

# Data Science, Artificial Intelligence, Machine Learning, and all that...



UNIVERSITY OF  
GOTHENBURG

**CHALMERS**

**Richard Johansson**

`richard.johansson@gu.se`

Lunchseminarium för datastudenter, 6 december 2018

# översikt

introduktion

utbildning inom ML, DS, AI

forskning och exjobb

vad är dessa buzzwords?

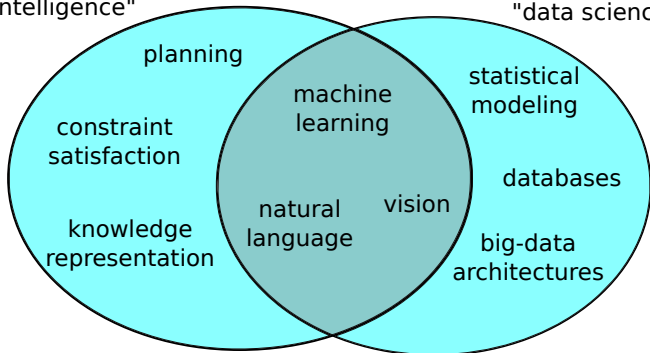


[källa]



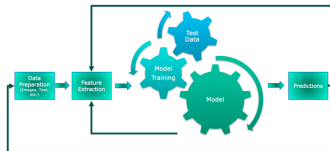
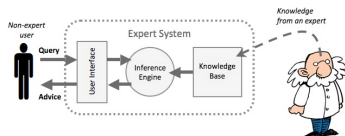
"artificial intelligence"

"data science"



# kunskapsdrivet eller datadrivet?

- ▶ **kunskapsdrivna system**: definiera exakt hur systemet funkar
- ▶ **datadrivna system**: specificera en modell → “träna” modellen



[källa]

# vad är målet med AI?

- ▶ **stark AI**: simulera den kognitiva processen
- ▶ **svag AI**: lösa ett problem som “kräver intelligens”



Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

John McCarthy,  
Dartmouth Workshop 1956

## Under the bonnet

How a self-driving car works

Signals from **GPS (global positioning system)** satellites are combined with readings from tachometers, altimeters and gyroscopes to provide more accurate positioning than is possible with GPS alone

**Lidar (light detection and ranging)** sensors bounce pulses of light off the surroundings. These are analysed to identify lane markings and the edges of roads

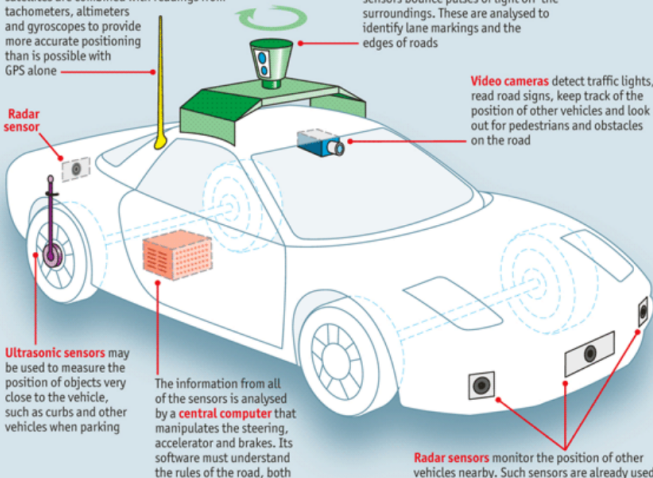
**Video cameras** detect traffic lights, read road signs, keep track of the position of other vehicles and look out for pedestrians and obstacles on the road

**Radar sensor**

**Ultrasonic sensors** may be used to measure the position of objects very close to the vehicle, such as curbs and other vehicles when parking

The information from all of the sensors is analysed by a **central computer** that manipulates the steering, accelerator and brakes. Its software must understand the rules of the road, both

**Radar sensors** monitor the position of other vehicles nearby. Such sensors are already used





## Riksdagens öppna data

Genom öppna data kan alla som vill fritt använda innehållet i riksdagens databaser. Utvecklare, journalister, forskare och andra intresserade kan bygga egna tjänster och ta fram statistik. Öppna data är ett viktigt verktyg för att ge insyn i riksdagens arbete och beslut.



### Dokument

I dokumenten ingår bland annat riksdagsbeslut, regeringens propositioner, ledamöternas motioner och protokoll från kammaren. Vissa dokumenttyper finns från 1961.



### Ledamöter

Information om alla ledamöter som sitter eller har suttit i riksdagen finns från omkring 1990. Det finns uppgifter om ledamoten, olika uppdrag och vad ledamoten sagt och gjort.



### Voteringar

När kammarens ledamöter är oense i en fråga kan det bli omröstning, votering. Uppgifter om voteringarna finns från riksmötet 1993/94 och framåt.

### Röstresultat perioden 1993/94–2001/02

Det finns nu ytterligare nio år av omröstningar tillgängliga som dataset. I dataseten kan man se hur varje enskild ledamot har röstat i olika ärenden.

▶ [Voteringar](#)

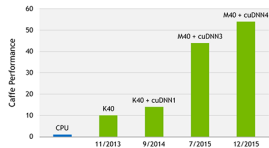
### Diariet i öppna data

Riksdagens diarium finns som dataset i öppna data. Där finns information om ärenden i diariet från den 12 september 2017.

▶ [Dokument](#)



## 50X BOOST IN DEEP LEARNING IN 3 YEARS

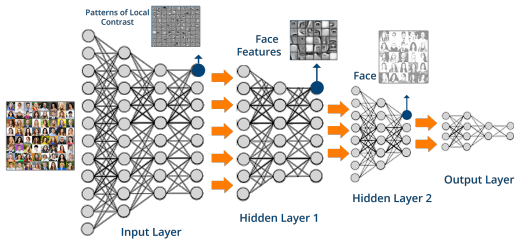


Alzavet training throughput based on 20 iterations.  
CPU: 1x E5-2680v3 12 Core 2.5GHz, 128GB System Memory, Ubuntu 14.04

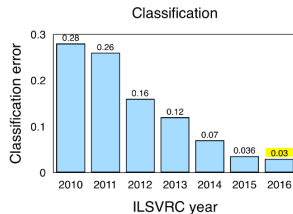
[källa]



# success stories: bildigenkänning

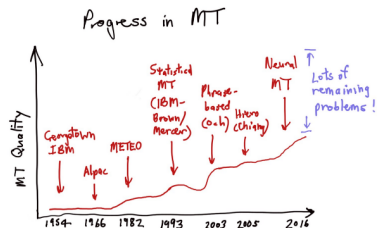
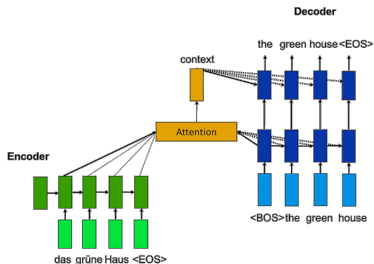
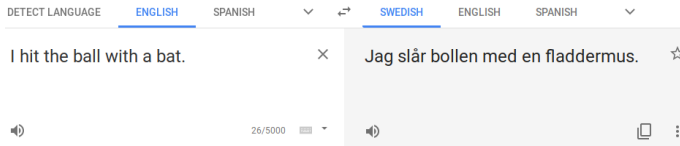


[källa]



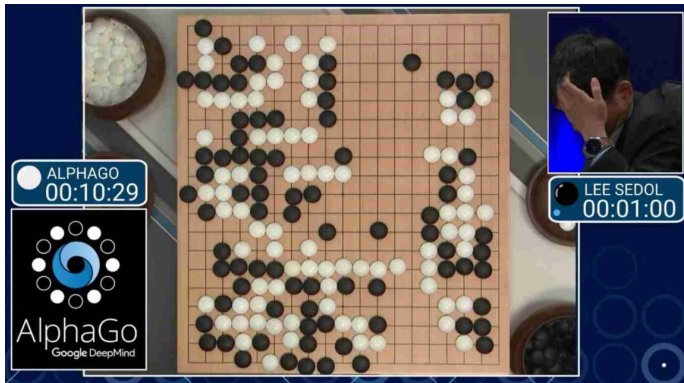
[källa]

# success stories: maskinöversättning



[image by [Chris Manning](#)]

# success stories: spel

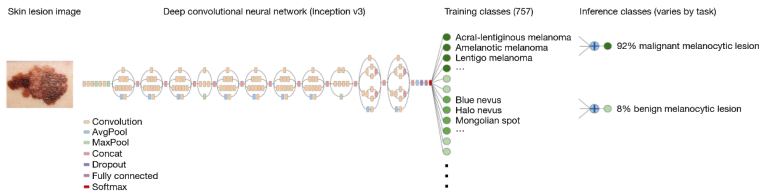


# tillämpningar: fordonsindustrin



[källa]

# tillämpningar: medicin

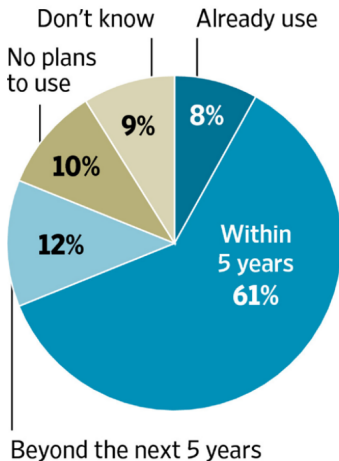


[källa]

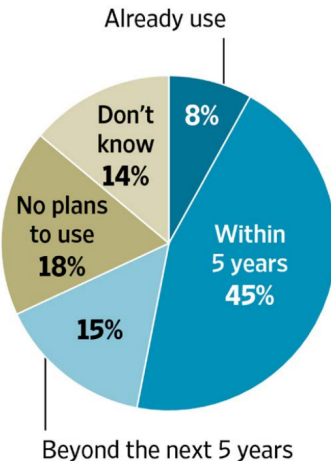


# framtidjobb?

## BUSINESS ANALYTICS



## MACHINE LEARNING



# sidospår: etiska utmaningar inom AI/ML

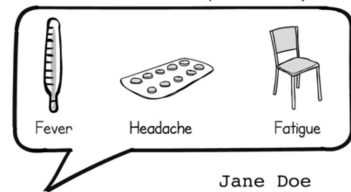
## Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

[källa] [data]

# tolkningsbarhet?

It's easier to trust a prediction if you understand the reasons for it.



Ok, I can trust you.

Jane Doe  
has the flu



[källa]



# stereotyper?

DETECT LANGUAGE   **ESTONIAN**   ▾   ↔   DANISH   SWEDISH   **ENGLISH**

---

Tema on insener.	×	He is an engineer.
Tema on arst.		He is a doctor.
Tema on meditsiiniõde.		She is a nurse.
Tema on lapsehoidja.		She is a nanny.

# översikt

introduktion

utbildning inom ML, DS, AI

forskning och exjobb

# nytt Chalmers-masterprogram i data science från 2019!

## Data Science

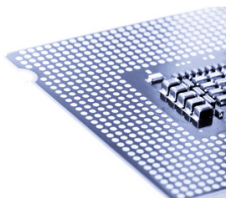
120 credits (MSc, 2 years )

With the digital revolution, Data Science and Artificial Intelligence (AI) has become an important part of our lives and in society as a whole. Vast amounts of data are available in science, industry and other organizations, ranging from molecular biology to social media, from quantum physics to business and retail. In addition, the quickly emerging technologies for processing large-scale data and machine learning is creating a wealth of opportunities, in analyzing such data for understanding and insight, and for building AI systems for new complex tasks, where automated decision making is becoming a reality. Skilled data scientists and AI engineers are in high demand everywhere. With a solid foundation in machine learning, and in understanding the diverse problems of Data Science and AI, you will have a wide range of career opportunities to choose from.



### What is Data Science and AI?

Data Science is a highly cross-disciplinary field concerned with how to extract useful knowledge from data, for deeper understanding and decision support. It is based on a blend of methods in statistics and machine learning, together with computational techniques and algorithms for handling large-scale data. Examples of application areas include biology and other sciences, healthcare, business, finance and different kinds of internet data. Computational methods range from algorithms for collecting and handling large-scale data, statistical methods such as Bayesian modelling, to machine learning techniques such as deep neural networks.





## Applied Data Science, Master's Programme

### About the programme

Big Data is taking centre stage in all areas – business and industry, public policy, the life sciences, natural sciences, humanities and social sciences, and people with knowledge of how to process and analyse large amounts of data has become an increasingly sought after category.

# enstaka kurser inom ML, DS, AI (CSE)

- ▶ databaser
- ▶ introduktion till AI
- ▶ beräkningsmetoder i bioinformatik
- ▶ tillämpad maskininlärning
- ▶ algoritmer för maskininlärning och inferens
- ▶ tekniker för storskaliga data
- ▶ (kommande) avancerad maskininlärning
- ▶ (kommande) maskininlärning för textbehandling



# översikt

introduktion

utbildning inom ML, DS, AI

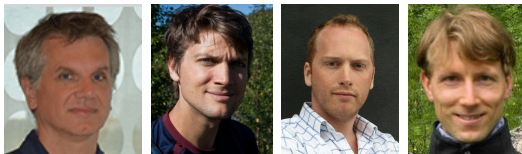
forskning och exjobb

# CSEs forskning inom DS/ML/AI

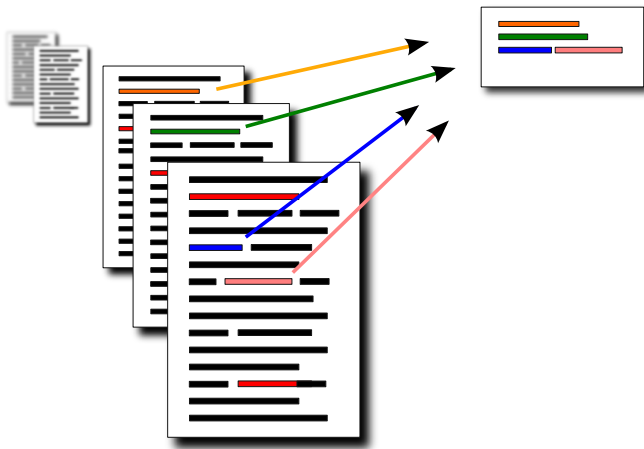
- ▶ **grundforskning:** nya metoder inom maskininlärning



- ▶ **tillämpningar:** biologi, medicin, fordon, text



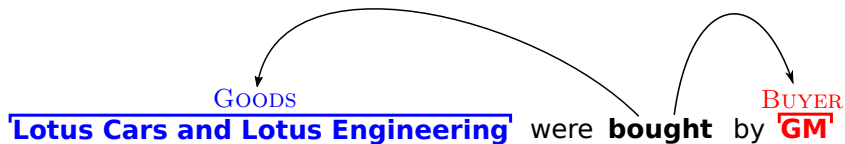
# exempel: textsammanfattning



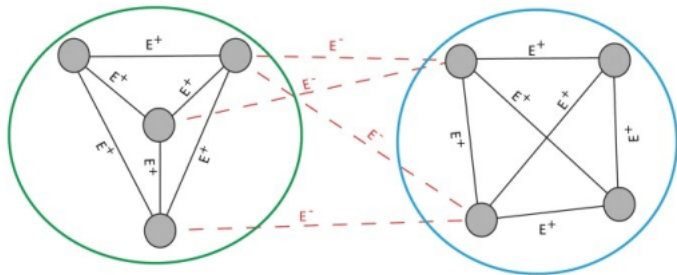
[källa]



## exempel: informationsextraktion



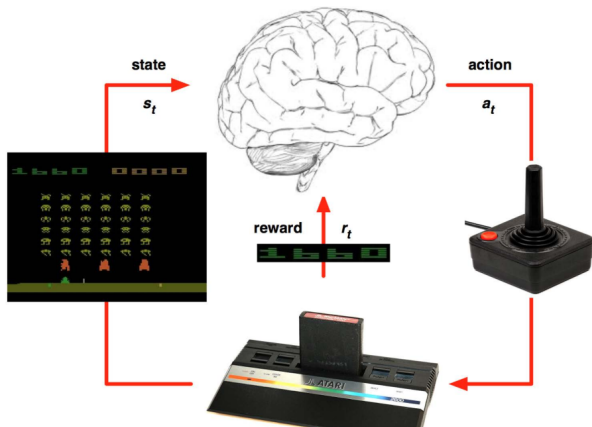
# exempel: korrelationsklustring



[källa]



# exempel: reinforcement learning



# industriexjobb: litet axplock

- Constructing a Context-aware Recommender System with Web Sessions (3Bits Consulting AB)
- Machine Learning for On-line Advertising Using Contextual Information (Admeta)
- The Identification of Target Proteins from Patents - Mining of biological entities from a full-text patent database (AstraZeneca)
- Browser Fingerprinting (Burt)
- Learning to rank, a supervised approach for ranking of documents (Findwise)
- Entity Disambiguation in Anonymized Graphs Using Graph Kernels (Recorded Future)
- Using Classification Algorithms for Smart Suggestions in Accounting Systems (SpeedLedger)
- Cluster User Music Sessions (Spotify)
- Extracting Data from NoSQL Databases - A Step towards Interactive Visual Analysis of NoSQL Data (TIBCO Software)
- Pattern Recognition in a Distributed Message Passing System (Volvo Technology)

tack för uppmärksamheten!

kontakt: [richard.johansson@gu.se](mailto:richard.johansson@gu.se)



- ▶ allmän AI: Russell & Norvig *Artificial Intelligence: A Modern Approach*
- ▶ allmän data science: Skiena *The Data Science Design Manual*
- ▶ maskininlärning introduktion: Daumé III *A Course in Machine Learning*
- ▶ maskininlärning, mer teoretiskt: Hastie, Tibshirani, Friedman *The Elements of Statistical Learning*
- ▶ neurala modeller, “deep learning”: Goodfellow, Bengio, Courville *Deep Learning*
- ▶ neurala modeller för text: Goldberg *Neural Network Methods for Natural Language Processing*

# länkar

- ▶ [scikit-learn](#), lättanvänt Python-baserat bibliotek för många typer av maskininlärning
- ▶ [Weka](#), lättanvänt Java-baserat bibliotek för många typer av maskininlärning
- ▶ [Keras](#), relativt lättanvänt bibliotek för neurala modeller (“legobitar”)
- ▶ [PyTorch](#) och [TensorFlow](#) är Facebooks respektive Googles bibliotek för implementation av neurala modeller