

# Machine Learning for Networking

## ML4N

Luca Vassio  
Gabriele Ciravegna  
Zhihao Wang  
Tailai Song

# Machine Learning (ML)

- “Design methodologies to extract patterns from data, ideally without much domain-specific expertise”

Mathematics for Machine Learning book

- “Algorithms that enables computers to learn and make decisions without being explicitly programmed”

ChatGPT

- Learn to answer questions based on data
- Fit models to data to make predictions

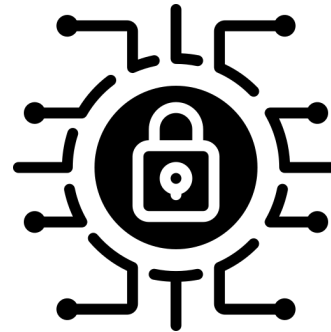
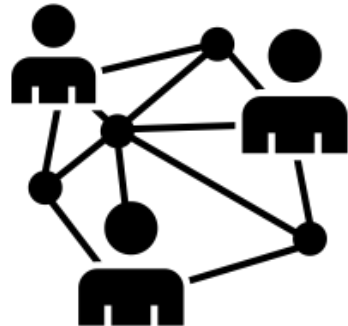
# Networking

- **Communication networks**

Nodes interconnected by telecommunications links used to exchange messages between the nodes

- **Cybersecurity applications**

Identification/protection of computer systems and networks from attacks



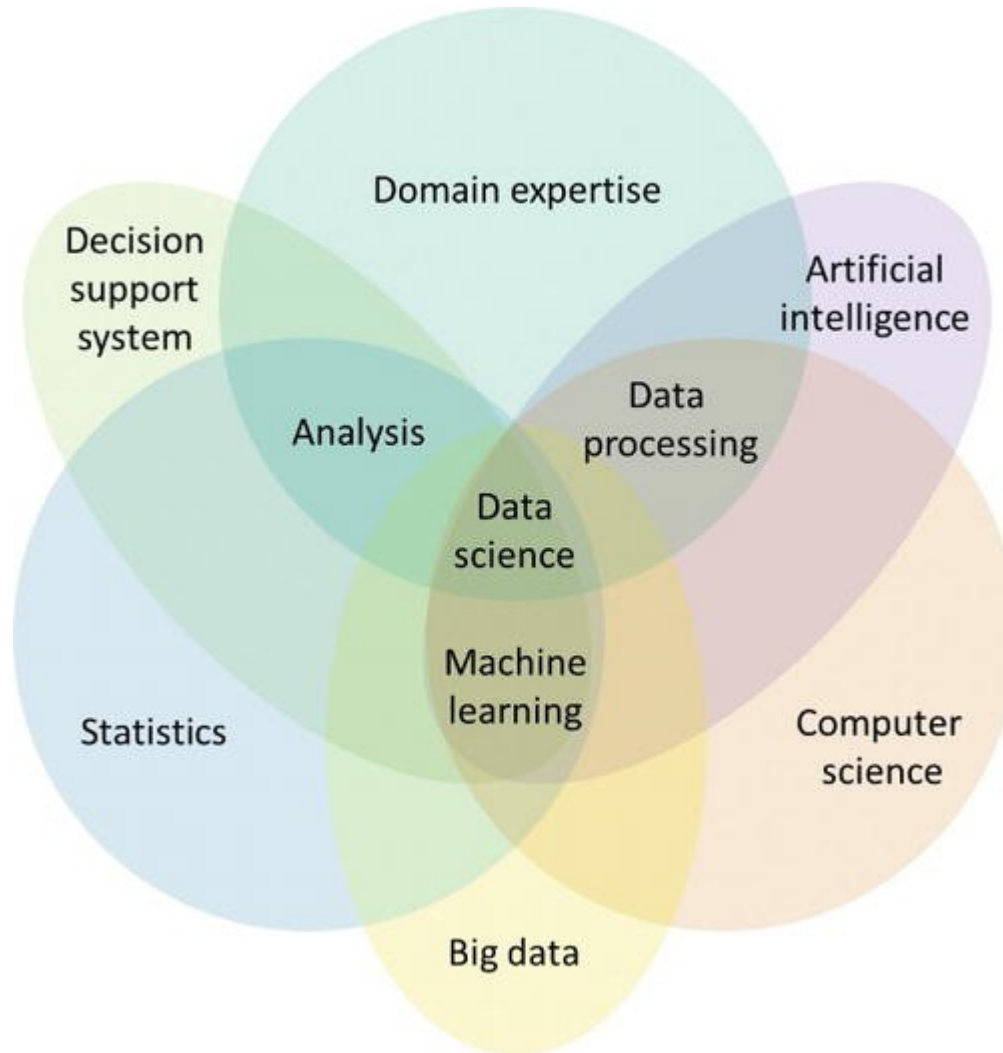
# ML4N course



- A Python course
- A computer science course
- A machine learning course
- A communication network course
- A modelling and statistic course
- A performance evaluation course
- A team-work course
- ...

Big data  
Data science  
Data mining  
Artificial intelligence  
Machine learning  
Pattern recognition  
Statistical learning  
Deep learning

...



# Big data

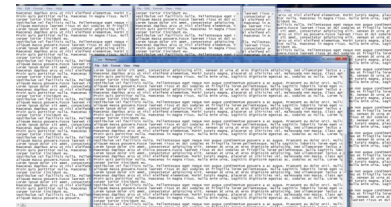
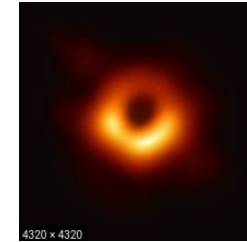
# Big data

- “Data whose scale, diversity and complexity require new architectures, techniques, algorithms and analytics to manage it and extract value and hidden knowledge from it”



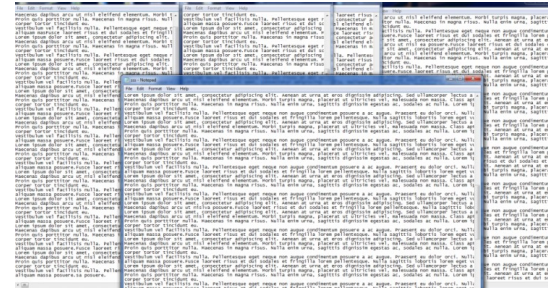
# Who generates big data?

- User-generated content on the Internet
- Health and scientific computing
- Log files
  - Web server log files, machine system log files
- Internet Of Things (IoT)
  - Sensor networks, RFID, smart meters



# Who generates big data?

- Log files
  - Web server log files, machine syslog files
- Internet Of Things
  - Sensor networks, RFID, smart meters



# 5 Vs of big data

- **Volume** – quantity of data
- **Velocity** – generation rate (streaming data)
- **Variety** – various formats and structures (audio, video, image, text, graph,...)
- **Veracity** – control data quality
- **Value** – decision-making capabilities

# Answers to big data challenges

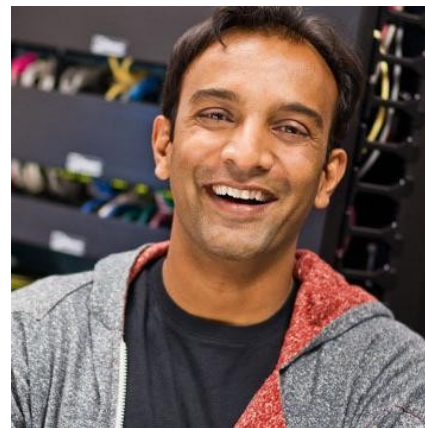
- Technology & infrastructure
  - New architectures, programming paradigms and techniques
  - Transfer the processing power to the data
  - Apache Hadoop/Spark ecosystem
- Data management & analysis
  - New emphasis on “data” → Data science



# Data Science

# Data Science

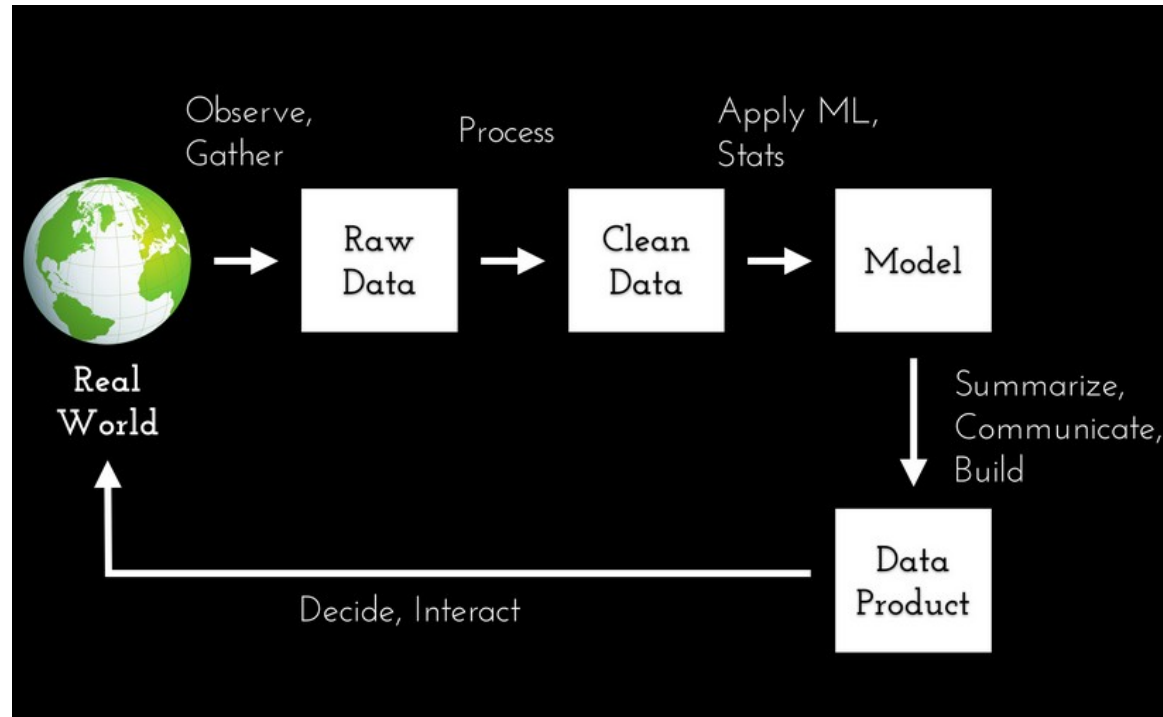
- “Extracting meaning from large quantities of data”



Dhanurjay Patil  
popularized the  
word *data scientist*

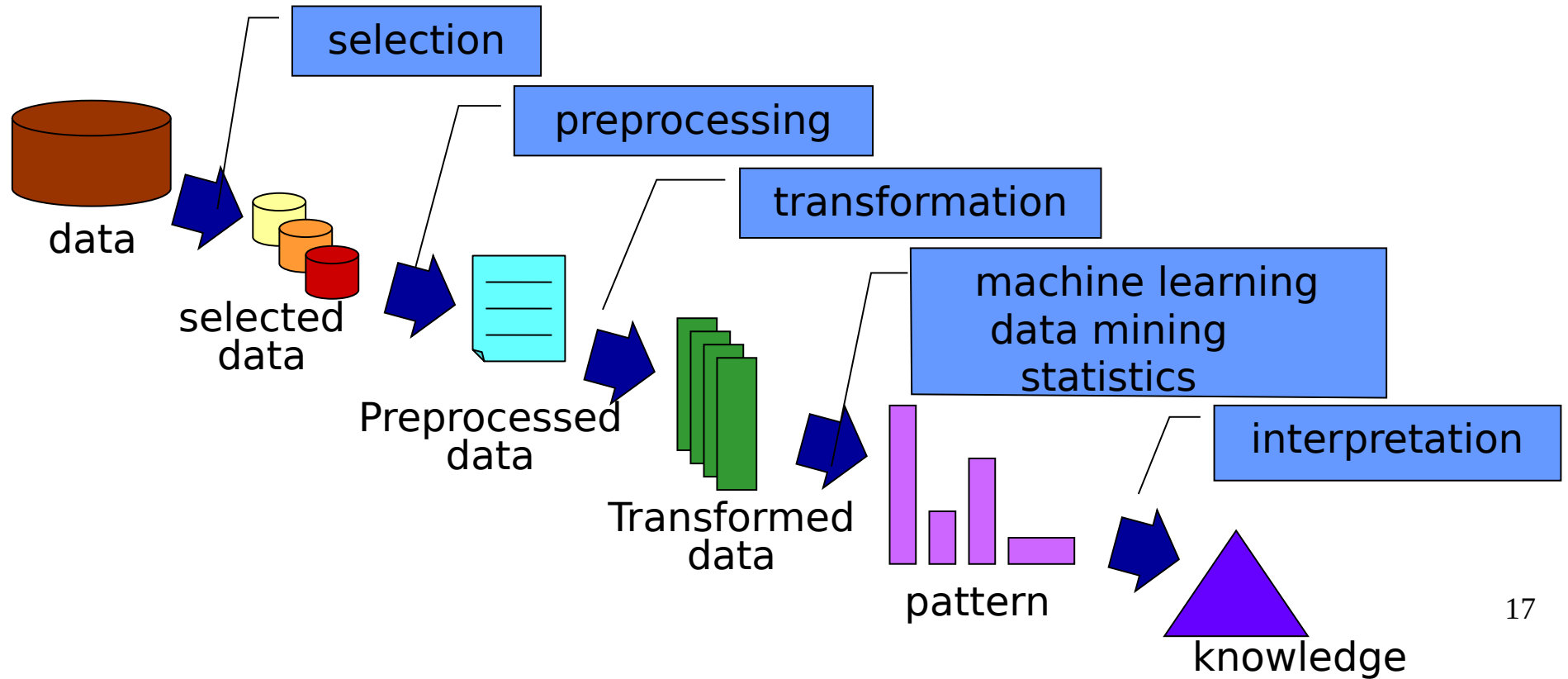
# The data science process

## - one possible pipeline



# The data science process

- ~KDD: Knowledge Discovery in Databases

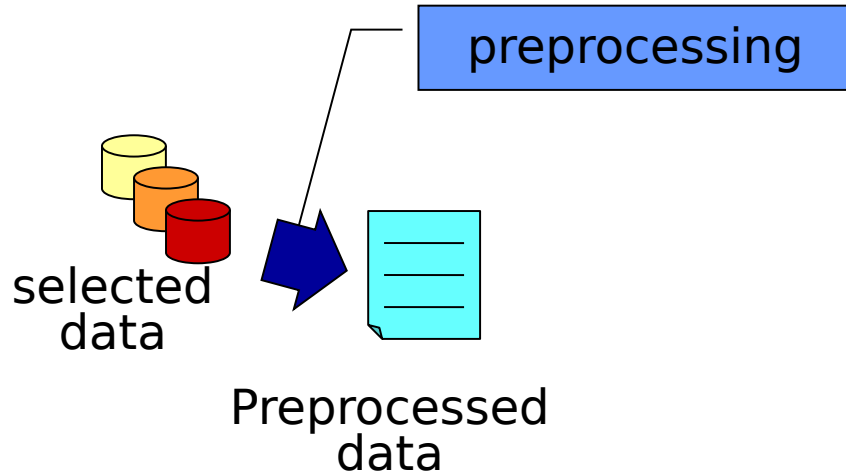




# Data storage

- Storage infrastructure
  - Storage technology, e.g., HDD, SSD
  - Networking architecture, e.g., DAS, NAS, SAN
- Data management
  - File systems (HDFS, Ceph), key-value stores (Memcached), column-oriented databases (Cassandra), document databases (MongoDB)
- Programming models
  - Map reduce, stream processing, graph processing

# Preprocessing



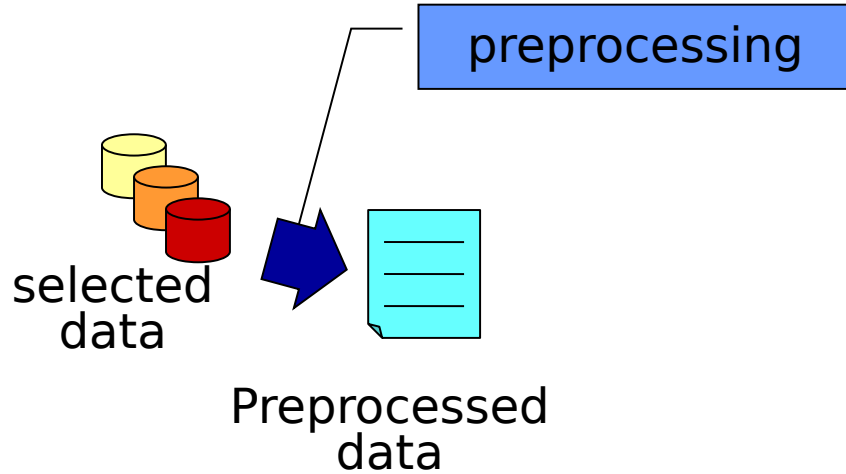
## Data cleaning

- ✂ reduces the effect of noise
- ✂ identifies or removes outliers
- ✂ solves inconsistencies

## Data integration

- ✂ reconciles data extracted from different sources
- ✂ integrates metadata
- ✂ manages redundancy

# Preprocessing



## Data cleaning

- ✂ reduces the effect of noise
- ✂ identifies or removes outliers
- ✂ solves inconsistencies

## Data integration

- ✂ reconciles data extracted

Real world data is “dirty”

Without good quality data, no good quality pattern

- ✂ manages redundancy

# Data analysis

- Objectives
  - Descriptive analytics, predictive analytics, prescriptive analytics
- Methods
  - Machine learning, data mining, statistical analysis
    - Text mining, graph data mining, association analysis, classification and regression, clustering,...
- Diverse domains call for customized techniques

# Machine learning and data mining common goal

- Extraction of **information** from **available** data
  - implicit
  - previously unknown
  - potentially useful
  - automatic (performed by algorithms)
- Extracted information is represented by means of **abstract models (pattern)**
- ML: use the **models** to make **predictions**

# The data science recipe

- Data expert
  - Data processing, data structures
- Data analyst
  - Machine learning, Data mining, statistics
- Visualization expert
  - Visual art design, storytelling skills
- Domain expert
  - Provide understanding of the application domain
- Business expert
  - Data driven decisions, new business models

# MODERN DATA SCIENTIST

## MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



## PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

## COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

# Artificial Intelligence



# Artificial Intelligence (AI)

- “AI is machines mimicking human intelligence to perform tasks”

ChatGPT

- “The goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and support for robotics.”

Wikipedia

# “Artificial Intelligence is the new Electricity!”



Prof. Andrew Ng during a talk at Stanford University

# Electricity fuels our planet



# AI fuels:

- Self-driving cars





# AI fuels:

- Real-time translation



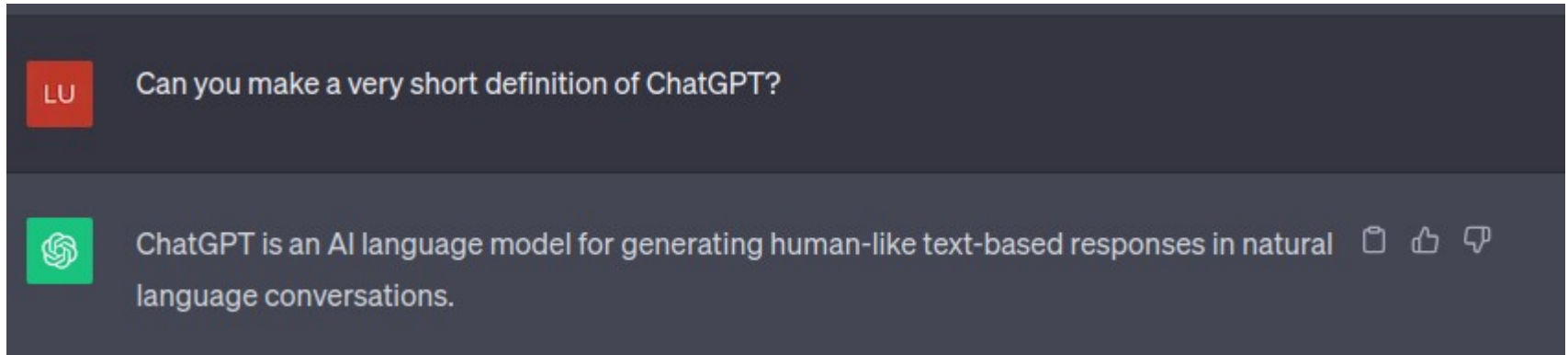
# AI fuels:

- Writing programming code

```
1
2 def common_prefix(a, b) :
3     """Return the common prefix of two lists."""
4     if len(a) < len(b) :
5         return common_prefix(b,a)
6     for i in range(len(a)) :
7         if a[i] != b[i] :
8             return a[:i]
9     return a
10
11 |
12
```

# AI fuels:

- Personal assistants



# AI fuels:

- Finding a job

## More jobs for you

Based on your profile and search history



### Lehtori metsätalous

Lapland University of Applied Sciences  
Rovaniemi, Lappi, Finland (Hybrid)

Promoted · 10 applicants



### Adjunkt/lektor til uddannelse inden for IT - systemudvikling

UCL Erhvervsakademi og Professionshøjskole  
Odense, South Denmark, Denmark (On-site)



Your profile matches this job

Promoted · 7 applicants



### Senior Lecturer, ICT, Robotics

Häme University of Applied Sciences, HAMK  
Riihimäki, Tavastia Proper, Finland (On-site)

Promoted · 20 applicants · Easy Apply



### Associate Professor of Ecological Macroeconomics

Roskilde University  
Roskilde, Zealand, Denmark (On-site)



1 school alum works here

Promoted · 2 applicants



### Assistant Professor in Physics with a focus on Experimental Quantum Technology

The Faculty of Engineering at Lund University  
Lund, Skåne County, Sweden (On-site)



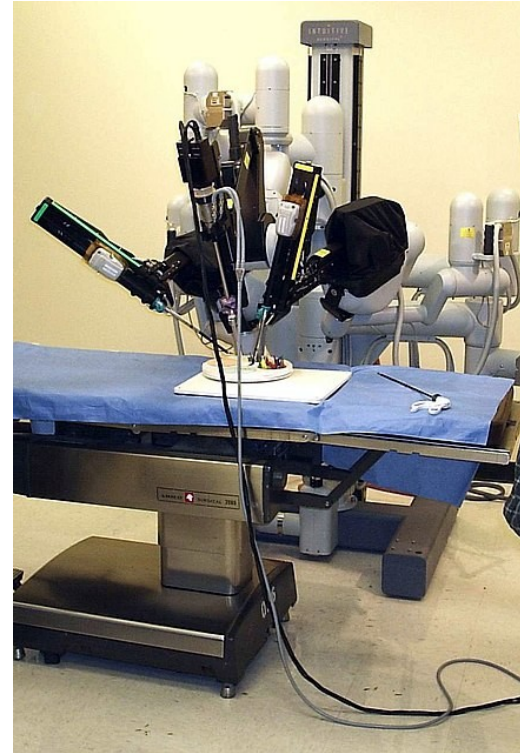
# AI fuels:

- Detect fake news



# AI fuels:

- Automated surgery:



A patient-side surgical arm of **Da Vinci Surgical System**

# AI fuels:

- ...

[https://en.wikipedia.org/wiki/Applications\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Applications_of_artificial_intelligence)

Intelligence involves learning

AI involves ML

# Machine Learning

# Machine Learning (ML)

- “Design methodologies to extract patterns from data, ideally without much domain-specific expertise”

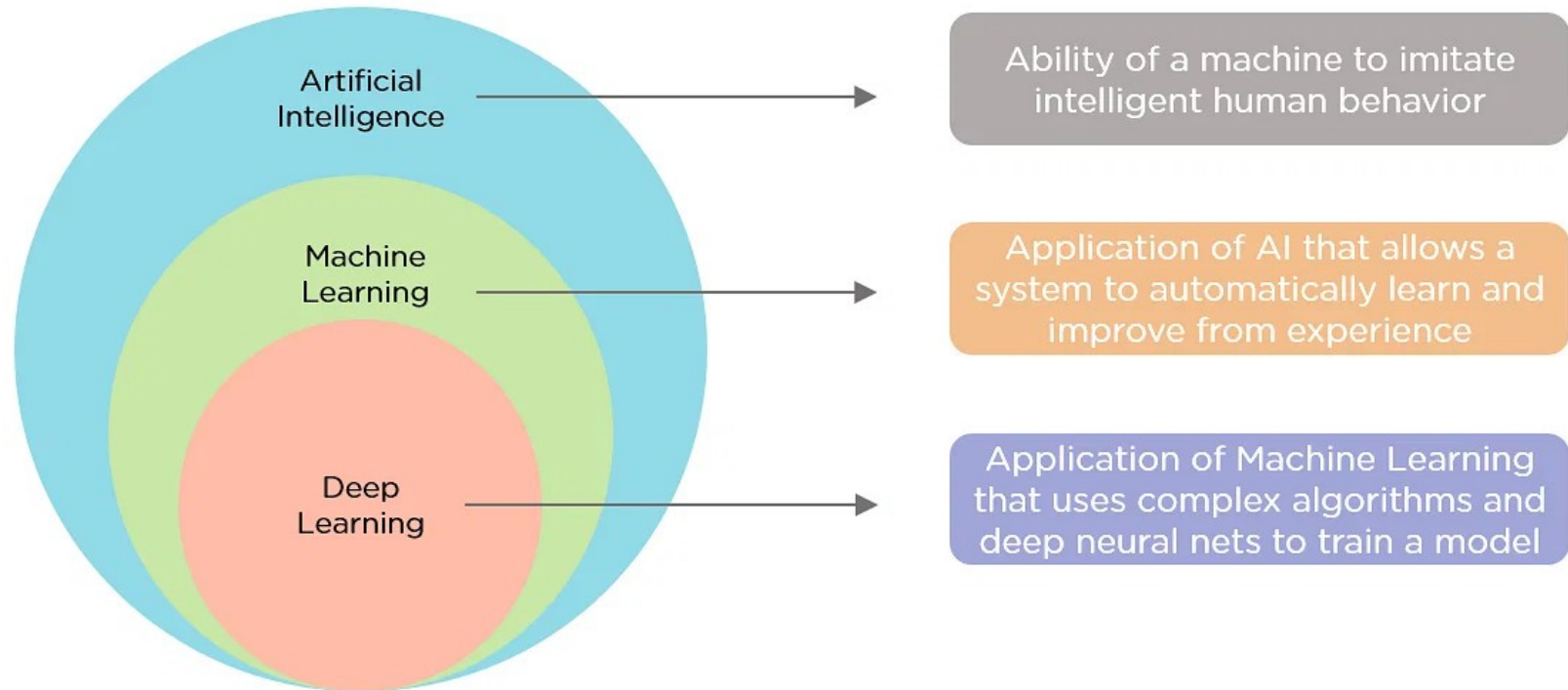
Mathematics for Machine Learning book

- “Algorithms that enables computers to learn and make decisions without being explicitly programmed”

ChatGPT

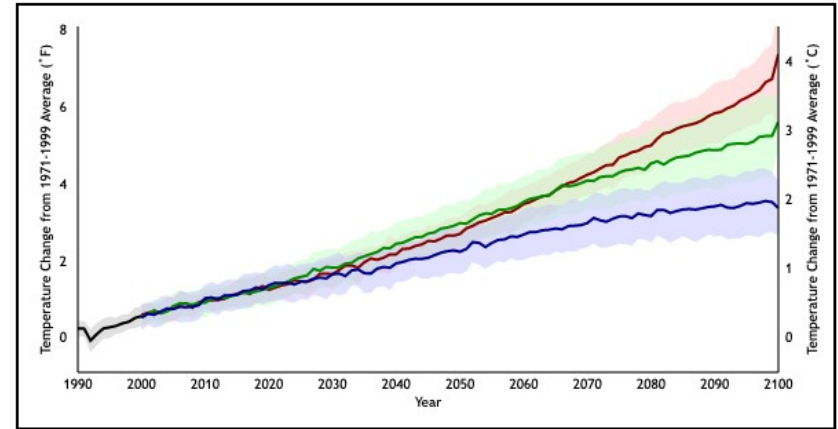
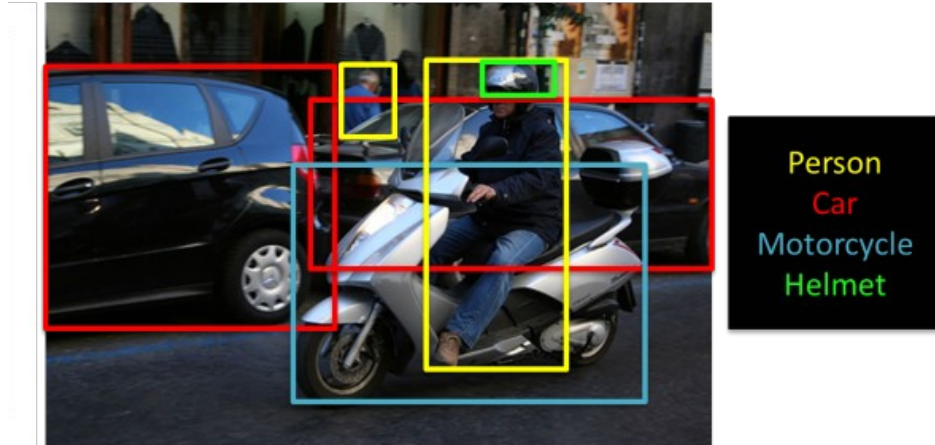
- Learn to answer questions based on data
- Fit models to data to make predictions

# Machine Learning and AI



# Supervised learning: Classification/Regression

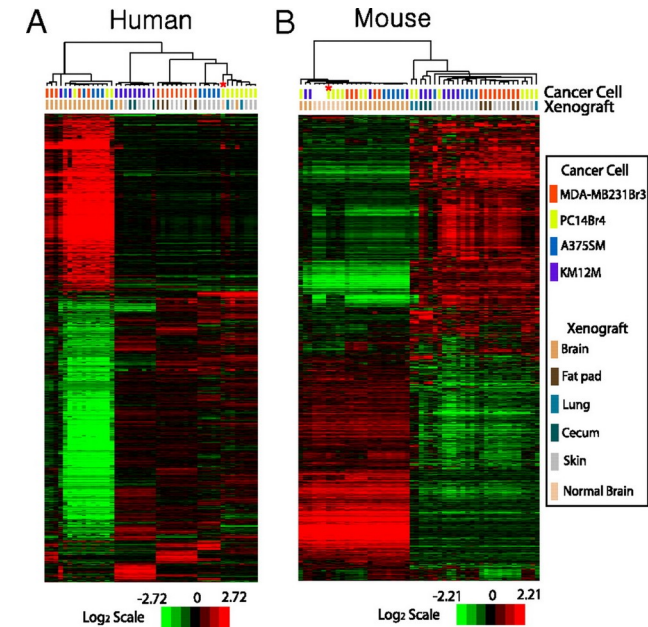
- Predict classes or values
- ~pattern recognition





# Unsupervised learning: Clustering

- Detecting groups of similar data objects
- Identifying exceptions and outliers

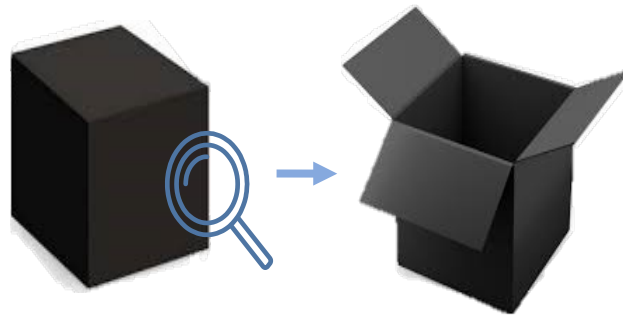


# Open issues

Social impact such as:

- Interpretability and transparency of the process

*The ability to explain or to present in understandable terms to a human*



Open the black box

# Open issues

Social impact such as:

- Bias in algorithms and data

## *Wrongfully Accused by an Algorithm*

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

<https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>



# Open issues

Social impact such as:

- Bias in algorithms and data

Algorithm can learn also prejudices and biases from data, leading to unfair outcomes

Watch Videos:

<https://www.netflix.com/title/81328723>

<https://www.codedbias.com/about>

<http://gendershades.org/>



# Open issues

Social impact such as:

- Privacy preservation

Strava released their global heatmap.  
13 trillion GPS points from their users



Security concerns have been raised after a fitness tracking firm showed the exercise routes of military personnel in bases around the world.

ML4N

# Prerequisites of the course

- Programming skills (whatever the language)
- Communication networks
- Probability theory and statistics
- Linear algebra
- Calculus
- Operational research (continuous optimization)

# Goals of ML4N



- You will learn:
  - How to use **Python** for data science and machine learning
  - Different phases of **data science** and **machine learning** process
  - **Theoretical** principles of machine learning
  - Models for **supervised and unsupervised** learning
  - **Properties, domains of application, and limitations** of different machine learning approaches
- You will be able to:
  - Design, implement and evaluate code in **Python**
  - Design, implement and evaluate a **machine learning pipeline**
  - Devise complete solutions for different **tasks in networking**
  - Critically evaluate which **strategies** are better suited



# Other Machine Learning courses in PoliTO

- Machine learning and Deep learning – Barbara Caputo

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TXFNG&p\\_a\\_acc=2024](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TXFNG&p_a_acc=2024)

- Advanced Machine Learning – Tatiana Tommasi

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01URWOV&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01URWOV&p_a_acc=2025)

- Applied AI and machine learning – Gianvito Urgese

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01VIAMY&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01VIAMY&p_a_acc=2025)

- Data science and machine learning for engineering applications – Tania Cerquitelli

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01DSTMW&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01DSTMW&p_a_acc=2025)

- Machine learning and pattern recognition – Sandro Cumani

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01URTOV&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01URTOV&p_a_acc=2025)

# Other Machine Learning courses in PoliTO

- Machine learning for IOT – Daniele Pagliari

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TXPSM&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TXPSM&p_a_acc=2025)

- Machine learning for vision and multimedia – Fabrizio Lamberti

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01URPOV&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01URPOV&p_a_acc=2025)

- Machine learning in applications – Santa Di Cataldo

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01URXOV&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01URXOV&p_a_acc=2025)

- Mathematics in Machine Learning – Francesco Vaccarino

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TXGSM&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TXGSM&p_a_acc=2025)

- Model Order Reduction and Machine Learning – Fabio Vicini

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01DTTNG&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01DTTNG&p_a_acc=2025)

# Other Machine Learning courses in PoliTO

- Optimization for machine learning – Giuseppe Calafiore

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TVOMV&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TVOMV&p_a_acc=2025)

- Data analysis and Artificial Intelligence – Tatiana Tommasi

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01USPLO&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01USPLO&p_a_acc=2025)

- Applied data science project – Giuseppe Rizzo

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TXXSM&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TXXSM&p_a_acc=2025)

- Data science lab: process and methods – Flavio Giobergia

[http://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01TWZSM&p\\_a\\_acc=2025](http://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01TWZSM&p_a_acc=2025)

- Large Language Models – Flavio Giobergia

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01HZNOV&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01HZNOV&p_a_acc=2025)

# Other Machine Learning courses in PoliTO

- Machine Learning for Mathematical Engineering – Carlo Masone

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01OHDNG&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01OHDNG&p_a_acc=2025)

- Robot Learning - Giuseppe Averta

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01HFNOV&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01HFNOV&p_a_acc=2025)

- Machine Learning – Sandro Cumani

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01HERUU&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01HERUU&p_a_acc=2025)

- Statistical learning and neural networks – Enrico Magli

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01SOVBH&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01SOVBH&p_a_acc=2025)

- Methods and tools for ICT – Diego Valsesia

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01DSFBH&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01DSFBH&p_a_acc=2025)

# Other Machine Learning courses in PoliTO

- Signal, image and video processing and learning – Diego Valsesia

[https://didattica.polito.it/pls/portal30/gap.pkg\\_guide.viewGap?p\\_cod\\_ins=01SOVBH&p\\_a\\_acc=2025](https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01SOVBH&p_a_acc=2025)

- ...

# Any questions?



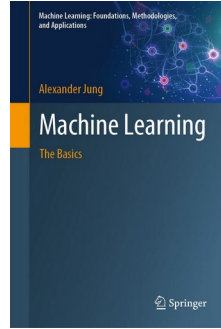
# Self-assessment quiz

- 1) What characterizes “Big Data”?
- 2) Make 5 examples of apps that (might) use Machine Learning and guess for which task
- 3) Write down your own definitions of Statistics, Computer Science, Data Science and Machine Learning

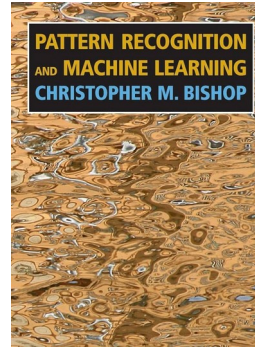
# References: readings



- Chapter 1



- Chapter 2 and Appendix B,C,D





# Slide acknowledgments



- Elena Maria Baralis – Politecnico di Torino
- Alexander Jung and Shamsiiat Abdurakhmanova – Aalto University