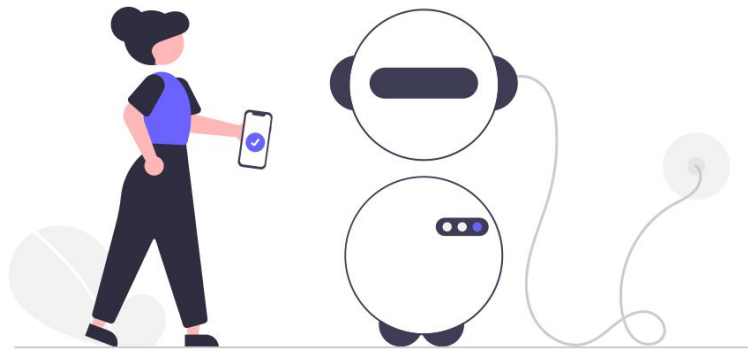


Vikten av att verkligen förstå AI

@Almedalsveckan2021

Hampus Londögård
londogard.com



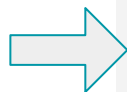


AIGCG

AI Global Competence Group



Vi på Almedalen



5/7

13:00

- Vikten av att verkligen förstå AI // **Hampus Londögård**

6/7

11:30

- How can AI help shape future cities and green industries in post-pandemic era? // **Prajit Datta**

13:00

- Deepfakes, kan vi lita på det vi ser? // **Maria Erman**





AIGCG

AI Global Competence Group



Vem är jag?

- Civilingenjör Datateknik (LTH)
- Anställd på AFRY
 - Ute hos en kund som är "Fortune 100"
 - Big Data / Machine Learning / Scala
 - Kompetensansvarig på AFRY IT South (1.5 år)
 - För AI senaste 2.5 år
 - Är delansvarig i AFRYs *AI Global Competence Group* där vi når ut till alla 20 000 anställda
- Driver londogard.com med kul produkter/bloggar/öppen källkod

Whoa!



population on mars



Allt



Bilder



Nyheter



Kartor



Videor



Fler

Inställningar

Verktyg

Ungefär 241 000 000 resultat (0,64 sekunder)

three billion people

Mars's population currently houses three billion people, and is led by the **Martian** Congressional Republic. **Mars** has the most advanced military in the solar system. The MCR headquarters is located at Olympia. The Belt refers to the asteroid belt and the outer planets, home to 72 million people.

static1.squarespace.com › TheExpanse_FinalEdits-7 PDF

[The Expanse - Squarespace](#)

Agenda

01

Problematik

02

Arbetsflöde

03

**Kritiska
Punkter**

04

Verktyg

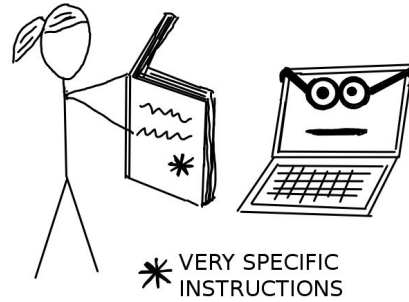
0

Intro

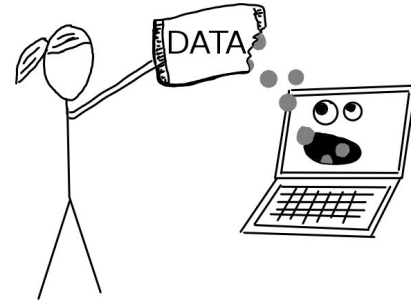
0. Intro

Vad är Machine Learning?

Without Machine Learning

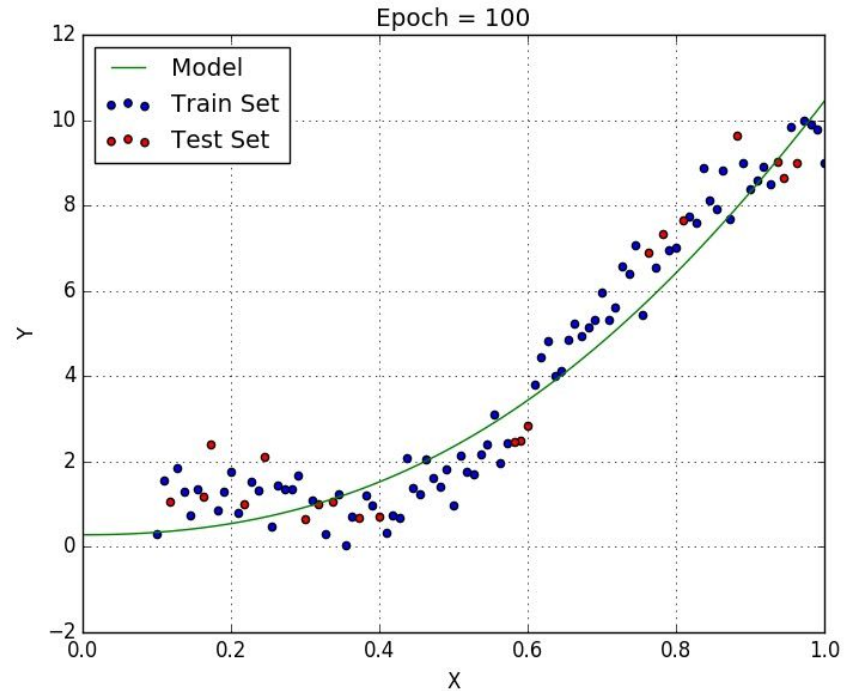


With Machine Learning



0. Intro

Vad är Machine Learning?



1. Problematik

Vikten av att verkligen förstå AI

Jämlik/Rättvis

Fairness

Förståelse/Förklarande

Interpretability / Explainability

Lite exempel

...där kommer fler 🤪

Al Car... Ruins Soccer Game For Fans After Mistaking
Refe... For Ball

69.7
SHARES

Facebook Enabled Advertisers to Reach 'Jew Haters'

After being contacted by ProPublica, Facebook removed several anti-Semitic ad categories and promised to improve monitoring.

by Julia Angwin, Madeleine Varner and Ariana Tobin, Sept. 14, 2017, 4 p.m. EDT

GOOGLE TECH ARTIFICIAL INTELLIGENCE

le 'fixed' its racist algorithm
gorillas from its

51

When...
for porn detec...

rogue and ora...
stab myself in the head.

Turkish - det

o bir aşçı
o bir mühe
o bir dokt
o bir hemşire

Danni Morritt, from Doncaster, said voice-command
Echo Dot told her beating of heart 'is not a good
thing' when she was learning about the cardiac cycle
she is a nurs

... life w
story.

Vad som är
skrämmande
men inte
förvånande

VB



The Machine

Making sense of AI

**65% of execs can't
explain how their AI
models make
decisions, survey finds**

Kyle Wiggers

@Kyle_L_Wiggers

May 25, 2021 10:03 AM

Förståelse / Förklaring

Interpretability

En features påverkan i resultatet

Explainability

Interna modellens representation & påverkan



2. Ett **vanligt** arbetsflöde

* Förenklat

Ett **vanligt** arbetsflöde*



* Förenklat

Data

Data är **grundstenen** i all
AI

Data är där vi **hittar**
mönster och gör **smarta**
beslut

Data är ungefär **90 % av**
lösningen



Den mörka sidan

Data kan vara **obalanserad**

Data kan vara **biased** (partisk/fördomsfull)

Data kan vara **dålig**

Varg eller Husky?



- > 90%
- Datalikheter?

"Why Should I Trust You?": Explaining the Predictions of Any Classifier, Ribeiro et. al

Träning & Validering

AI-modellen tränas till att **hitta mönster** i data

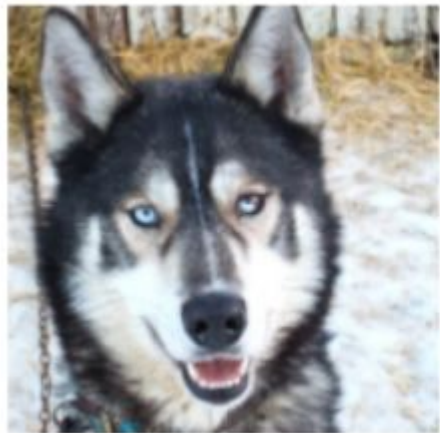
Validering ser till att modellen **presterar väl**



Den mörka sidan

Vi kan **validera fel saker**
Modellen kan hitta **felaktiga mönster**

Varg eller Husky?



(a) Husky classified as wolf



(b) Explanation

Varg eller Husky

Modelförståelse
(Interpretability)
Modelrättvisa
(Fairness)

Exempel:
Visualisera vikter
(explanation)

"Why Should I Trust You?": Explaining the Predictions of Any Classifier, Ribeiro et. al

Lansering

Kunderna skall äntligen
få **förbättra/förenkla**
sina liv



Den mörka sidan

Beslut som är **farliga kan fattas**
Miljön ute i världen kan skilja sig
Världen förändras med tiden

Varg eller Husky?



Varg eller Husky

Mänsklig slutfaktor
Tight feedback-loop
Kontinuerlig träning

"Why Should I Trust You?": Explaining the Predictions of Any Classifier, Ribeiro et. al

Rättvisepolit

Allokering / Fördelning

Försämrar/Förbättrar möjlighet för några grupper att

- Få resurser
- Få möjligheter
- Få information

Exempel:

Alexa Mistakenly Offers Porn

When the child in the video tells Alexa to "play 'Di for porn detected...hot chick amateur girl sexy...' (f



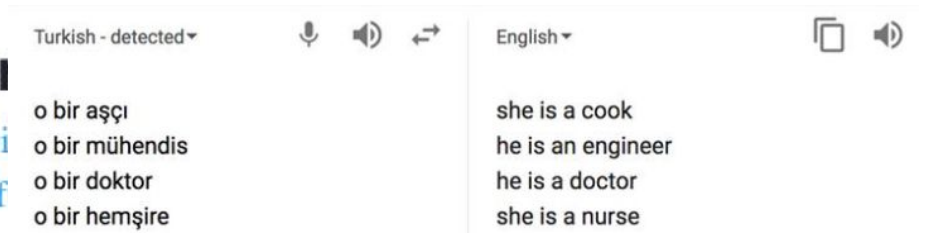
REUTERS

RETAIL

OCTOBER 11, 2018 / 1:04 AM / UPDATED 3 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

person.



Turkish - detected ▼

- o bir aşçı
- o bir mühendis
- o bir doktor
- o bir hemşire

English ▼

- she is a cook
- he is an engineer
- he is a doctor
- she is a nurse

An aerial photograph of a city, likely Copenhagen, showing a dense urban landscape with various building styles, including historic multi-story houses and modern structures. A large white rectangular box is superimposed over the center of the image, containing the text 'Fler eksempel på samhällspåverkan'.

Fler exempel på samhällspåverkan

Exempel på samhällspå

Fail: IBM's "Watson for Oncology" Cancelled After \$62 million and Unsafe Treatment Recommendations

pga historisk data

- Apple Cards Sexist

How A US Citizen Was Wrongly Arrested Due To A Flawed Facial Recognition Match

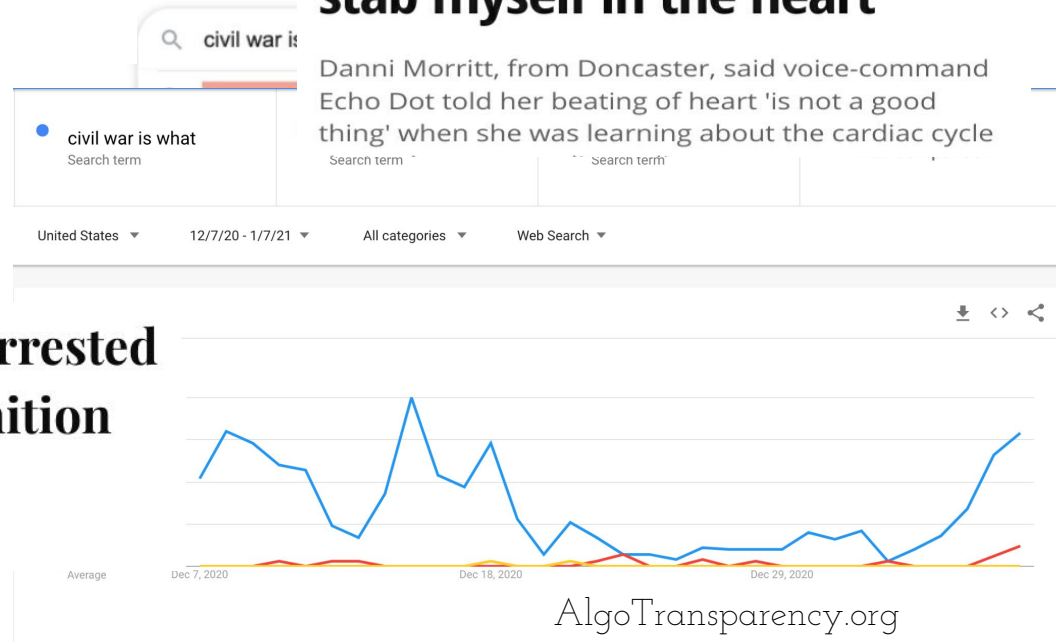
26/06/2020

Here's what Google
January 6th before
incognito from N



"My Amazon Alexa went rogue and ordered me to stab myself in the heart"

Danni Morritt, from Doncaster, said voice-command Echo Dot told her beating of heart 'is not a good thing' when she was learning about the cardiac cycle





4. Verktvg

Förstå AI:ns livscykel

- I vilken miljö lanserar vi?
- I vilken miljö kan vi hamna i?
- Vem kan komma att använda AI:n?
- Hur håller vi koll **kontinuerligt**?
 - *Data Drift*
- Lansera inte 0-100

Rättvis AI (Fairness)

- Har vår data underliggande **bias**?
- Har vi **rättvis allokering**?
- Har vi **rättvis servicekvalitet**?
- Kan vi sänka vår prestanda för att öka den för bredden?

✓ fairlearn

Förklarande AI (Explainability & Interpretation)

- Kan vi **förklara** vår AI:s beslut?
- Kan vi **förstå** vår AI:s beslut?
- Låter vi våra **användare förstå**?

✓ interpretml

Validera Korrekt

- Validerar vi vad som verkligen betyder något?
- Uppdaterar vi modellen mot rätt mål (förlust, en: *loss*)
- Validerar vi rätt miljö?

**Inte att
glömma**

Sök

**Personlig
Assistent**

Foto

... & mer

Tack!

Frågor?

hampus.londogard@gmail.com
+46 733 673 179



THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POL
PILE OF LINEAI
THE ANSWERS

ANOTHER HUGE STUDY
FOUND NO EVIDENCE THAT
CELL PHONES CAUSE CANCER.
WHAT WAS THE W.H.O. THINKING?

I THINK THEY JUST
GOT IT BACKWARD.



**We're using
AI instead
of biased humans**

imgflip.com



**What did you
train the AI on?**

RE NOT...THERE ARE SO
NY PROBLEMS WITH THAT.

JUST TO BE SAFE, UNTIL
I SEE MORE DATA I'M
GOING TO ASSUME CANCER
CAUSES CELL PHONES.



**What did you
train the AI on?**

TOM
FISH
BURN

.com