

UNIVERSITETET I OSLO

Autotekst.uio.no (Whisper)

Transcribing research data and
video

DEIC 2023

Dagfinn Bergsager
Assistant Director IT at UiO



In Norway, we want a society where everyone can participate.
Therefore, universal design of ICT is a legal requirement
for both public and private sector.

www.uutilsynet.no/

Timeline: UiO Autotekst beginnings

- ⌘ Attempt to “solve” speech-to-text for UiO
 - ✦ autumn of 2020
- ⌘ Google speech-to-text API
- ⌘ Launched Autotekst.uio.no in 2021
- ⌘ Lots of work after processing
 - ✦ Boring



**Google Cloud
Speech API**



21. September 2022

New opportunities



September 21, 2022
7 minute read

Introducing Whis

We've trained and are open-sourcing a that approaches human level robustness speech recognition.

READ PAPER

VIEW CODE

VIEW MODEL CARD

Whisper examples:



REVEAL TRANSCRIP

openai / whisper Public

Code Pull requests 10 Discussions

main

Go to file

jongwook invoking __call__ in... 6 days ago

data	initial commit	2 mo
notebooks	initial commit	2 mo
tests	initial commit	2 mo
whisper	invoking __call__ instead ...	6 c
.gitignore	initial commit	2 mo
LICENSE	initial commit	2 mo
MANIFES...	initial commit	2 mo
README...	invoking __call__ instead ...	6 c
approach...	initial commit	2 mo
language...	add section Available mo...	2 mo
model-ca...	fixes the link to the model...	2 mo
requirem...	Add conda environment.y...	2 mo
setup.py	Add package metadata to...	las

README.md

Robust Speech Recognition via Large-Scale Weak

Alec Radford^{*1} Jong Wook Kim^{*1} Tao Xu¹ Greg Brockman¹ Christine Me

Abstract

We study the capabilities of speech processing systems trained simply to predict large amounts of transcripts of audio on the internet. When scaled to 680,000 hours of multilingual and multitask supervision, the resulting models generalize well to standard benchmarks and are often competitive with prior fully supervised results but in a zero-shot transfer setting without the need for any fine-tuning. When compared to humans, the models approach their accuracy and robustness. We are releasing models and inference code to serve as a foundation for further work on robust speech processing.

1. Introduction

Progress in speech recognition has been energized by the development of unsupervised pre-training techniques exemplified by Wav2Vec 2.0 (Baevski et al., 2020). Since these methods learn directly from raw audio without the need for human labels, they can productively use large datasets of unlabeled speech and have been quickly scaled up to 1,000,000 hours of training data (Zhang et al., 2021), far more than the 1,000 or so hours typical of an academic supervised dataset. When fine-tuned on standard benchmarks, this approach has improved the state of the art, especially in a low-data setting.

These pre-trained audio encoders learn high-quality repre-

methods are exceedingly a training dataset which boot from the same dataset. How brittle and spurious and de and distributions. In a parti ford et al. (2021) docume classification accuracy whe model on the ImageNet da without observing any im when classifying the same image datasets. A model t formance when trained on basic errors when evaluate because it is exploiting th humans are oblivious to (C

This suggests that while un proved the quality of audio of an equivalently high-qu bined with a recommended tuning, is a crucial weakne and robustness. The goal should be to work reliably of environments without re a decoder for every deploy

As demonstrated by Naraya et al. (2020), and Chan et al tems that are pre-trained in datasets/domains exhibit h much more effectively to

UiO Autotekst

Finally speech-to-text that works reasonably well in Norwegian



UNIVERSITETET
I OSLO

Last opp til transkribering



Dra og slipp filen din her, eller klikk for å velge fil

Inndataspråk

Automatisk gjenkjenn

Oversett transkriberingen til engelsk

Transkriberinger

Se alle Filer i kø Ferdige

test_av_arbeidsmetoder_2.mp3 Lagt i kø 27. okt. 2022

test_av_arbeidsmetoder_2.mp3 Lagt i kø 27. okt. 2022

test_av_arbeidsmetoder_2.mp3 Transkriberer 27. okt. 2022

Forelesning_1_jernalder.m4a Ferdig 27. okt. 2022

Last ned tekstfil

Forelesning_1_jernalder.m4a Ferdig 27. okt. 2022

Last ned tekstfil

Transkriber med Autotekst

Med Whisper fra OpenAI for transkripsjoner av høy kvalitet

- Last opp lyd- eller videofilen din og motta transkribert fil i ønsket format
- Velg å motta transkripsjonen din på det opprinnelige språket eller oversatt til engelsk
- Tjenesten kjører på sikre servere ved Universitetet i Oslo

Logg inn via Feide

Last opp til transkribering



Dra og slipp filen din her, eller klikk for å velge fil

Inndataspråk

Automatisk gjenkjenn ▾

Oversett transkriberingen til engelsk

Transkriberinger

Se alle Filer i kø Ferdige

test_av_arbeidsmetoder_2.mp3 Lagt i kø 27. okt. 2022

test_av_arbeidsmetoder_2.mp3 Lagt i kø 27. okt. 2022

test_av_arbeidsmetoder_2.mp3 Transkriberer 27. okt. 2022

Forelesning_1_ømalder.m4a Ferdig 27. okt. 2022 Last ned tekstoffil ↓

Forelesning_1_ømalder.m4a Ferdig 27. okt. 2022 Last ned tekstoffil ↓



TEK5130-
introduc...per.mp3

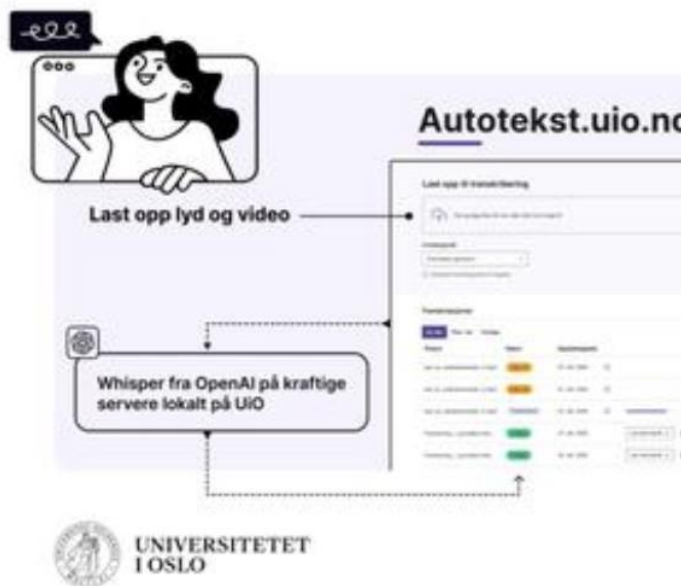
Feedback so far on the Whisper version
- a lot better than alternatives



Svein Stølen

9. februar kl. 19:59 · 🌐

IT-avdelingen ved [Universitetet i Oslo \(UiO\)](#) USIT - er svært kompetent. Nå bruker de automatisk transkribering av lyd og video. **gratulerer.**



UIO.NO

Autotekst i ny versjon: God og enkle på norsk - Universitetet i Oslo

👍 Dagfinn Bergsager, Ingrid Stensland og 10 andre

👍 Liker

💬 Kommenter

digi.no

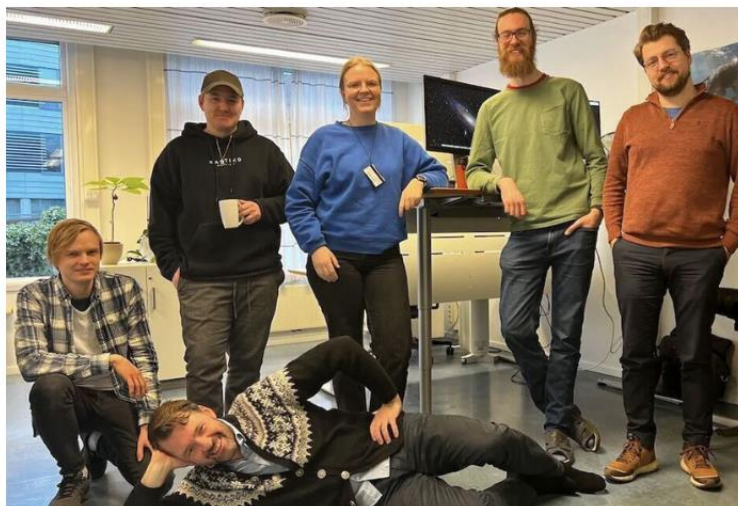


tomm.eriksen@usit.uio.no

DIGITALISERING OG OFFENTLIG IT

Bygde tjeneste som sparer dem for 20 millioner i året: - Det er ny og sjokkerende teknologi

Universitetet i Oslo (UiO) har tatt i bruk en ny og revolusjonerende tjeneste. Med nyvinning sparer de både tid og millioner av kroner.



Teamet bak OpenAI-teknologien som ligger lokalt på UiO



Oslo · 1d

Funker dette bedre enn den lignende funksjonen til Word? 😂



nært · 1d

@2 Ja den er mye bedre enn Word. Også veldig bra på mange dialekter



langt borte · 18h

Naais, går på uiS, skal skrive master neste år. Ser uiS ikke tilbyr noe slikt. Vurdere å melde meg opp til fag på UiO for å få tilgang 🙄

Må bare si at det strømmer inn gode ord om autoteksttjenesten etter oppgradering til whisper! Det er stor glede over hvor bra den har blitt, og så langt har jeg kun hørt om tekster som har vært så og si 100 % korrekt, så med andre ord svært lite etterarbeid. All kred til dere!!

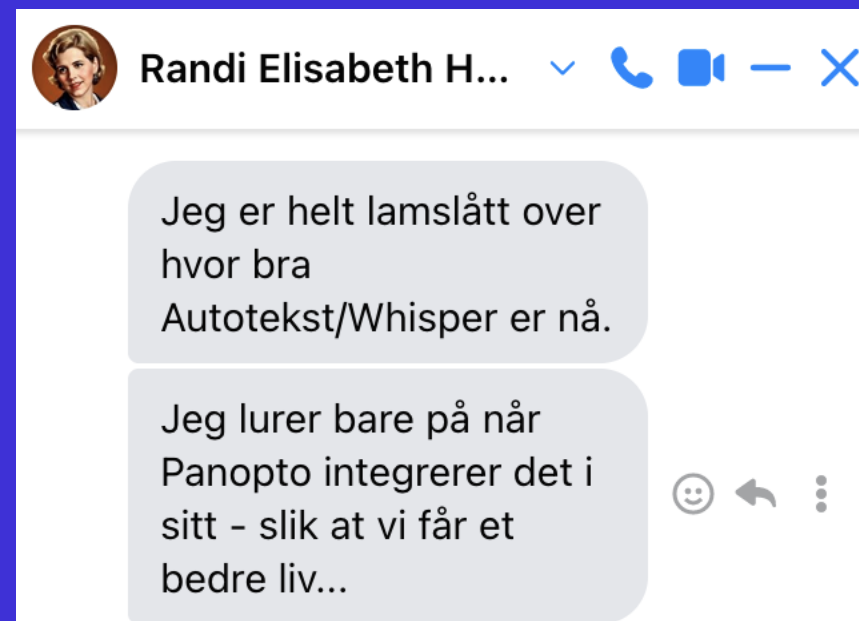
– Marte Veierød, nettredaktør UiO

Nå har jeg allerede fått testet med en video som vi har her, med dårlig lyd og møredialekt. Dette er et kvantesprang

– Kjetil Knarlag, HK-dir

Det er virkelig helt fantastisk at dere har laget dette programmet, det letter en enorm byrde både i tid og penger fra mine og andre forskeres skuldre!

– Eva Stambøl Postdoc UiO

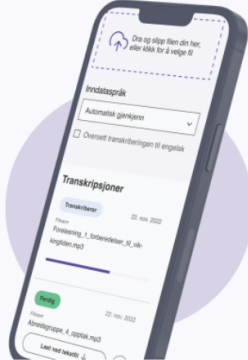


Experience with Autotekst Whisper version so far

- Better than alternatives for Norwegian
- Works well with many dialects
- Cleans up false starts and incomplete sentences
- Many errors are consistent, easier to clean up after
- Quite good at guessing language if not specified
- Works with audio and video files
- Does not distinguish between multiple speakers
- Sometimes bad results: if so, try again!

UNIVERSITETET
I OSLO

Autotekst



Transkriber med Autotekst

Med Whisper fra OpenAI for transkripsjoner av høy kvalitet

[Logg inn via Feide](#)

Har du en Educloud bruker? [Logg inn med Educloud](#)

- Last opp lyd- eller videofilen din og motta transkribert fil i ønsket format
- Velg å motta transkripsjonen din på det opprinnelige språket eller oversatt til engelsk
- Tjenesten kjører på sikre servere ved Universitetet i Oslo

Statistics ⊕

From

01/02/2023 📅

To

07/11/2023 📅

Completed

21898

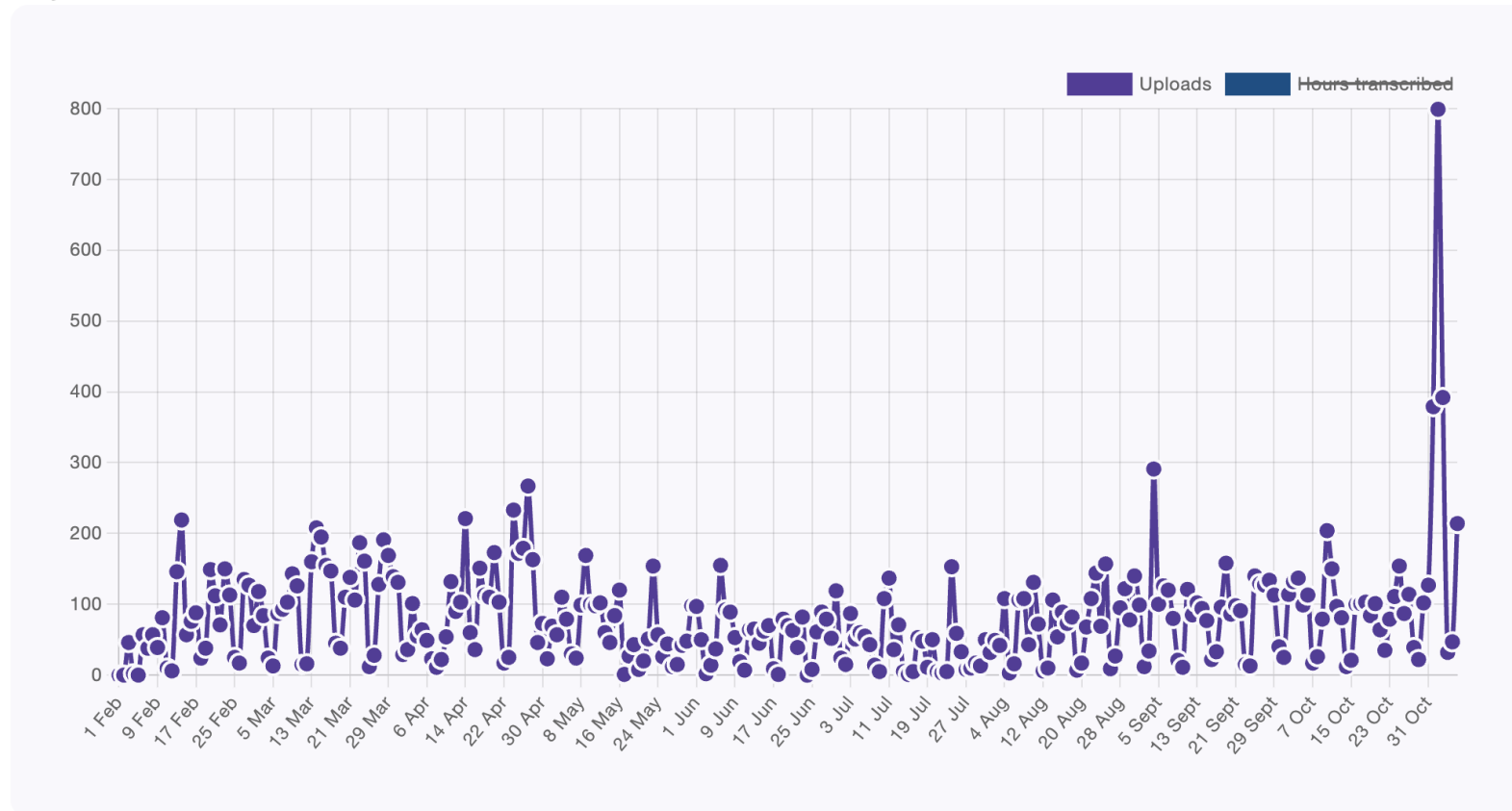
Error

507

Hours transcribed

12039

Day ∨






Last opp lyd og video

Autotekst.uio.no

Last opp til transkribering

 Dra og slipp filen din her, eller klikk for å velge fil






Inndatapråk

Automatisk gjenkjenn

Oversett transkripsjonen til engelsk

Transkripsjoner

Se alle Filer i ka Ferdige

Filnavn	Status	Opplysningsdato	
test_av_arbeidsmeloder_2.mp3	Lagt i ka	27. okt. 2022	
test_av_arbeidsmeloder_2.mp3	Lagt i ka	27. okt. 2022	
test_av_arbeidsmeloder_2.mp3	Transkribert	27. okt. 2022	
Forklaring_1_jemkoder.m4a	Ferdig	27. okt. 2022	Last ned tekst! 
Forklaring_1_jemkoder.m4a	Ferdig	25. okt. 2022	Last ned tekst! 



Whisper fra OpenAI på kraftige servere lokalt på UiO



Transkriberte filer
.vtt .txt .srt



UNIVERSITETET
I OSLO

Takk for oss!



"A picture of a dice showing 6"

Laget av Bing bildeskaper

Leveres av DALL·E 3


Integration with Nettskjema.no

Logg inn

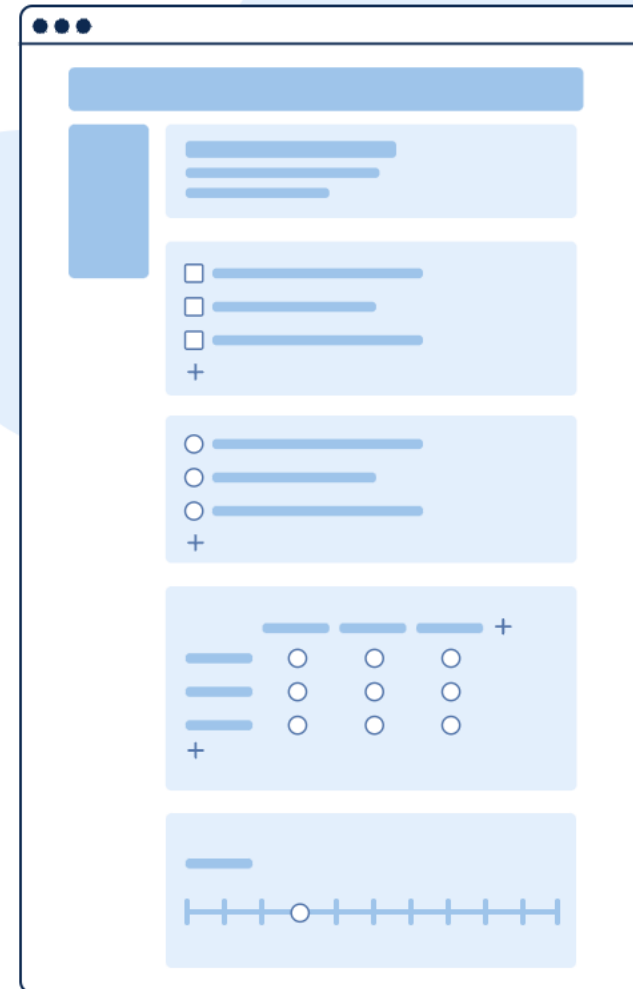
I Nettskjema kan du logge inn på flere ulike måter.
Velg én under:

 Logg inn med FEIDE

 Logg inn med TSD

 Logg inn med ID-porten

[Søk om ID-portenbruker](#)



Nytt opptak

Lagrer i skjema: 257830



00:00:03

00:03

-01:29:56

Vis Bygg skjema Kodebok Innstilling

< Tilbake til se resultater

Opptak

Kommentar Levert svar

Tekstlab 12.09.23 14:38 Lukk

00:00 Så kan jeg ta opp sånn som dette, så kan jeg snakke inn i mobilappe levere dette.

Last ned

00:00

Hastighet 1.00 x

5 sek 5 sek

Volum

Kinesisk 12.09.23 11:51 Lukk

00:00 Ecto, ecto, ecto, taun, pau, taun, taun, pau, teto.

Last ned

.vtt

.vtt

.txt

00:00

Hastighet 1.00 x

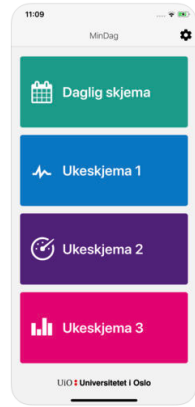
5 sek 5 sek

Volum

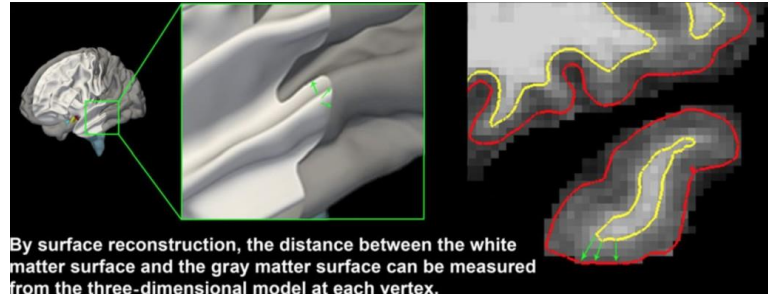
Kinesisk 12.09.23 11:51 Åpne

Test 12.09.23 11:50 Åpne

Scalable safe storage

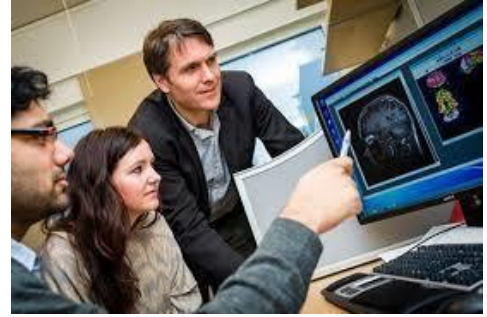


Self Service



By surface reconstruction, the distance between the white matter surface and the gray matter surface can be measured from the three-dimensional model at each vertex.

Compute, analyse and store data



Collaborate / share



Data capture

Nettskjema.no

Consent Service



Analysis

Feedback services



Panopto™

video platform

Whisper at UiO

- ✓ TSD (sensitive data)
- ✓ Autotekst
- ✓ Nettskjema
- ✓ Panopto
- ✓ Additional local versions

Technology for Whisper backend and API server

& Backend

- ✧ Backend server is a “left over” IBM Power server shared with Nettskjema/Autotekst etc.
 - 8 V100 Tensor Core GPU cards running Whisper
 - 2 TB RAM
- ✧ RabbitMQ job que
- ✧ Private S3 storage

& API server

- ✧ HAProxy load balancer
- ✧ Private OKD (Kubernetes OpenShift)
 - nginx
 - Node.js
- ✧ PostgreSQL database





Further plans

Further plans

- ⌘ Whisper X
 - Faster Whsiper
 - Pyannote.audio for "diarization"
- ⌘ Nasjonalbibliotekets model
- ⌘ Scaling
- ⌘ Sensitive data
- ⌘ Norsk språkmodell

Norwegian language model?

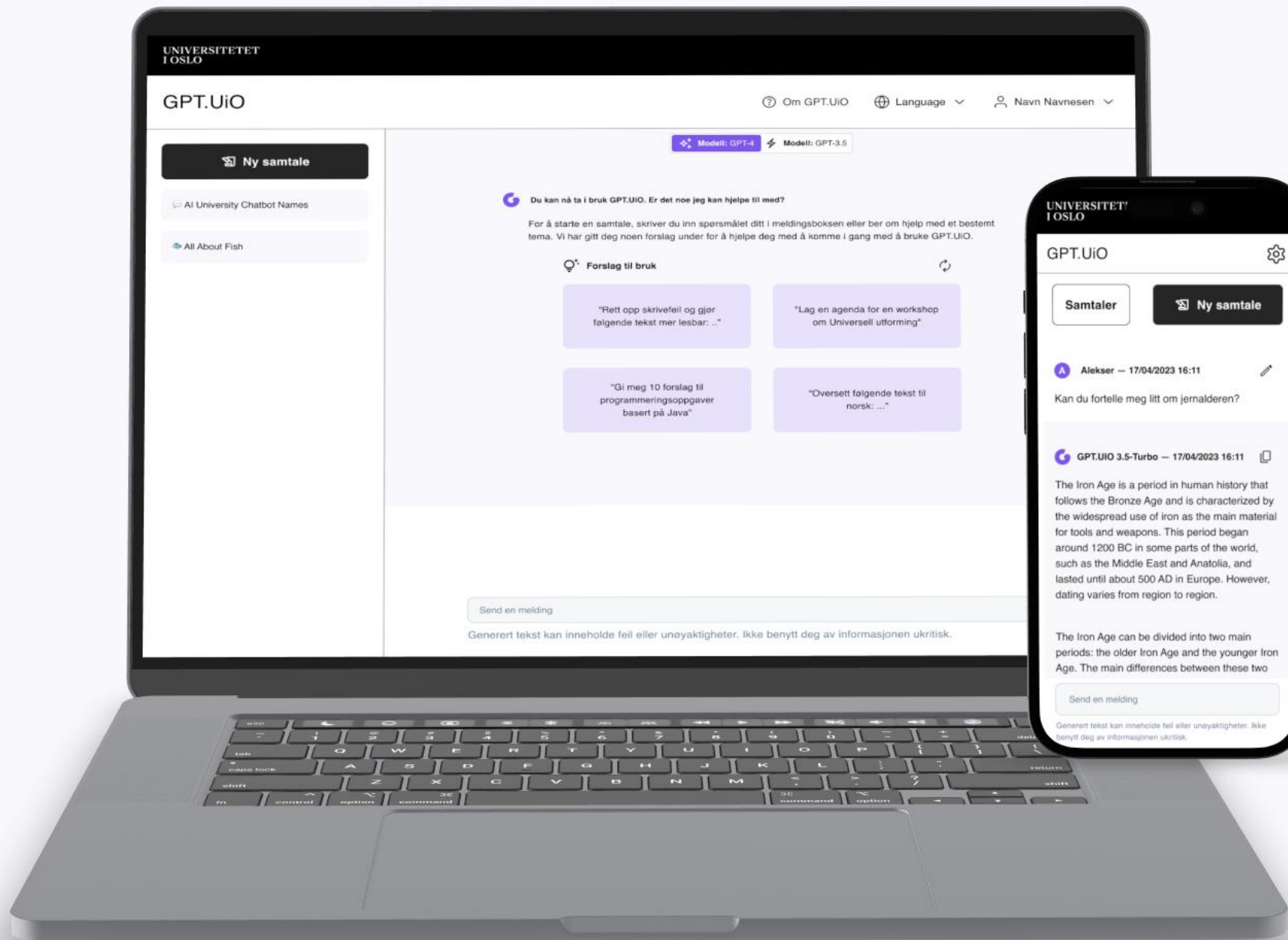
[linkedin.com/in/bergsager](https://www.linkedin.com/in/bergsager)

bonusfoiler

UNIVERSITETET I OSLO

GPT UiO





Ny samtale

AI University Chatbot Names

All About Fish

Modell: GPT-4 Modell: GPT-3.5

Du kan nå ta i bruk GPT.UiO. Er det noe jeg kan hjelpe til med?

For å starte en samtale, skriver du inn spørsmålet ditt i meldingsboksen eller ber om hjelp med et bestemt tema. Vi har gitt deg noen forslag under for å hjelpe deg med å komme i gang med å bruke GPT.UiO.

Forslag til bruk

"Rett opp skrivefeil og gjør følgende tekst mer lesbar: ..."

"Lag en agenda for en workshop om Universell utforming"

"Gi meg 10 forslag til programmeringsoppgaver basert på Java"

"Oversett følgende tekst til norsk: ..."

Send en melding

Generert tekst kan inneholde feil eller unøyaktigheter. Ikke benytt deg av informasjonen ukritisk.

Samtaler

Ny samtale

Aleksr — 17/04/2023 16:11

Kan du fortelle meg litt om jernalderen?

GPT.UiO 3.5-Turbo — 17/04/2023 16:11

The Iron Age is a period in human history that follows the Bronze Age and is characterized by the widespread use of iron as the main material for tools and weapons. This period began around 1200 BC in some parts of the world, such as the Middle East and Anatolia, and lasted until about 500 AD in Europe. However, dating varies from region to region.

The Iron Age can be divided into two main periods: the older Iron Age and the younger Iron Age. The main differences between these two

Send en melding

Generert tekst kan inneholde feil eller unøyaktigheter. Ikke benytt deg av informasjonen ukritisk.

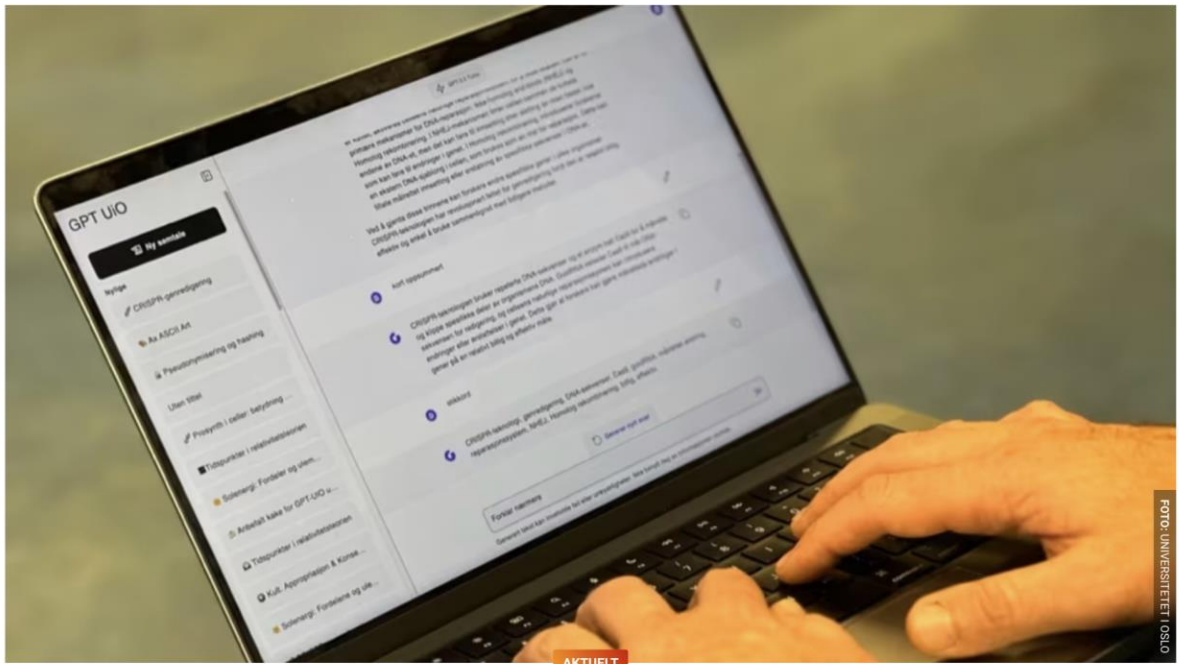
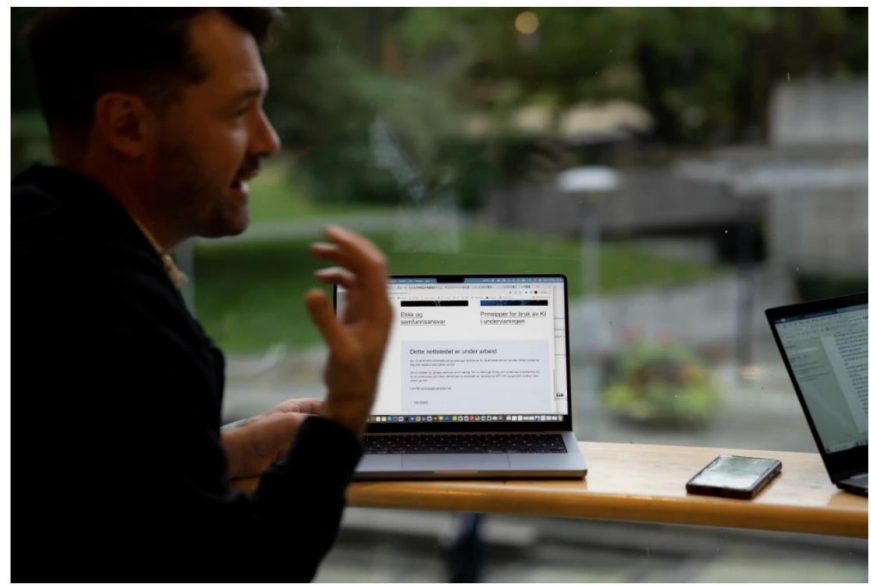


FOTO: UNIVERSITETET I OSLO

AKTUELT

Tusenvis av studenter bruker den nye tjenesten: – Positiv respons

NTNU utelukker ikke at det blir nye «ChatGPT»-løsninger hos dem etter hvert



KI: UiO hiver seg på kunstig intelligens-bølgen.
Foto: Isabel Svendsen Berge

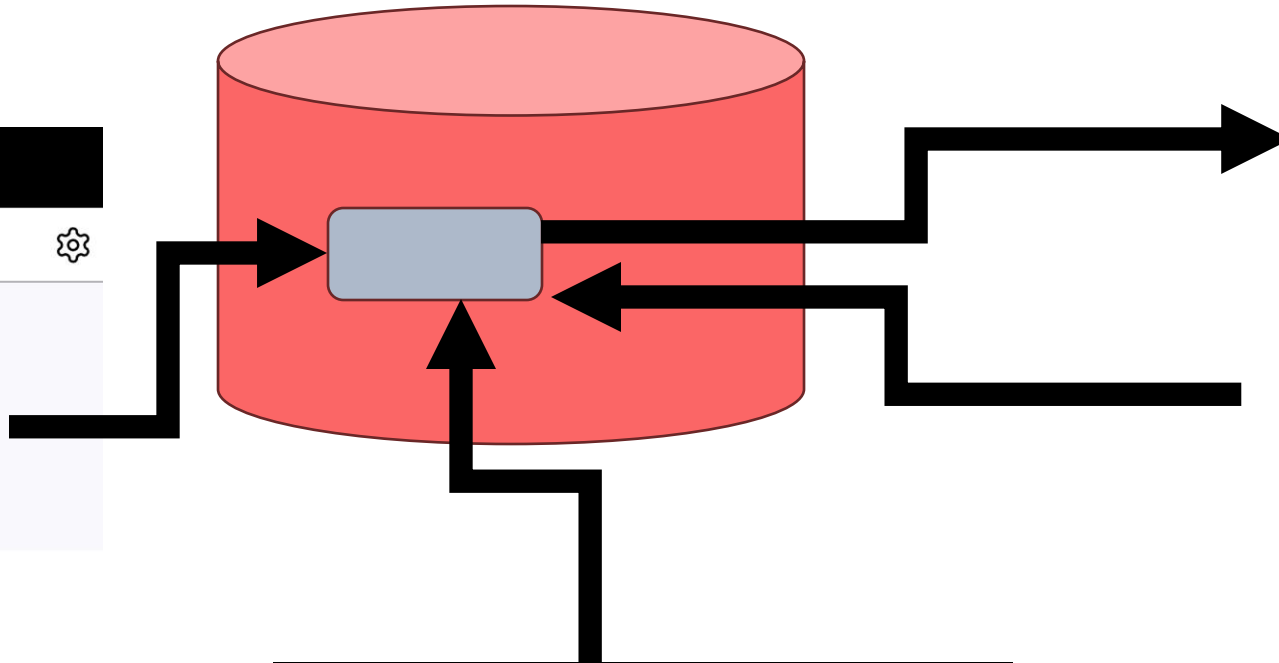
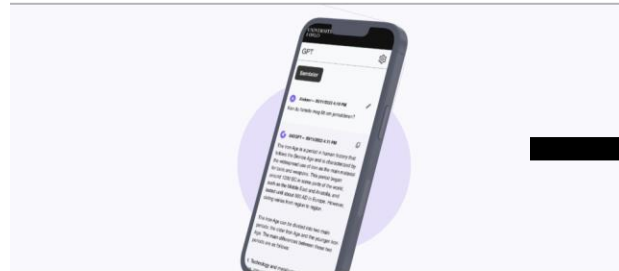
UiO lanserer ChatGPT

Kunstig intelligens har blitt sluppet fri på Blindern. På tross av tendenser til løgn og halvsannheter er prosjektets prøvekaniner positivt overrasket.

Opptak fra åpent møte om GPT UiO

Oppsummering av juridiske vurderinger i utviklingen av GPT UiO, samt gjennomgang av teknisk oppsett fra webinarer vi hadde i oktober. Se demo av GPT UiO ca. 24 minutter ut i [videoen](#).









GPT UiO

- Tilpasset ansatte og studenter ved UiO
- Kan brukes til behandling av grønne og gule data


Logg inn via Feide

GPT UiO


Ny samtale

- Tal på indonesisk: 1-10
- Heltallsforko...   
- God IT-lagspiller
- Sykefraværsspørsmål i inte...
- Mobbing in English?
- Blockchain Beyond Crypto
- Akselerasjon vs. hastighet 
- Max. Sausages?


hva er absolutt null

UIOGPT — 30/08/2023, 21:54 


Absolutt null er det matematiske uttrykket for tallet 0 i positiv form. Absoluttverdien til et tall angir avstanden mellom tallet og origo på tallinjen uten å ta hensyn til retning. Siden avstanden mellom 0 og 0 er 0, er absoluttverdien av null lik 0, og dermed kan vi si at absolutt null er 0.

Dagfinn — 30/08/2023, 21:55 

Hva er N forkortelse for

UIOGPT — 30/08/2023, 21:55 

N er forkortelsen for de naturlige tallene. Naturlige tall er alle positive heltall, det vil si fra og med 1 og oppover. Derav betegnelsen "naturlige", siden de ofte brukes til å telle objekter i naturen eller hverdagen vår. Symbolene for de naturlige tallene er vanligvis N eller \mathbb{N} .

Send en melding 



GPT-3.5 Turbo og GPT-4 for studenter og ansatte



Forslag på bruk



Personvern i fokus. Data sendes ikke utenfor Europa.

The screenshot shows the GPT UiO chat interface. On the left is a sidebar with a 'Ny samtale' button and a list of 'Favoritter' (Solenergi: fordeler og ulemper) and 'Nylige' (Prinsipper i maskinlæring og n..., Prinsipper bak maskinlæring o..., Tidspunkter i relativitetsteorien, CRISPR-genredigering, Akselerasjon vs Hastighet) items. The main chat area shows a conversation with GPT-3.5 Turbo. The user asks for an explanation of machine learning and neural networks. The AI responds with a detailed explanation of neural networks, their structure, and their applications. The user then asks for more details on neural networks, and the AI explains the layers and training process. At the bottom, there is a 'Generer nytt svar' button and a 'Send en melding' field.

GPT UiO

Ny samtale

Favoritter

- Solenergi: fordeler og ulemper

Nylige

- Prinsipper i maskinlæring og n...
- Prinsipper bak maskinlæring o...
- Tidspunkter i relativitetsteorien
- CRISPR-genredigering
- Akselerasjon vs Hastighet

GPT-3.5 Turbo

K

Forklar meg prinsippene bak maskinlæring og nevralt nettverk

Prinsippene bak maskinlæring involverer å bruke algoritmer og statistisk modellering for å lære å utføre spesifikke oppgaver basert på data. Istedenfor å programmere en algoritme for å utføre en oppgave direkte, lærer maskinlæringsystemet å utføre oppgaven gjennom trening med store sett av data og feedback. En av de mest populære metodene i maskinlæring er nevralt nettverk.

Nevrale nettverk er et sett med algoritmer som er inspirert av strukturen og funksjonen til hjernen. De er satt sammen av mange enkle prosesseringselementer, eller "neuroner", som samhandler gjennom tilkoblinger vektet av numeriske verdier. Disse tilkoblingene og vektene justeres gjennom trening med data for å produsere ønskede utganger fra inndata. Ved hjelp av bakoverstryking kan nettverkene justeres slik at de kan generalisere og utføre korrekt når de blir presentert med nye data.

Denne egenskapen, sammen med evnen til å håndtere store mengder data, gjør nevralt nettverk spesielt egnet til å løse komplekse problemer innen bilde- og talegjenkjenning, naturpråkprosessering og automatisering av fysiske systemer.

Kan du utdype mer om nevralt nettverk

Selvfølgelig!

Nevrale nettverk består av et stort antall enkle prosesseringselementer, eller "neuroner", som er pakket sammen i layers. Hver neuron i et lag er koblet til alle andre neuroner i det forrige laget og sender signaler videre til det neste laget. Disse signalene mellom neuronene er vektet med tall som representerer styrken på koblingen mellom dem.

Trening av et nevralt nettverk innebærer å presentere det med et sett med treningsdata, som består av en inndata og en ønsket utgang. Nettverket justerer vektene mellom neuronene for å redusere differansen mellom den faktiske utgangen og ønsket utgang, basert på en feilfunksjon. Dette kalles bakoverpropagering.

Jo mer treningsdata som brukes, desto bedre vil det trænede nettverket være for å generalisere og produsere riktige utganger når det oppdager nye data. Til slutt kan det nevralt nettverket brukes til å utføre oppgaver ved å presentere inndata og produsere en utgang.

Nevrale nettverk har vist seg effektive i en rekke områder, inkludert bilde- og talegjenkjenning, oversettelse, smarte taleassistenter, autonome kjøretøy og finansprognoser. Ettersom teknologien utvikler seg, vil trolig bruksområdene for nevralt nettverk fortsette å ekspandere.

Generer nytt svar

Send en melding

Generert tekst kan inneholde feil eller unøyaktigheter. Ikke benytt deg av informasjonen ukritisk.



eee

eee



Hva er bedre med GPT UiO?

- Avtaler
- Dataflyten
- UU
- Tilpasninger i GUI
 - Promts
 - Risks...

