Alpaca - Marts 2023
- OpenLlama - Maj 2023
- LLaMA 2 - Juli 2023
- DanskGPT offentlig - Oktober 2023
- DanskGPT er bedre end ChatGPT 3.5 - Februar 2024
- DanskGPT er bedre end GPT 4o - Juli 2024



Hvorfor?

- Privatlivshensyn
- Troede ikke at store sprogmodeller kunne lære et nyt sprog
- Ingen havde gjort det før



03

Hvordan er den trænet?

Data indsamling

For at lære en sprogmodel et sprog, kræver det store mængder data.

Data til sprogmodeller er sjovt nok tekst.

Webscraping af:
Wikipedia, lex.dk, folketinget, retsinformation, skat.dk, h-sø, twitter, open subtitles, europarl, common-crawl, gutenberg, wikibooks, wikisource, danavis, reddit

Efterbehandling af data inkluderer:

- Fjerne dupliketter
- Fjerne tekster uden ord
- Fjerne tekster med en længde på under 40 karakterer

Alt i alt giver det ca. 3 millarder ord.

Nyeste udgave af DanskGPT er trænet på syntetisk data.

Tid, Penge & Regnekraft

DanskGPT er trænet over to stadier:

1. Lære modellen dansk
    a. 3 milliarder danske ord
    b. 17 dages compute på 8x A100
2. Besvare instruktioner
    a. Ca. 800.000 instruktion/svar par
    b. 10 dages compute på 1x A100

Arbejdet med DanskGPT har indtil videre taget ca. 950 timer, og har produceret 40 modeller. Omkostningerne til regnekraft har været ca. 100.000 kr.

At generere instruktioner/svar par er en iterativ proces.

Gratis compute

04

Hvad kan den bruges til?

Prompt?

"...kort besked på en computerskærm der i forbindelse med en mulighed for indtastning angiver fx hvilken type input der forventes, eller hvilken tilstand programmet aktuelt befinder sig i".

Kort sagt fortælles sprogmodellen hvordan den skal agere til en given opgave.

"Du er en AI-assistent der oversætter. Brugeren vil give dig en tekst på Engelsk som du skal oversætte til dansk."

Hvordan bliver jeg bedre til at prompte?

Du skal bare læse vores guide.

syv.ai/prompting-guide

05

AI act & etik

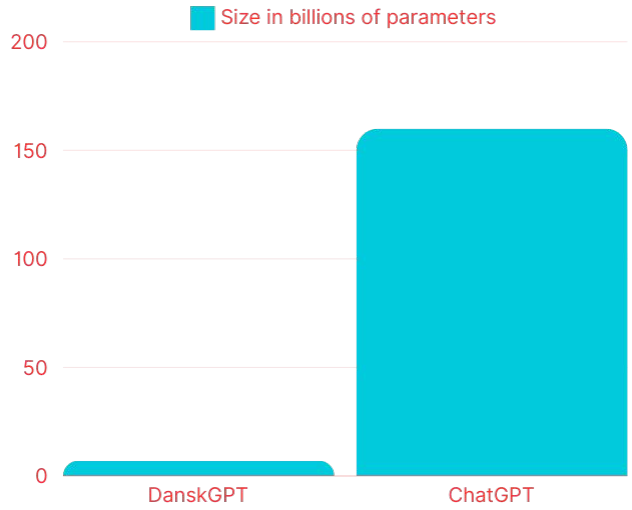
AI Act og etik

AI Act er en forordning fra EU, der er foreløbigt vedtaget. Den regulerer brugen af AI.

1. Unacceptable risk
2. High risk
3. Limited risk
4. Minimal risk

DanskGPT kan alt efter use-case kan ligge mellem High risk og minimal risk.

HOW DOES IT COMPARE IN BENCHMARKS?



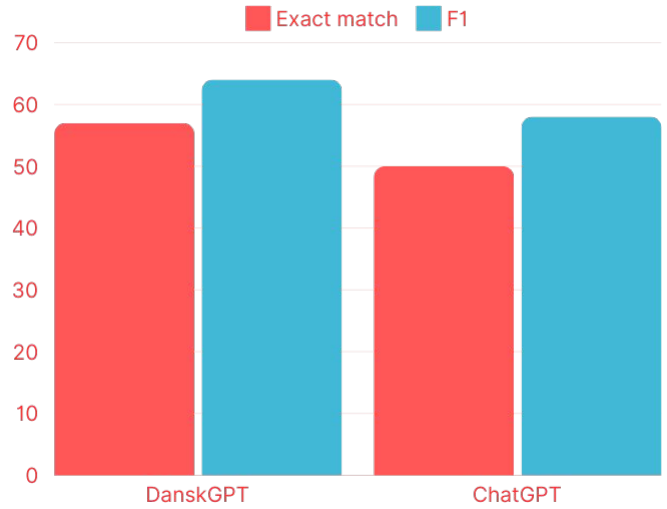
MORE BENCHMARKS?

Classification

of AngryTweets

Danish knowledge

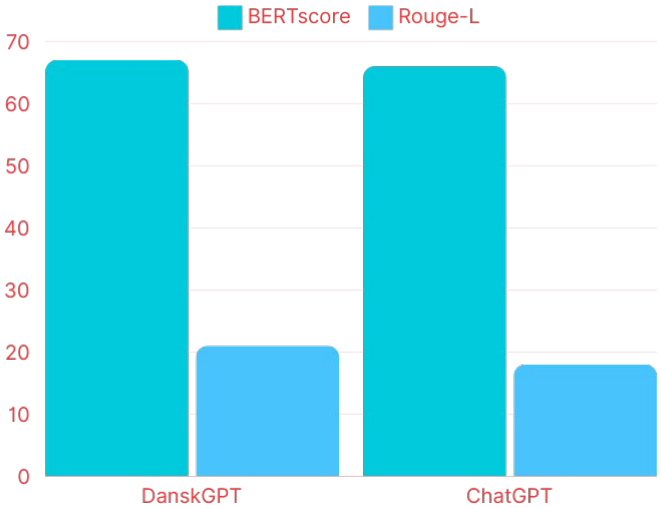
of ScandiQA-da



MORE?

Summarization

of Nordjylland News



Speed

at batch size = 1

06

Teste sprogmodeller

Testmetoder

1. Givet et input er output 1:1 hvad vi forventer?
2. Måle afstand af semantik (embeddings)
3. Sikre at modellen er ens = Logit bias
4. LLM-in-the-loop

1. str = str
2. str -> List[float, ...] ≈ List[float, ...]
3. List[float, ...] = List[float, ...]
4. str -> int

Tokens = Hun : 1, n: 2 en: 3, dø: 4, r: 5, er: 6
Hunnen dør = 1, 2, 3, 4, 5
Hun er en dør = 1, 6, 3, 4, 5

Demo af DanskGPT

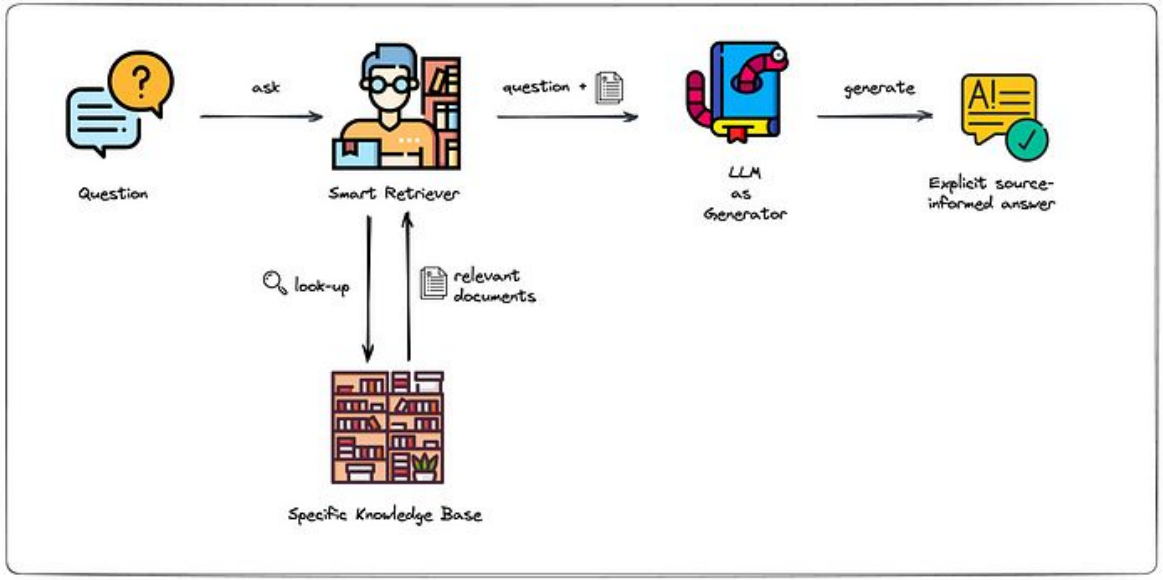
Hop ind og prøv

https://chat.danskgpt.dk

RAG

Retrieval Augmented Generation

RAG er en metode hvortil virksomheder kan få svar på deres egen data uden at skulle træne en model.



Spørgsmål?

Hviske

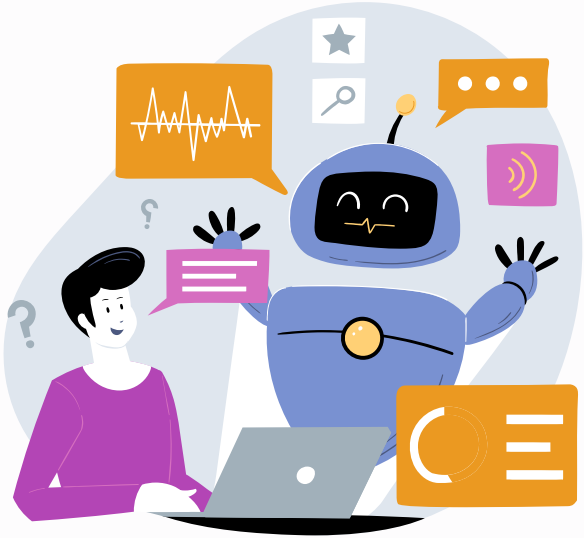
Hviske

Hviske er bygget ovenpå OpenAI's Whisper v3.

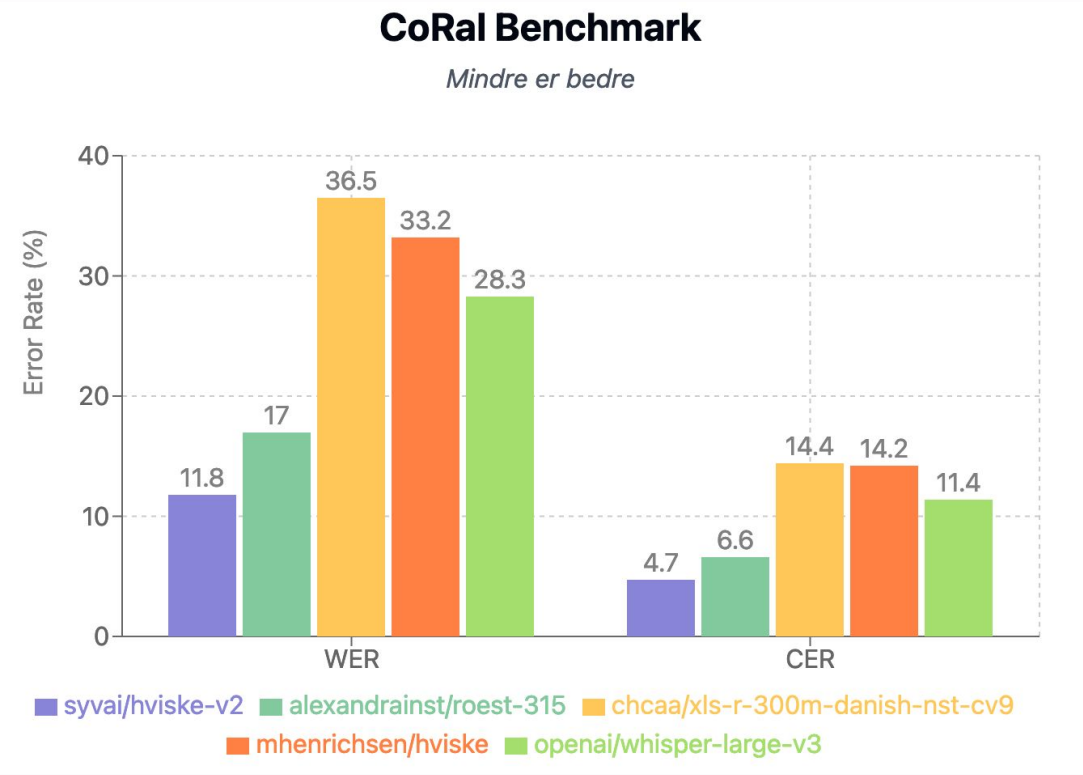
Hviske er en transskriberingsmodel, der ligesom DanskGPT, er optimeret til det danske sprog.

Hvordan virker det?

- Lyd -> 16khz
- 16kHz -> tokens
- tokens -> model
- model -> tokens
- tokens -> tekst



Flot, men er den god?



Lad os prøve den

https://ludwig.syv.ai/

Hvad nu?

ailex

TRÆN JERES EGEN LLM

DANSKGPT

SYV.AI

LUDWIG.SYV.AI

MADS HENRICHSEN
MADS@SYV.AI