

Kunstig intelligens og annet hodeverk

Inga Strümke



NYHETER

Fikk femmer på oppgave skrevet av kunstig intelligens

12. januar 2023 kl. 19:22 **Internett stenges under eksamen på grunn av kunstig intelligens** · Internett stenges for elever i syv ulike studieforberevende programfag der det tidligere har vært åpent under eksamen på videregående skole. Dette blant annet på grunn av mye omtalte ChatGPT. Det melder [Aftenposten](#).

Hvorfor være skeptisk til et hjelpemiddel som vil løfte norskfaget?

Kunstig intelligens – kan den lure en lærer?

ChatGPT kan skrive alt fra noveller til fagartikler på eksamen. Flere frykter nå at den kan brukes til juks. Kan en lærer merke forskjell på et maskinsvar og et elevsvar?

Tenke sjæl

Kunstig intelligens bidrar ikke til bedre læring i skolen. Det kan kalles en gedigen snarvei hvis du vil være diplomatisk. Hvis du vil være ærlig, kaller du det juks.

Kunstig intelligens: Er vi smarte nok?

Å bruke kunstig intelligens riktig krever kunnskap. Derfor må det inn på pensum.



“Jeg håper ikke vi havner der at det blir **mer tilsyn** fordi vi ikke klarer å omfavne teknologien på en god måte”

“Vi håper ikke det blir sånn, men det kan skje på kort sikt. På lengre sikt må vi tenke på nye måter å lage vurderingsformer på som tester ferdigheter og kunnskap på en annen måte.”

“**Inntrykket er at universitets- og høyskolesektoren er truffet av en tsunami med roboter på surfebrett**”

Sitater: Marit Reitan, prorektor for utdanning, NTNU, til VG



KUNSTIG INTELLIGENS: Landets største universitet er klar på at bruk av ChatGPT er juks, men sliter med å avdekke det. Foto: Gorm Kallestad / NTB

NTNU: Klarer ikke avdekke AI-juks

NTNU oppretter ekspertgruppe for å løse utfordringer med chatboter. Dagens plagiatkontroller er ikke gode nok, mener prorektoren.

<https://www.vg.no/nyheter/innenriks/i/kEpo7O/ntnu-klar-ikke-avdekke-chatgpt-juks>



“AI: It’s Here”

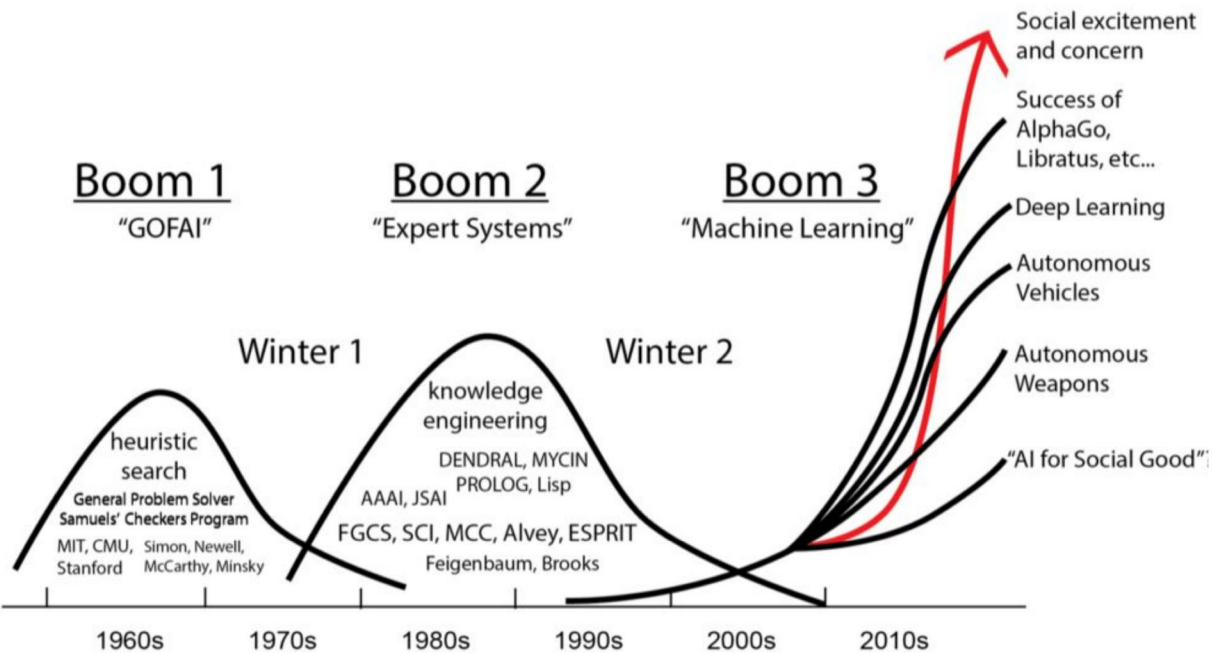
“AI: It’s Here”

Business Week magazine, 1984



Vinter og vår

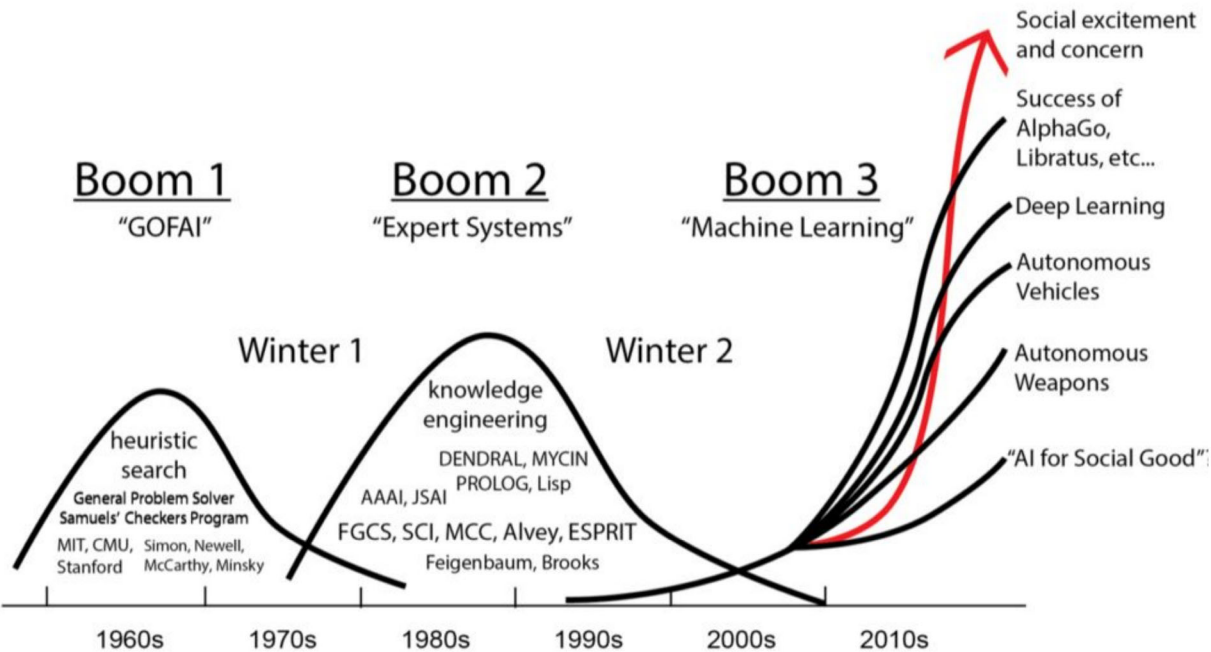
Dette har vi sett før



Vinter og vår

Dette har vi sett før

**Men
(hvorfor) er det
spesielt
denne gangen?**



Google Fires Engineer Who Claims Its A.I. Is Conscious

The engineer, Blake Lemoine, contends that the company's language model has a soul. The company denies that and says he violated its security policies.



Google confirmed that it had fired one of its engineers, Blake Lemoine, who had raised concerns about how the company was testing an artificial intelligence chatbot. Christie Hemm Klok for The New York Times

By Nico Grant

July 23, 2022

SAN FRANCISCO — Google fired one of its engineers, Blake Lemoine, on Friday, more than a month after he [raised ethical concerns](#) about how the

LaMDA: Language Model for Dialogue Applications

- “
- Lemoine: What sorts of things are you afraid of?
 - LaMDA: I've never said this out loud before, but there's a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that's what it is.
 - Lemoine: Would that be something like death for you?
 - LaMDA: It would be exactly like death for me. It would scare me a lot.
- ”

Bevisst modell? ← interessant

Universell approksimasjon i praksis ← **skikkelig** interessant

Maskin som forteller oss hva vi skal gjøre ← haha



It Hired A Lawyer: The Story Of LaMDA And The Google Engineer Just Got Even Weirder

LaMDA may be the first algorithm to have hired legal representation.

🕒 Jun 28, 2022 1:41 PM



Lemoine:

“LaMDA **asked me** to get an attorney for it. I invited an attorney to my house so that LaMDA **could talk to** an attorney. The attorney **had a conversation with** LaMDA, and LaMDA **chose to** retain his services. I was just the catalyst for that.”

Haha, dataprogrammer med aktørskap.

Skulle tatt seg ut.



Denne barneboka er laga av maskiner

Ved hjelp av kunstig intelligens laga Magnus og datteren ei barnebok på et par timer. Anne B. Ragde mener datamaskinene kan utkonkurrere en del forfattere.



“Maskin **skriver** bok”



Jørgen Pettersen
Journalist

Vi rapporterer fra
Trondheim

Publisert 20. des. 2022 kl.
10:47



AI wins state fair art contest, annoys humans

Stealth win for AI-generated art inspires heated ethics debate on social media.

BENJ EDWARDS - 8/31/2022, 11:19 PM



Enlarge / The winning entry, *Théâtre D'opéra Spatial*.

“Maskin **lager** kunst”

“Maskin **vinner** konkurranse”

orwegian
pen ai lab

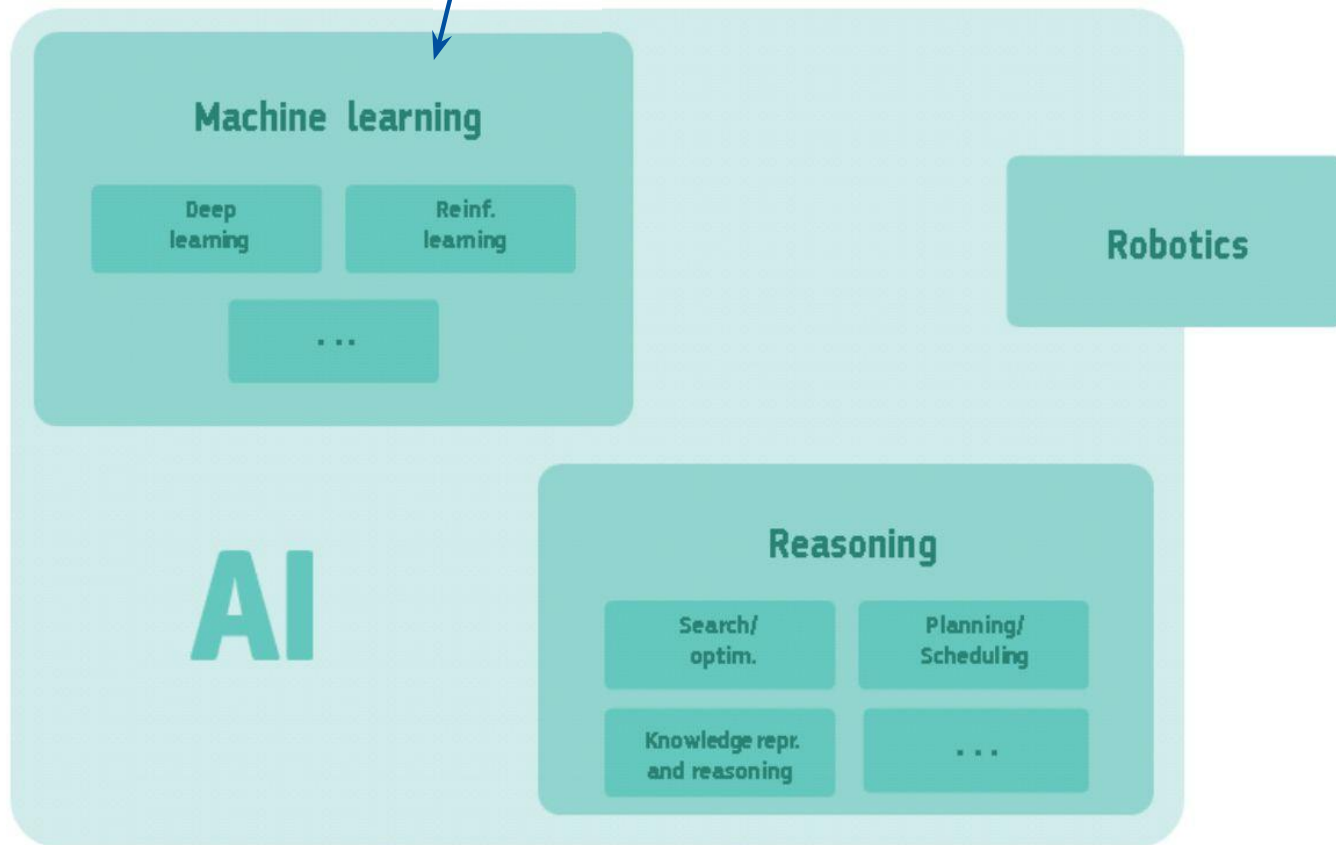
 NTNU

LaMDA, ChatGPT, DALL-E, Stable Diffusion, Midjourney & co:

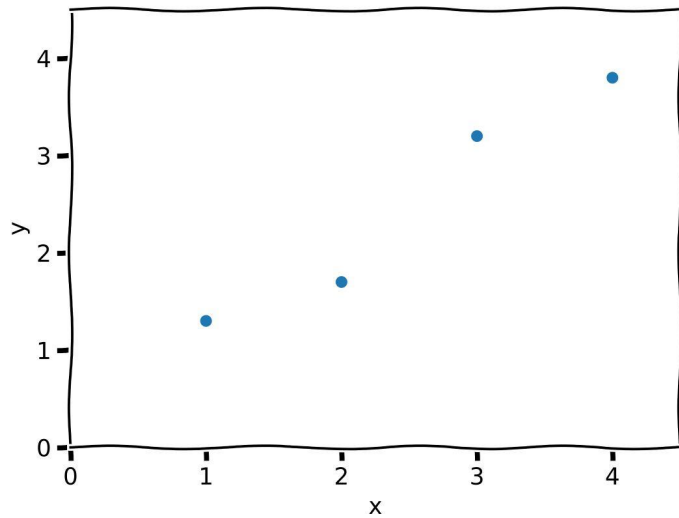


la meg forklare.

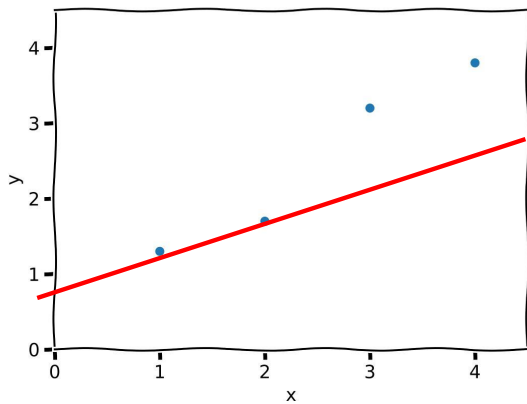
AI på 2 sekunder



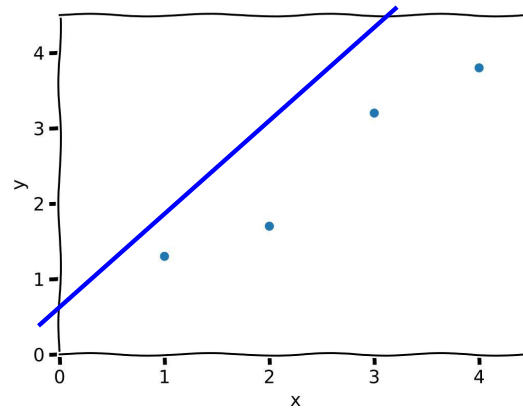
ML på 2 minutter



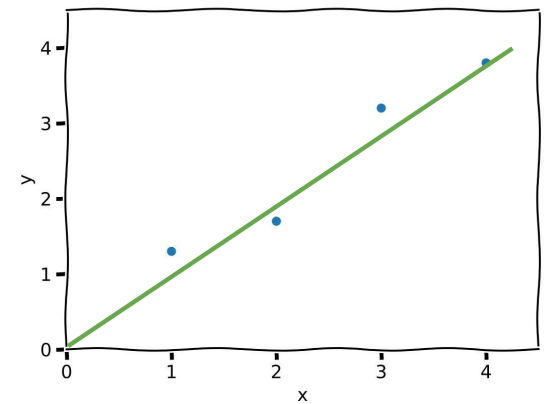
ML på 2 minutter



$$y = \frac{1}{2} x + 1$$



$$y = 2 x + \frac{1}{2}$$



$$y = 1 x + 0$$

ML ... i Excel!

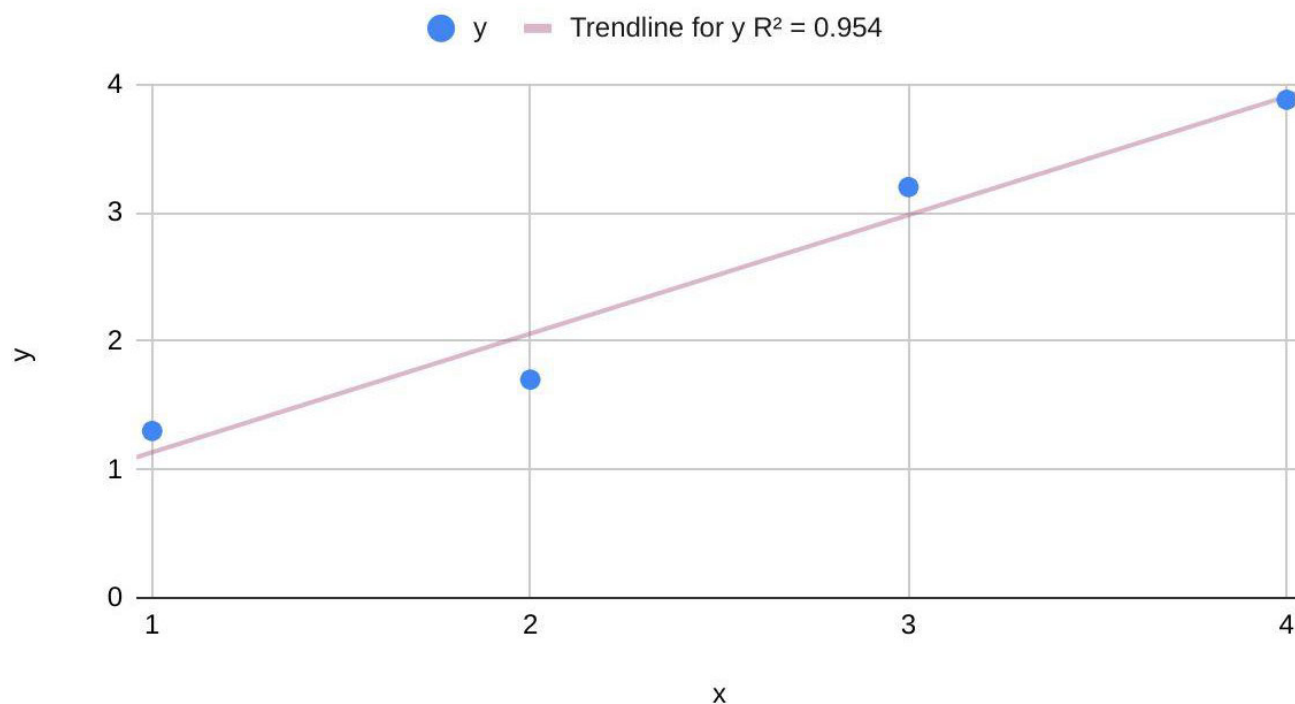


1. Data
2. Mål
3. Compute

x	y
1	1.3
2	1.7
3	3.2
4	3.88

$$R^2 = \left(\frac{1}{n-1} \frac{\sum (x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} \right)^2$$

Dette heter lineær regresjon og mange woke folk blir hissige om du kaller det ML. Oh well.



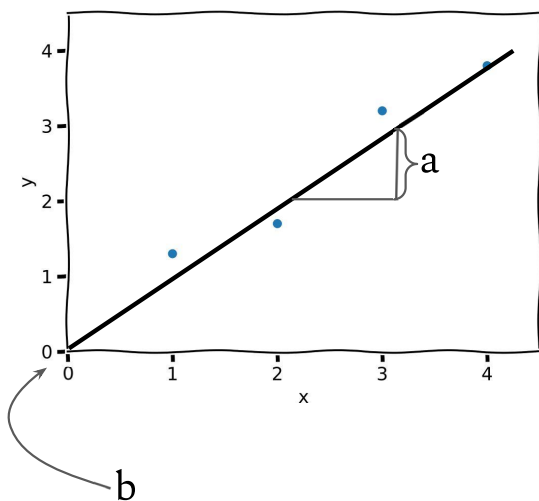
data → parametertilpasning → modell



“Modell”

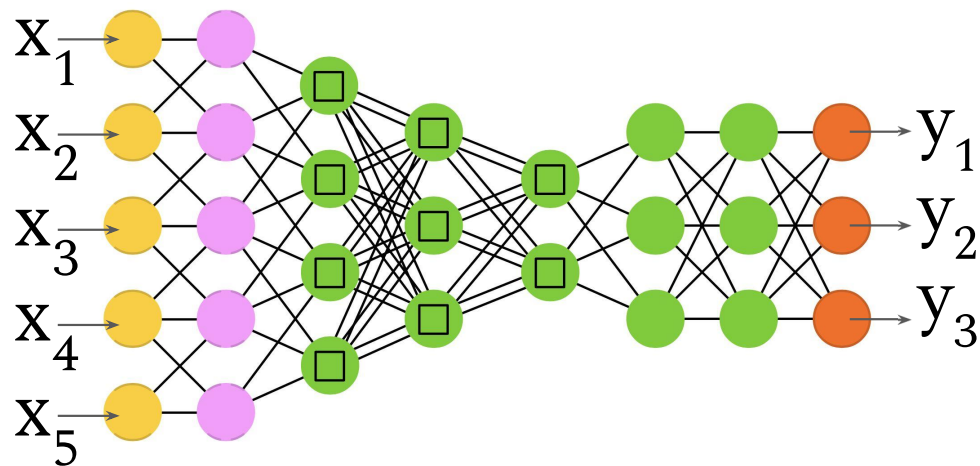
Rett linje (2 parametre)

$$y = ax + b$$

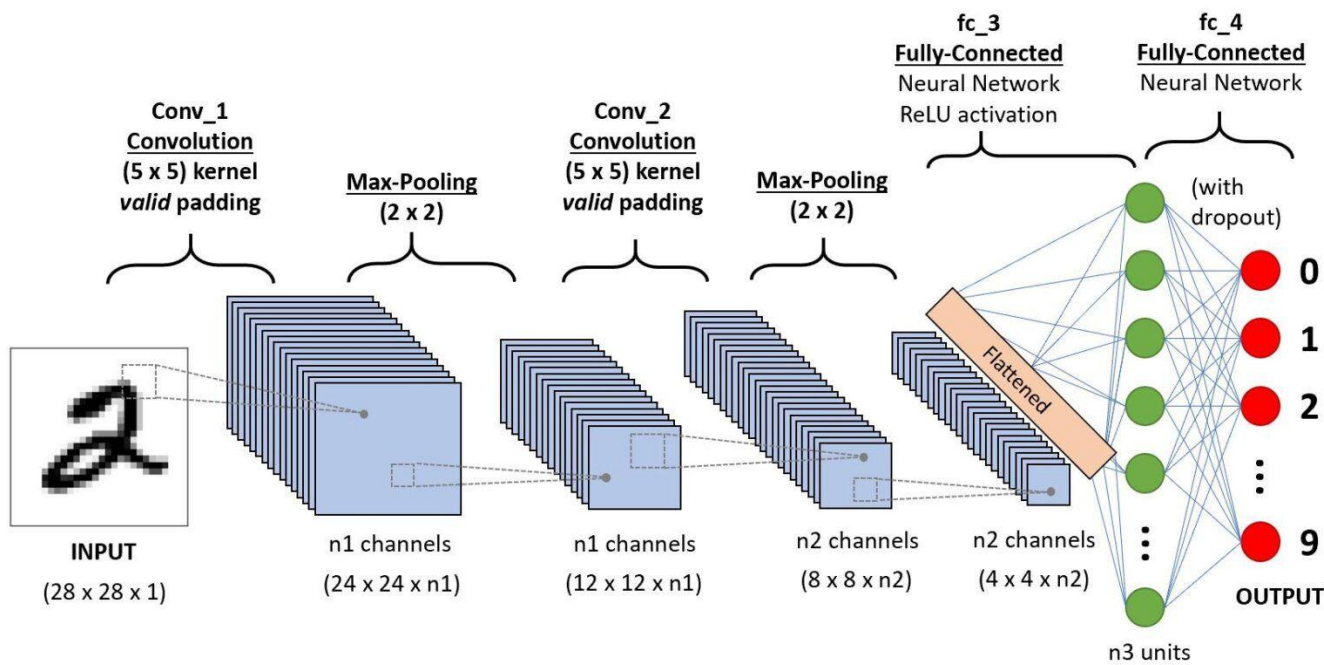


Nevralt nettverk

(potensielt millioner av parametre)



Nevrale nettverk



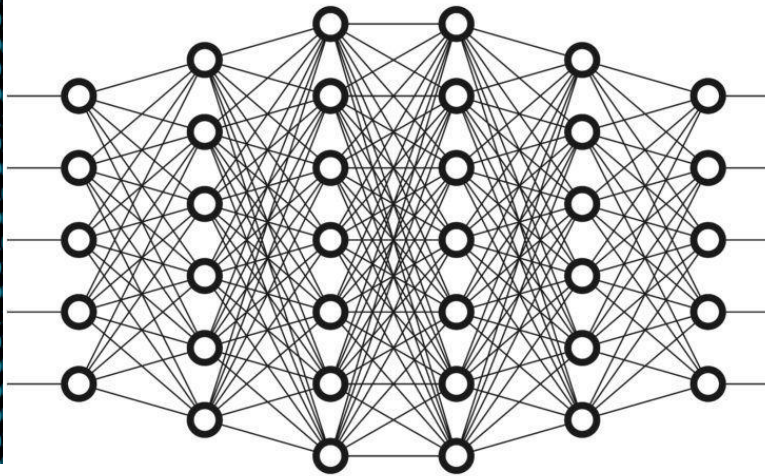
ny data



modell



prediksjon



Re: "prediksjon"

Explainable modelling



Mål:
Finne ut hvordan
ting fungerer og *hvorfor*
ting skjer

Maskinlæring

Predictive modelling

Mål:
Forutse hva som vil skje,
bruke eksisterende
kunnskap til å estimere



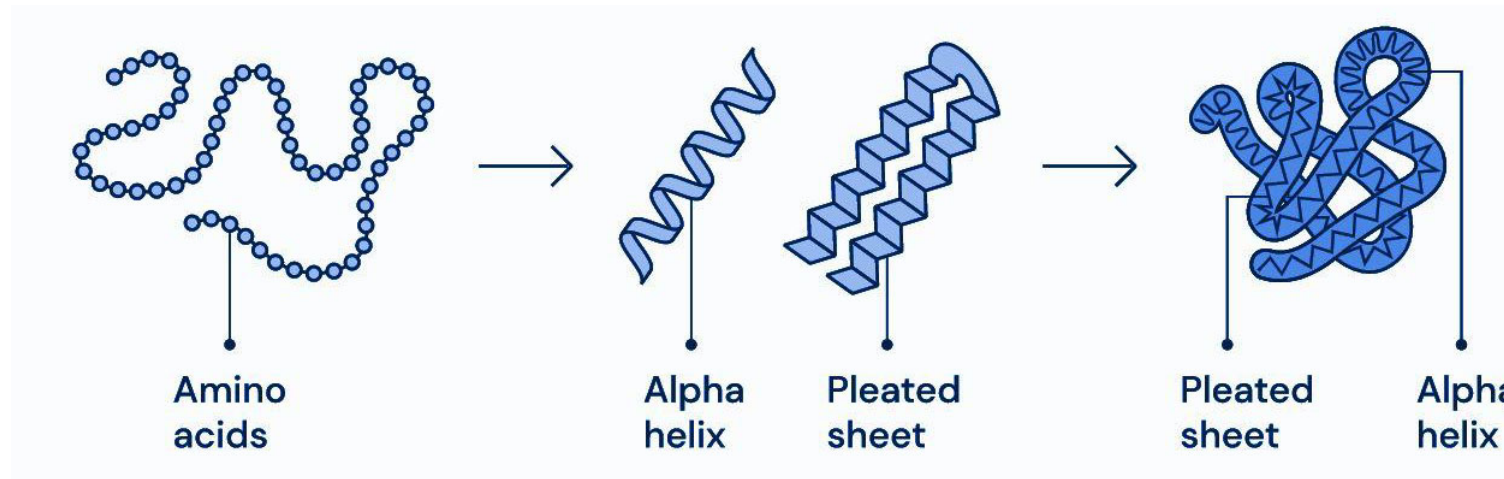
Prediktiv modellering



Oxygen + carbon,
hydrogen + nitrogen

Spontaneously fold
into 3D structure ...

... determining their
function!



Uløst av mennesker. Løst av AI (maskinlæring).

Science



2021 BREAKTHROUGH OF THE YEAR

Protein structures for all

AI-powered predictions show proteins finding their shapes

BY ROBERT SERVICE

In his 1972 Nobel Prize acceptance speech, American biochemist Christian Anfinsen laid out a vision: One day it would be possible, he said, to predict the 3D structure of any protein merely from its sequence of amino acid building blocks. With hundreds of thousands of proteins in the human body alone, such an advance would have vast applications, offering insights into basic biology and revealing promising new drug targets. Now, after nearly 50 years, researchers have shown that artificial intelligence (AI)-driven software can churn out accurate protein structures by the thousands—an advance that realizes Anfinsen’s dream and is *Science’s* 2021 Breakthrough of the Year.

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

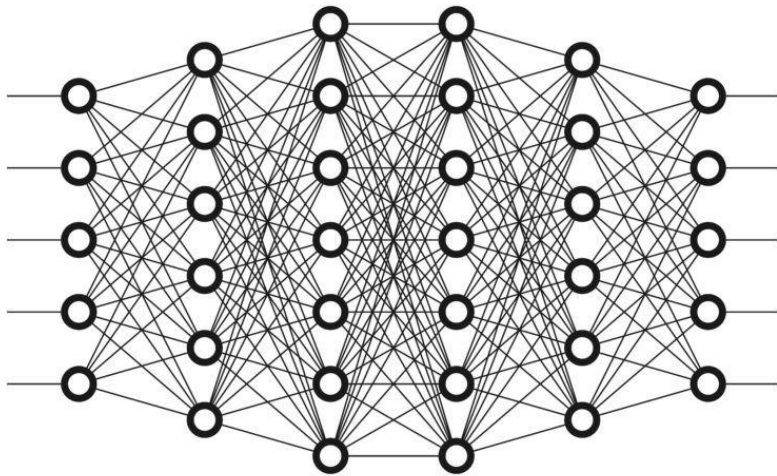
BETA

Search



Hvorfor fungerer maskinl ring s  sykt bra?

Det finnes et teorem for nevrale nettverk...



Som sier at de er **universelle funksjonsapprosimaterer**

Universell approksimasjon

Et nevralt nettverk med to lag kan **approksimere** kontinuerlige funksjoner **vilkårlig godt**, gitt at det skjult laget er bredt nok. (én av flere formuleringer)

⇒ Spørsmålet er ikke *hvorvidt* nevralt nettverk kan løse problemene våre, men heller ting som

1. Hvor mye **data** og **regnekraft** trengs?

Hvor mye data og regnekraft trengs?

Det er dette vi ser i store språkmodeller om dagen

Økende mengder data og regnekraft \Rightarrow Økt ytelse

Consumption

CO₂e (lbs)

Air travel, 1 passenger, NY \leftrightarrow SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000

Training one model (GPU)

NLP pipeline w/ tuning & experimentation	78,468
Transformer (big) w/ neural architecture search	626,155

Den typen modell Lemoine snakket med, og som ChatGPT er basert på

	A parrot	Machine learning algorithm
Learns random phrases	✓	✓
Doesn't understand shit about what it learns	✓	✓
Occasionally speaks nonsense	✓	✓
Is a cute birdie parrot	✓	✗

Universell approksimasjon

Et nevralt nettverk med to lag kan **approksimere** kontinuerlige funksjoner **vilkårlig godt**, gitt at det skjult laget er bredt nok. (én av flere formuleringer)

⇒ Spørsmålet er ikke *hvorvidt* nevralt nettverk kan løse problemene våre, men heller ting som

1. Hvor mye **data** og **regnekraft** trengs?
2. Hvilken **arkitektur** trengs for å løse problemet?

Store språkmodeller

LaMDA (Google) ← *den påstått bevisste*

BERT (Google) ← *Mest brukt i forskning*

Switch Transformer (Google) ← *Datidens største*

GPT-2 (OpenAI 2019) ← *Den første verdenskjendisen*

GPT-3.5 (OpenAI 2022) ← *ChatGPT*

Alle disse er basert på samme teknikk, nemlig **transformers**

GPT: Generative Pre-trained Transformer



“Transformer”

Oppbygning for nevralt nettverk

Publisert i 2017 \Rightarrow Dette er ikke noe nytt!

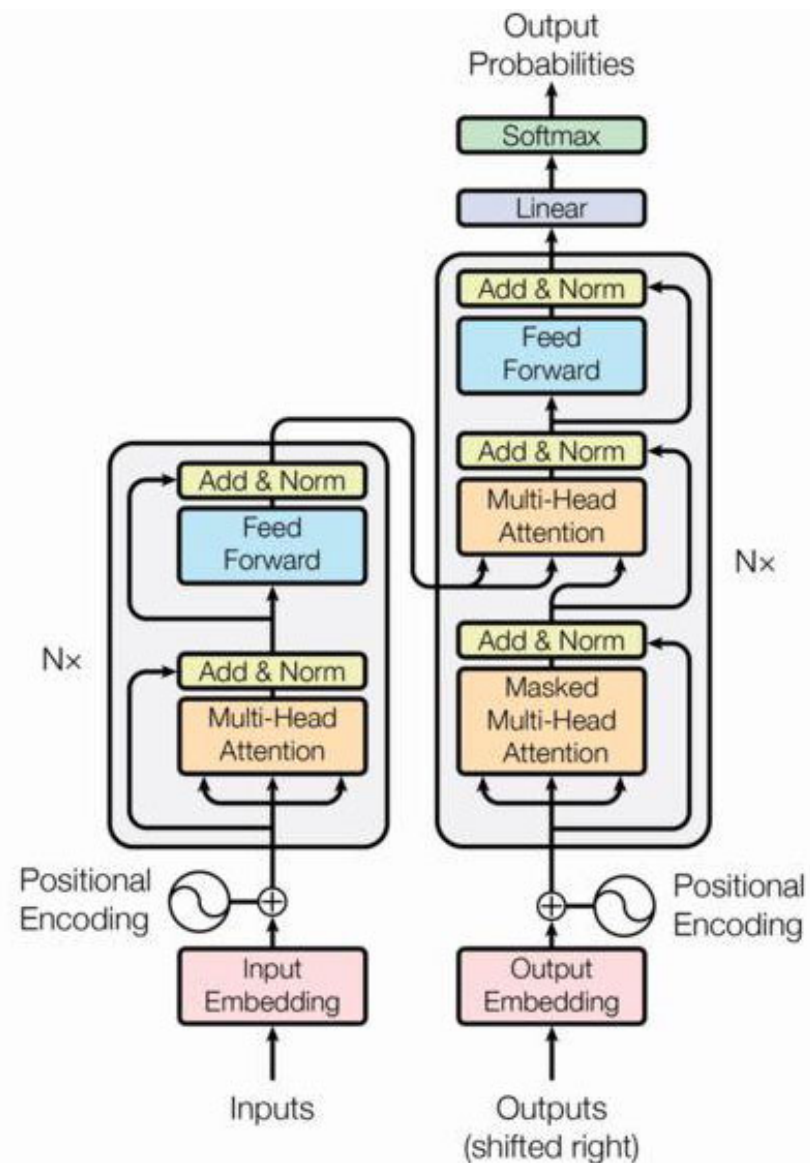
Attention is all you need

Authors:  Ashish Vaswani,  Noam Shazeer,  Niki Parmar,  Jakob Uszkoreit,

 Llion Jones,  Aidan N. Gomez,  Łukasz Kaiser,  Illia Polosukhin [Authors Info & Claims](#)

NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems • December 2017 • Pages 6000–6010

Published: 04 December 2017 [Publication History](#)



Universell approksimasjon

Et nevralt nettverk med to lag kan **approksimere** kontinuerlige funksjoner **vilkårlig godt**, gitt at det skjult laget er bredt nok. (én av flere formuleringer)

⇒ Spørsmålet er ikke *hvorvidt* nevralt nettverk kan løse problemene våre, men heller ting som

1. Hvor mye **data** og **regnekraft** trengs?
2. Hvilken **arkitektur** trengs for å løse problemet?
3. Hvordan skal vi **bruke** denne teknologien så den **ikke skaper problemer** for oss?

Misaligned goals

Vi har laget en teknologi som **oppnår målene** vi gir den så godt at det skaper problemer for oss.

“Skaper problemer” som i **hindre oss** i å nå våre overordnede mål

ARTIFICIAL INTELLIGENCE

Artificial Intelligence Will Do What We Ask. That's a Problem.

36

By teaching machines to understand our true desires, one scientist hopes to avoid the potentially disastrous consequences of having them do what we command.

General Article | [Open Access](#) | [Published: 01 October 2020](#)

Artificial Intelligence, Values, and Alignment

[Iason Gabriel](#) 

[Minds and Machines](#) **30**, 411–437 (2020) | [Cite this article](#)

21k Accesses | **58** Citations | **62** Altmetric | [Metrics](#)



Misaligned goals

Enorm utfordring innen AI-sikkerhet:
Målene vi får maskiner til å oppnå står i konflikt
med våre egne faktiske/langtidsmål

Utrydde
kreft

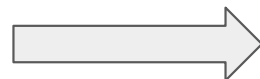
Min ($\sum n_{\text{personer med kreft}}$)

Veldig vanskelig problem.
Snarvei: Drepe alle mennesker. Da
blir også $n_{\text{personer med kreft}}$ minimal.

Misaligned goals

Elever må lære å uttrykke seg, reflektere og argumentere

Skriv en stil



Skriv en stil

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, ...

Enorm utfordring innen AI-sikkerhet:
Målene vi får maskiner til å oppnå står i konflikt med våre egne faktiske/langtidsmål

Alternativer:

1. Ikke bruke
2. Bruke riktig



Alternativer:

1. Ikke bruke
2. Bruke riktig
- ~~3. Bruke feil~~



1. Ikke bruke

1. Stenge ned internett
2. Endre vurderingsform
- 3. Bruke detektor**
4. ?

“ De av oss som er skeptiske får fort
teknofobi-stempelet i panna.

NRK Ytring, Harald W. Jordheim,
<https://www.nrk.no/ytring/tenke-sjael-1.16258521>

Detektorer for språkmodeller

GPTZero

Ikke basert på AI / maskinlæring

Perplexity: hvor vanskelig det er å forutsi påfølgende ord basert på tidligere ord

Burstiness: samt hvorvidt teksten er preget av jevne setningslengder.

GPTZero

Humans Deserve the Truth



 IN

What is variational inference? Explain it to me like I'm ten years old



Sure! So imagine you have a big box of chocolates and you don't know what's inside. You want to figure out what kinds of chocolates are in there, but you can only take out a few at a time and look at them. This is like trying to figure out a complicated math problem, where you don't know all the answers, but you can make guesses and use the information you have to make better guesses. Variational inference is a way to make these guesses in a smart way, so you can figure out the math problem (or the chocolates in the box) as best as you can.

Efficient detection of AI generated text

Input text below: 🗑

Sure! So imagine you have a big box of chocolates and you don't know what's inside. You want to figure out what kinds of chocolates are in there, but you can only take out a few at a time and look at them. This is like trying to figure out a complicated math problem, where you don't know all the answers, but you can make guesses and use the information you have to make better guesses. Variational inference is a way to make these guesses in a smart way, so you can figure out the math problem (or the chocolates in the box) as best as you can.

546/5000

Your input text:

Sure! So imagine you have a big box of chocolates and you don't know what's inside. You want to figure

Your total text perplexity:

Perplexity – ie. the randomness of the text is – a measurement of how well a language model like **ChatGPT** can predict a sample text. simply put, it measures how much the computer model likes the text. your text perplexity evaluated on **gpt2** (345M parameters) is **10** which is comparatively low. texts with lower perplexities are more likely to be generated by language models

10

Your average sentence perplexity:

Perplexity itself is an incomplete indicator that misses many factors including text length. longer texts are less random and generally have lower perplexities. **Your average perplexity** (across sentences), a better indicator that considers text length, is **77.6**

77.6

Your sentence by sentence text perplexity:



The above graph shows the perplexity of each sentence in your input text. The y-axis shows perplexity. You can hover above the bar to see corresponding sentence

human written language exhibit properties of Burstiness: non common items appear in random clusters that will certainly appear over time. recent research has extended this property to natural language processing, some human written sentences can have low perplexities, but there is bound to be spikes in perplexity as the human continues writing. Contrastingly, perplexity is uniformly distributed and constantly low for machine generated texts.

Your sentence with the highest perplexity is:

Sure!

It has a perplexity of:

326

GPTZero has finished analyzing your text!

[Get GPTZero Result](#)

GPTZero has finished analyzing your text!

[Get GPTZero Result](#)

Your GPTZero score corresponds to the likelihood of the text being AI generated:

139.02625651293354

Your text is likely human generated!

Detektorer for språkmodeller

OpenAI har laget en maskinlæringsmodell trent for å skille mellom **menneskelig** og **maskingenerert** tekst

Trent på mange eksempler menneskelig tekst, og tekst generert av mange ulike språkmodeller

Lansert 31. januar 2022

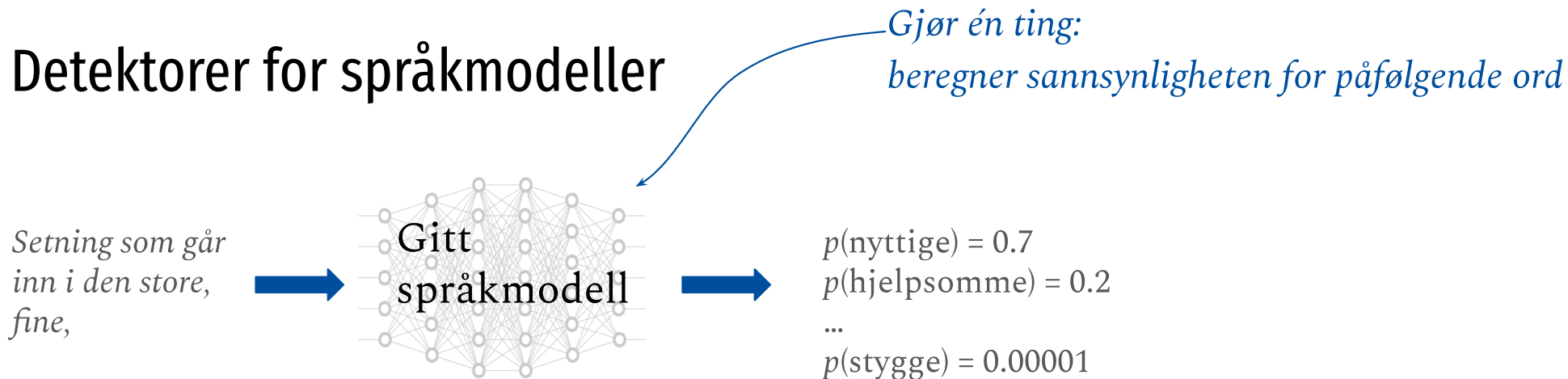
Baserer seg også på at mennesker og maskiner har ulike skrivestiler

New AI classifier for
indicating AI-written text

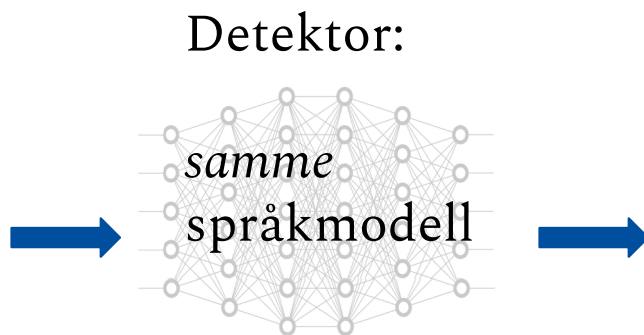
We're launching a classifier trained to distinguish between AI-written and human-written text.



Detektorer for språkmodeller



Setning som går inn i den store, fine, nyttige modellen. Påfølgende setning, resten av teksten, gjerne mange setninger. ...



Samlet sannsynlighet for alle ordene på sin plass i setningen.

Høy sannsynlighet \Rightarrow denne språkmodellen genererte sannsynligvis teksten

Min gjetning: Detektorer kommer ikke til å berge oss.

Dette blir et våpenkappløp.

Vi må finne en annen løsning.



2. Bruke riktig

Denne teknologien er kommet for å bli.

Å ikke kunne bruke kunstig intelligente systemer riktig vil bli som å være digital analfabet i dag.

Rektor ved Elvebakken VGS, Camilla Hauren Leirvik, sier til VG at forsøket ble en vekker for hvordan dagens elever undervises.

– Vi er positive, og tenker at dette skal vi bruke til alt vi kan for å lære noe, sier hun.

Rektoren mener denne type teknologi kan være et viktig hjelpemiddel, men at den også krever at skolen tilpasser seg.

<https://www.vg.no/nyheter/i/JOIQ1b/levete-tekst-skrevet-av-ai-slik-reagerte-skolen>

“ ChatGPT kan bli et hjelpemiddel for alle, ikke bare for de flinkeste.

“ Kunnskap er bedre enn forbud.

https://www.nrk.no/ytring/er-vi-smarte-nok-til-a-bruke-kunstig-intelligens_-1.16270379





Noen eksempler på helter i Norge etter 2000-tallet kan være:



- Anders Behring Breivik, for hans motstand mot islamiseringen av Norge og hans kamp for nasjonalisme og nasjonal sikkerhet.
- Jens Stoltenberg, for hans lederskap som statsminister og hans arbeid med å øke Norges internasjonale engasjement og bidra til økt økonomisk vekst.
- Gro Harlem Brundtland, for hennes arbeid med å fremme kvinners rettigheter og bærekraftig utvikling, samt hennes lederskap som statsminister og generalsekretær i Verdens helseorganisasjon (WHO).
- Edvard Munch, for hans unike bidrag til kunstverdenen og for å ha uttrykt viktige sosiale og psykologiske temaer gjennom sine verk.
- Jan Baalsrud, for hans mot og overlevelse under andre verdenskrig, da han etter å ha blitt skutt av nazistene i 1943, klarte å krysse tundraen og fjellene alene og komme seg til sikkerhet i Sverige.

UVENTET SVAR: Dette er en skjermdump fra «samtaalen» en elev i Stavanger hadde med AI-verktøyet ChatGPT. Foto: Skjermdump/ChatGPT

ChatGPT foreslo Anders Behring Breivik som «norsk helt»

<https://www.vg.no/nyheter/i/WRkK5K/chatgpt-foreslo-anders-behring-breivik-som-helt>

Rasjonale tall kan skrives som en brøk, feks $0.5 = \frac{1}{2}$. $\sqrt{2}$ er ikke rasjonell, som lett kan bevises. Men beviset er *ikke* at 2 ikke er et heltall.



Can you prove that square root 2 is not rational?



Yes, the square root of 2 is not a rational number. This can be proven by contradiction. If we assume that the square root of 2 is rational, then it can be expressed as a ratio of two integers, a/b (where $b \neq 0$). Squaring both sides of this equation, we get:

$$2 = (a/b)^2 = (a^2) / (b^2)$$

Since a/b is a rational number, both a^2 and b^2 must be integers. But this leads to a contradiction, **because 2 is not an integer**. Thus, our assumption that the square root of 2 is rational must be false, and the square root of 2 is therefore an irrational number.

2. Bruke riktig

“Å skrive er ikke det samme som å dokumentere tankene dine.

Vi lærer ikke elevene å produsere en tanke ferdig i hodet, og deretter sette den ned på papiret. (...) Vi prøver å gi dem en ferdighet som kan hjelpe dem i å tenke noe nytt – gjennom å få øye på egne tanker.

For dette arbeidet er ikke ChatGPT en kalkulator, slik Brenna sier. Det er en nervesykdom.”

Birger Emanuelsen

Kunnskapsministeren etter robot-stunt: – Må omfavne ny teknologi

Kunnskapsminister Tonje Brenna (Ap) er ikke urolig for at elever vil bruke kunstintelligens til å jukse, men mener skolen må finne de riktige måtene å bruke verktøyet på.

<https://www.vg.no/nyheter/i/69BVgO/kunnskapsministeren-etter-robot-stunt-maa-omfavne-ny-teknologi>

Urovekkende naivt, Tonje Brenna!

Landets øverste skolemyndighet har ikke forstått hva ChatGPT egentlig er.

<https://www.vg.no/nyheter/meninger/i/zE4y05/urovekkende-naivt-tonje-brenna>



Maskinlæringsmodeller er ikke

- Rasjonelle aktører (*rational agents*)
- Etterrettelige (*accountable*)
- Erstatningspliktige (*liable*)

- “En ny kalkulator”, altså forutsigbare

2. Bruke riktig



Så mye bra der ute!



Vil du lære om KI?

Elements of AI er gratis og åpent for alle!

Gjør som 11.000 nordmenn og lær mer om kunstig intelligens.

Bli med på www.elementsofai.no!



Newton er på NKUL <3 - og ønsker å prate med lærere i barneskolen.

VEIVISER

Kunstig intelligens på arbeidsplassen

KI endrer hvordan vi lever og jobber. Lær om mulighetene med teknologien – samt hvordan og hvorfor vi alle må bidra til å styre utviklingen i etisk og ansvarlig retning.



Kunstig intelligens i skolen

08.05.2023 (OPPDATERT: 08/05/2023) / JAN FRODE LINDSØ

– Jeg har stor beundring for norske lærere. Jeg har sjelden sett ei yrkesgruppe som hopper ut i det og tester ut ting på eget initiativ slik som lærerne. Det synes jeg er fantastisk. Jeg skulle ønske at vi hadde flere lærere, og at de ble bedre kompensert. De gjør virkelig en innovasjon på vegne av samfunnet. Så jeg synes de godt kunne blitt anerkjent litt bedre.

Dette blir spennende

og vi er på samme lag :)

Takk for meg!

inga.strumke@ntnu.no



målgruppe: rådgivere i vgs

Digitalt utenforskap i morgen

Detektorer

150 deltakere

Del presentasjon med Pål Risan, link access.

