SARDIGNA CHIRCAS
SARDEGNA RICERCHE

open:campus

Artificial Intelligence for developers

8 weekend per diventare Machine Learning Specialist

Natural Language Processing

Maurizio Atzori
Università degli Studi di Cagliari
atzori@unica.it

February 9-10 , 2024

Intelligent Data Exploitation

Dipartimento di Matematica ed Informatica

https://web.unica.it/unica/en/intelligent_data.page

Faculty:

Maurizio Atzori - **atzori@unica.it**

Cecilia Di Ruberto

Giovanni Puglisi

Barbara Pes

Andrea Loddo

Manuela Sanguinetti

+ Research Fellows / PhD / students

Topics of the Lab:

- **Artificial Intelligence and Natural Language Processing**
- Knowledge Graphs and Semantic Web
- Biomedical Image Analysis, Precision Agriculture, Image Retrieval
- Computer Vision, Multimedia Forensics
- Data Mining and Machine Learning
- High-dimensional Data Analysis and Feature Selection

Featured on:

Wikimedia Research Newsletter

Google Faculty Research Awards

NewScientist

ZDNet

WIRED IT

KURZWEIL ACCELERATING INTELLIGENCE

COMMUNICATIONS OF THE ACM

IEEE Artificial Intelligence & Knowledge Engineering 2019 (IEEE AIKE 2019)
https://aike2019.unica.it/ *3-5 Giugno 2019 all'Hotel Regina Margherita*

Sistemi Evoluti per Basi di Dati (SEBD 2024)
https://sebd2024.unica.it/ *June 23-26 at Tanka Resort Villasimius*

Outline of the course

- Intro on AI, ML and NLP
- Text Processing
- Words and Corpora
- Lexical similarity
- Language Modeling
- Text Classification
- Semantic similarity
- Knowledge Graphs
- Intro to Large Language Models

Teaching Material: Books

⇨ **Speech and Language Processing by Dan Jurafsky and James H. Martin**

- **Free** online version (draft Feb 2024): https://web.stanford.edu/~jurafsky/slp3/

- Introduction to Natural Language Processing, The MIT Press 2019 by Jacob Eisentein (Georgia Tech, now Google AI)
  - **Free** online version (2018): https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf
- Goldberg: A Primer on Neural Network Models for Natural Language Processing
  - **Free** at http://u.cs.biu.ac.il/~yogo/nnlp.pdf
- Artificial Intelligence: Foundations of Computational Agents, second edition, Cambridge University Press 2017 by David L. Poole and Alan K. Mackworth (University of British Columbia, Vancouver)
  - **Free** online: https://artint.info/2e/html/ArtInt2e.html

Teaching Material: other resources

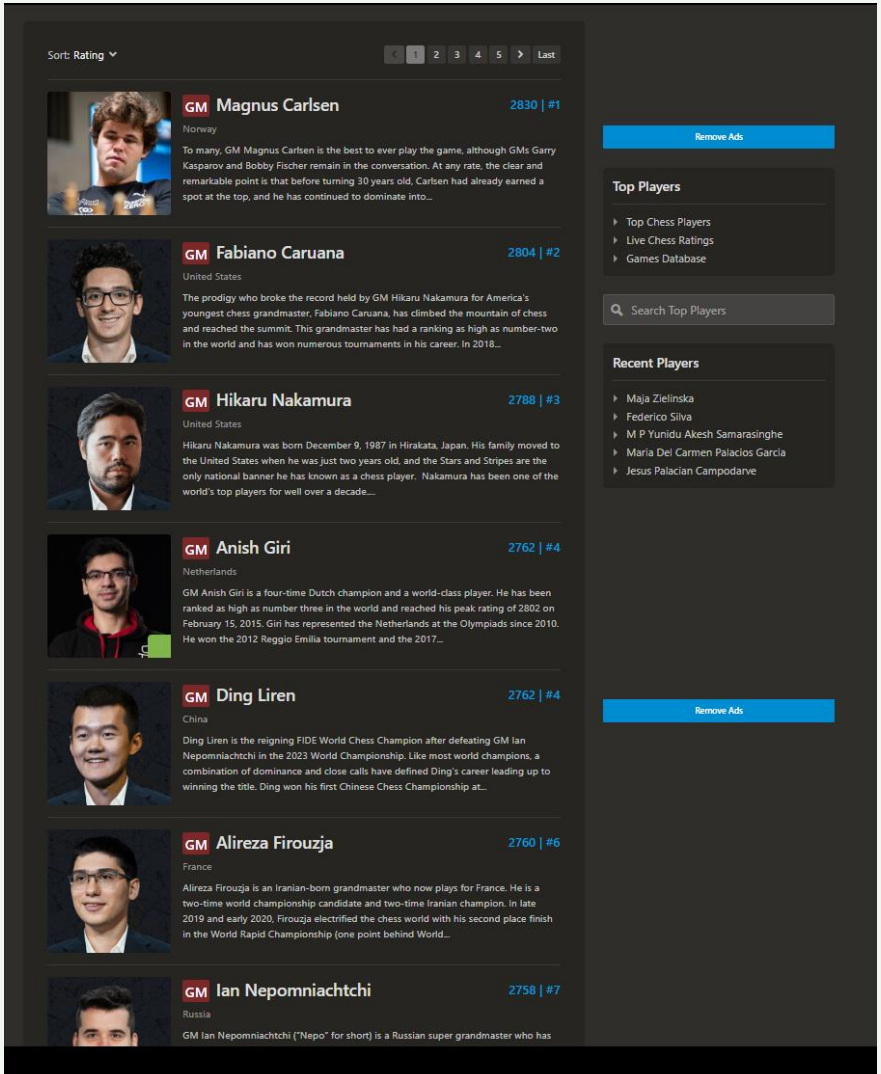
- Free Python libraries:
  - Spacy https://spacy.io/
  - NLTK http://www.nltk.org/book/
  - Sklearn (Scikit-learn), Numpy, Scipy, Pandas
  - PyTorch https://pytorch.org/
  - Textblob
  - Transformers (huggingface)

What are they?

Intro on AI, ML and NLP

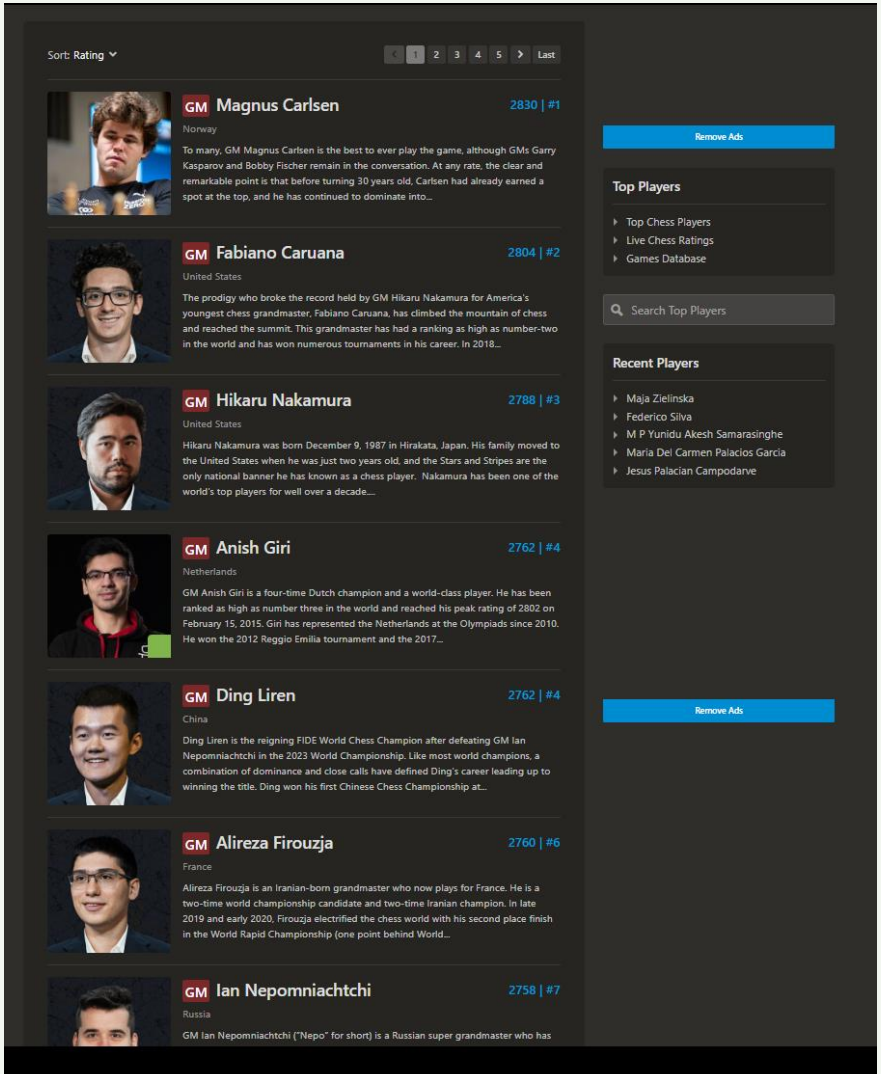
1y

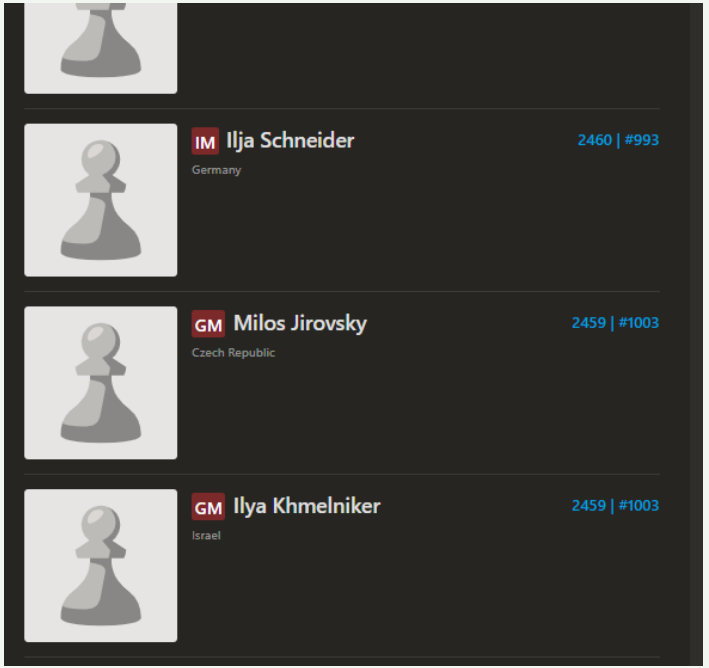
Top Chess players Feb 2024



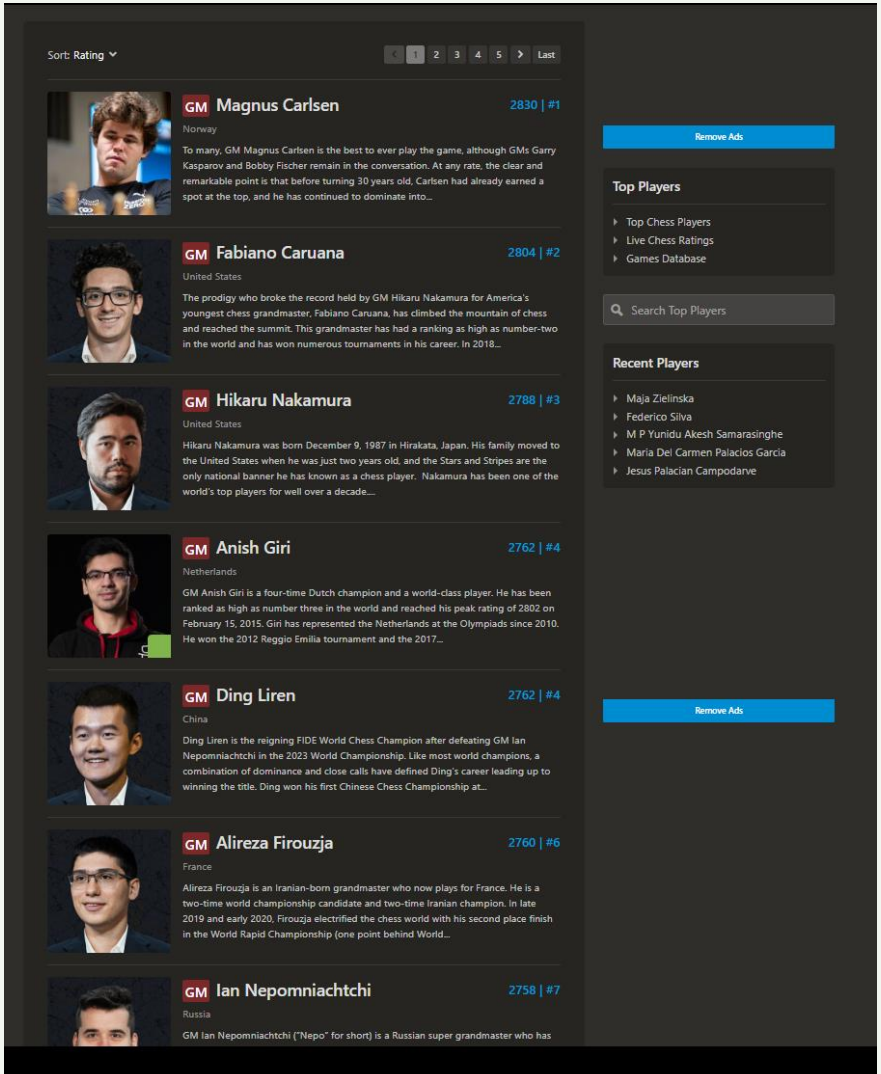
Source: chess.com

Top Chess players Feb 2024





Source: chess.com



As of June 2023, Stockfish is the highest-rated engine according to the computer chess rating list (CCRL), with a rating of approximately 3530

Source: chess.com

Artificial Intelligence: some definitions

It is disputed

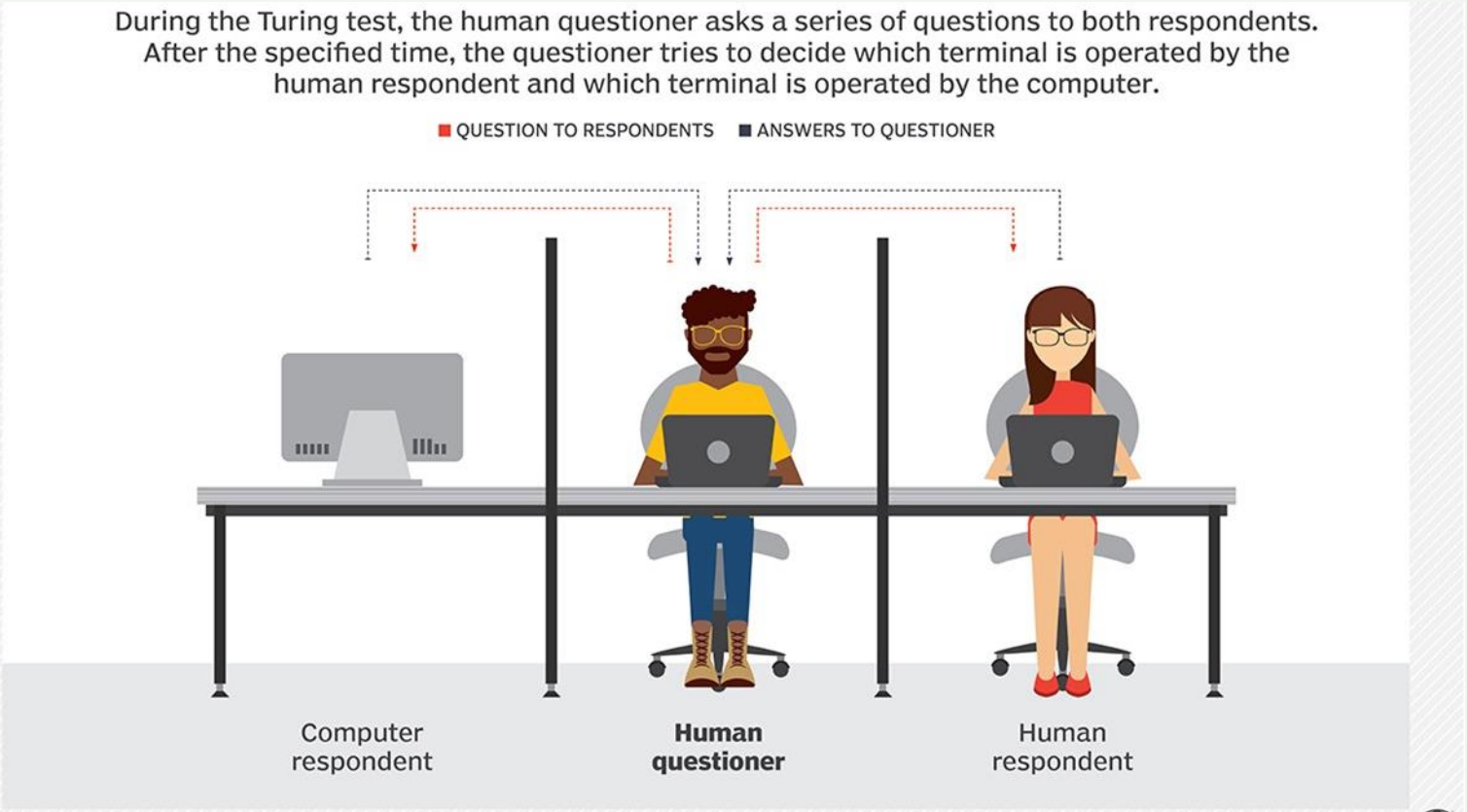
- The goal of artificial intelligence is to build software and robots with the same range of abilities as humans (Russell and Norvig, 2009)
- AI is whatever hasn't been done yet
- Intelligence is the computational part of the ability to achieve goals in some world
- If it can pass some tests (e.g. Turing Test)

AI: the beginning

After WWII (1939-1945), a number of people independently started to work on intelligent machines. The English mathematician **Alan Turing** may have been the first. He gave a lecture on it in 1947. He also may have been the first to decide that AI was best researched by programming computers rather than by building machines. By the late 1950s, there were many researchers on AI, and most of them were basing their work on programming computers.

https://www.kurzweilai.net/what-is-artificial-intelligence

Turing Test (1950)



Other tests for intelligence

- The **Marcus Test-** In which a program which can ‘watch’ a television show is tested by being asked meaningful questions about the show's content.
- **The Lovelace Test 2.0-** Which is a test made to detect AI through examining its ability to create artwork (https://arxiv.org/abs/1410.6142),
  - "Until a machine can originate an idea that it wasn’t designed to, it can’t be considered intelligent in the same way humans are." Ada Lovelace
  - e.g. “tell a story in which a boy falls in love with a girl, aliens abduct the boy, and the girl saves the world with the help of a talking cat.”
- **Winograd Schema Challenge-** a test that asks multiple-choice questions in a specific format

Ada Lovelace

first computer programmer

Winograd schemas and Reasoning

The trophy doesn’t fit into the brown suitcase because **it** is too [small/large].

Q: What is "it" referring to?

Solving this example requires spatial reasoning; other schemas require **reasoning** about actions and their effects, emotions and intentions, and social conventions.

Considerations on Tests for intelligence (1/2)

- Turing Test (as well as others) is a restrictive definition of AI, since it does not admit other **goal-oriented behavior** such as insect behavior as intelligence.

This is why there is now a distinction on different levels of AI:

Artificial General Intelligence (**AGI**, aka **strong AI**, **full AI**) is the intelligence of a machine that can understand or learn *any intellectual task* that a human being can (currently very far from solving this problem)

Artificial Intelligence or **Applied AI** (aka **weak AI**, **narrow AI**) the use of software to study or accomplish specific problem solving or reasoning tasks. It does not attempt to perform the full range of human cognitive abilities (e.g. Siri, self-driving cars, machine translation, ...)

Considerations on Tests for intelligence (2/2)

- These tests on intelligence all rely on a very characteristic ability of human beings: **natural language understanding**

Natural language understanding not only falls into AI topics, but it is also considered **AI-complete**, that is, solving it would solve any other AI task. In other words, it is the most difficult problem in AI

Eliza (1966)

"Psychotherapist" by Joseph Weizembaum

Nice online demo at https://www.masswerk.at/eliza/

http://psych.fullerton.edu/mbirnbaum/psych101/eliza.htm

Natural Language Processing (NLP)

Natural Language Processing is the set of methods for making human language accessible to computers… or, "whatever you do with NL (text/voice)"

Some examples:

- speech recognition,
- natural language understanding (NLU), and
- natural language generation (NLG).

ELIZA is an example of NLP application which mimic NLU

NLP: applications

Real examples of NLP applications:

- **automatic machine translation** is ubiquitous on the web/media
- **text classification** keeps our email inboxes free of spam
- **search engines** have moved beyond string matching, IR, and network analysis to a high degree of linguistic sophistication (with **knowledge graphs**)
- **chatbots/dialog systems** provide an increasingly common and effective way to get and share information
- **voice assistants** such as Alexa help to accomplish daily tasks at home

Google

list of nlp tasks

Immagini | Video | Notizie | Siti per offerte di lavoro | Libri | Finanza | Tutti i filtri

Circa 29.800.000 risultati (0,66 secondi)

NLP tasks

Da fonti sul Web

- Named-entity recognition
- Sentiment analysis
- Speech recognition
- Stemming
- Keyword extraction
- Coreference resolution
- Customer service autom...
- Similarity
- Summarization
- Part-of-speech tagging
- Chatbots
- Text generation
- Language identification
- Search results
- Email filtering
- Dependency parsing
- Machine translation
- Text classification
- Question answering
- Lexical analysis
- Spam detection
- Translation
- Identifying stopwords
- Sentence segmentation

Mostra meno

NLP-progress

Repository to track the progress in Natural Language Processing (NLP), including the datasets and the current state-of-the-art for the most common NLP tasks.

Tracking Progress in Natural Language Processing

Table of contents

English

NLP: related areas

These diverse applications are based on a common set of ideas, drawing on algorithms, linguistics, logic, statistics.

Computational Linguistics: the object of the study is the language itself, and it's done by means of computational methods (e.g., as in computational astronomy)

Artificial Intelligence: the object of the study of reproducing intelligence of any form, by means of several techniques such as reasoning over knowledge bases

Machine Learning (ML): the object of the study is automatic learning, usually from examples. Currently most NLP approaches are based on ML, either supervised or unsupervised

Machine Learning: a brief introduction

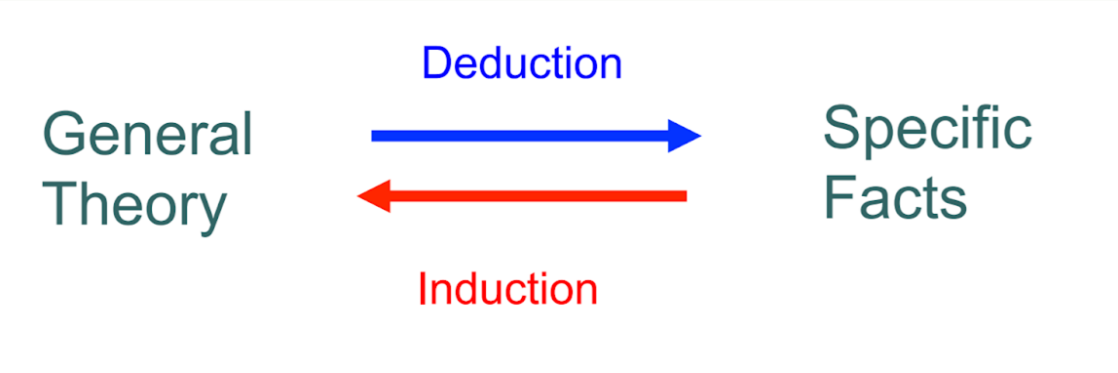
What is Machine Learning?

- Programs that get better with experience given a task and some performance Measure… some examples:
  - Learning to classify news articles
  - Learning to recognize spoken words
  - Learning to play board games
  - Learning to navigate a virtual world
  - Usually involves some sort of inductive reasoning step
- Usually involves some sort of **inductive reasoning** step

Next slides from https://homepage.cs.uri.edu/faculty/hamel/courses/2015/spring2015/csc481/lecture-notes/ based on work by D. Poole and A. Mackworth (see their book)

Reasoning

- Deductive reasoning (rule based reasoning)
  - From the general to the specific (All Cretans are liars, Epimenides is a cretan)
- Inductive reasoning
  - From the specific to the general (Epimenides is a liar, Kresilas is a liar, Nearchus is a liar)



Arrows represent inference. **Inference** is the act or process of drawing a conclusion based solely on what one already knows.

Note: very different from Mathematical Induction!

Inductive Reasoning: example

- Facts: every time you see a swan you notice that the swan is white.
- Inductive step: you infer that all swans are white.

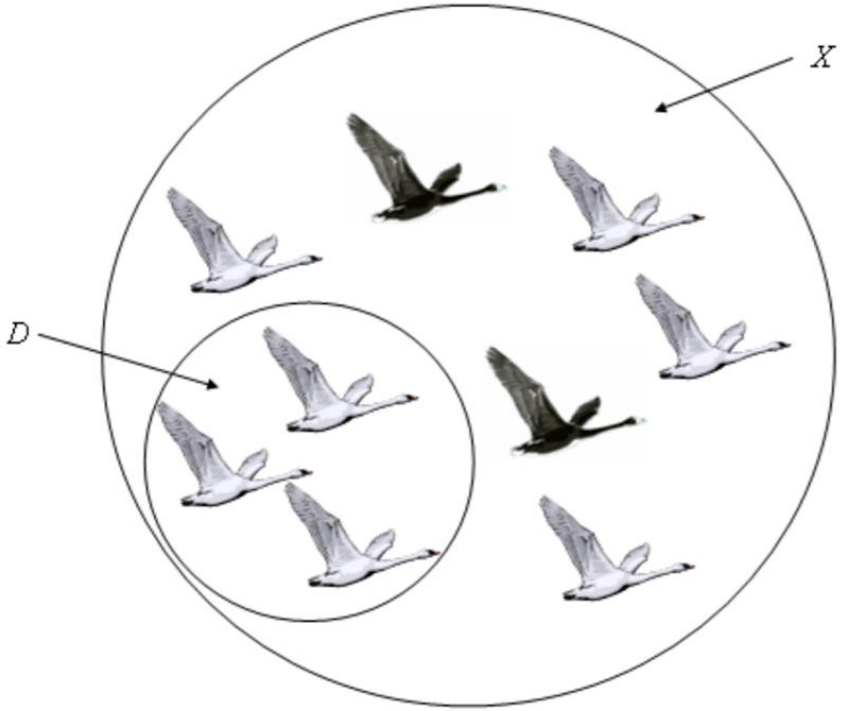
Observed Swans are white.

Induction

All Swans are white.

Reasoning

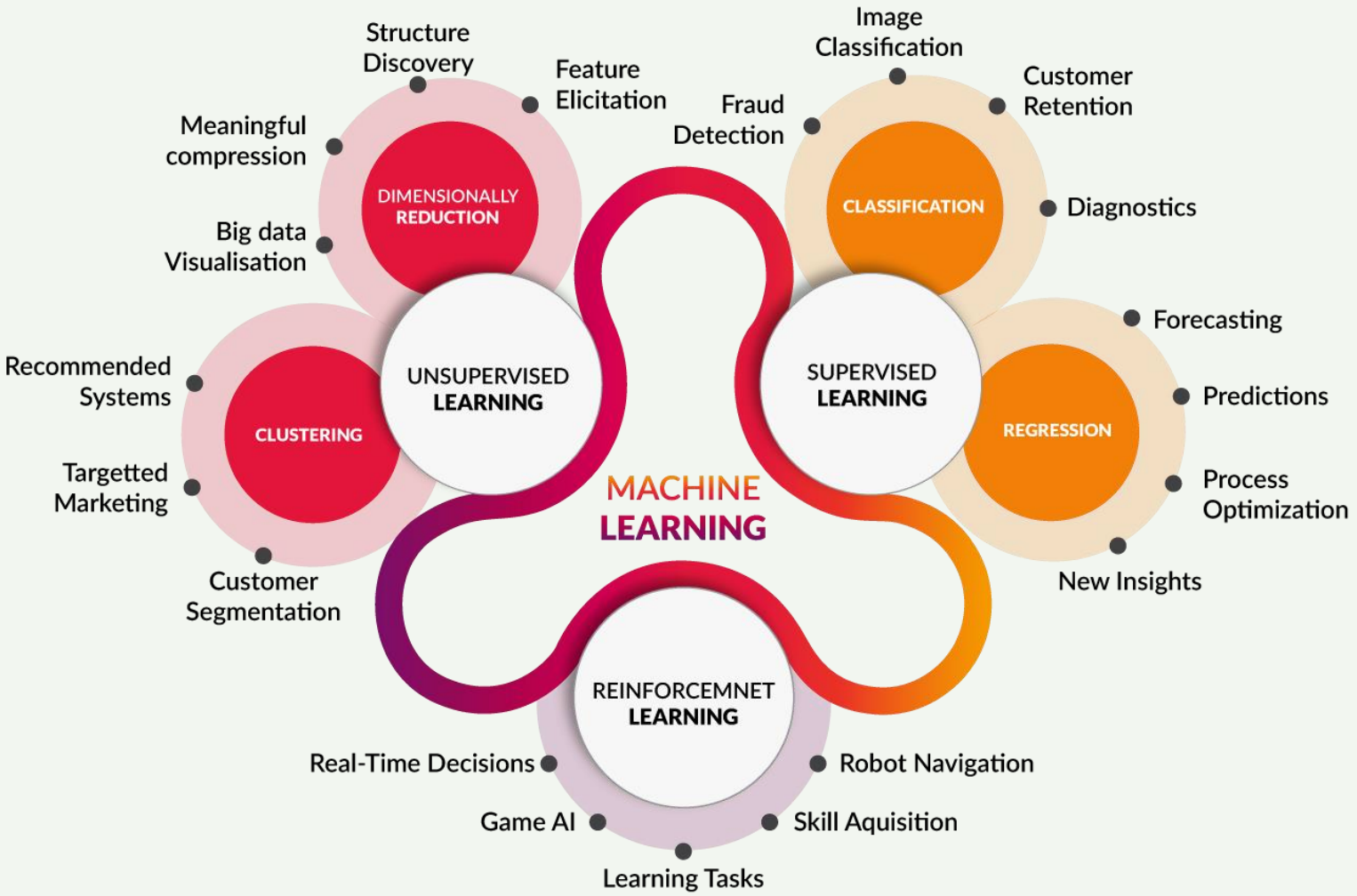
- Deduction is “truth preserving”
  - If the rules employed in the deductive reasoning process are sound, then, what holds in the theory will hold for the deduced facts
  - There exist also probabilistic deductive reasoning frameworks
- Induction is NOT “truth preserving”
  - It is more of a statistical argument
  - The more swans you see that are white, the more probable it is that all swans are white. But this does not exclude the existence of black swans



D ≡ observations
X ≡ universe of all swans

Different styles of ML: supervision

- Supervised Learning
  - The learning needs explicit examples of the concept to be learned (e.g. white swans…)
- Unsupervised Learning
  - The learner discovers autonomously any structure in the domain that might represent an interesting concept
- Reinforcement Learning
  - The learner gets feedbacks from his own past behaviour and learn from it. E.g. it plays against itself, improving time over time


MACHINE LEARNING
UNSUPERVISED LEARNING
DIMENSIONALLY REDUCTION
Structure Discovery
Feature Elicitation
Meaningful compression
Big data Visualisation
CLUSTERING
Recommended Systems
Targetted Marketing
Customer Segmentation
SUPERVISED LEARNING
CLASSIFICATION
Image Classification
Customer Retention
Fraud Detection
Diagnostics
REGRESSION
Forecasting
Predictions
Process Optimization
New Insights
REINFORCEMNET LEARNING
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

Different styles of ML: knowledge representation

- Symbolic Learners (transparent models), e.g.:
  - If-then-else rules
  - Decision trees
  - Association rules
- Sub-Symbolic Learners (non-transparent models), e.g.:
  - Neural Networks
  - Clustering (Self-Organizing Maps, k-Means)
  - Support Vector Machines

What are they?

Intro on AI, ML and NLP

Performance Evaluation

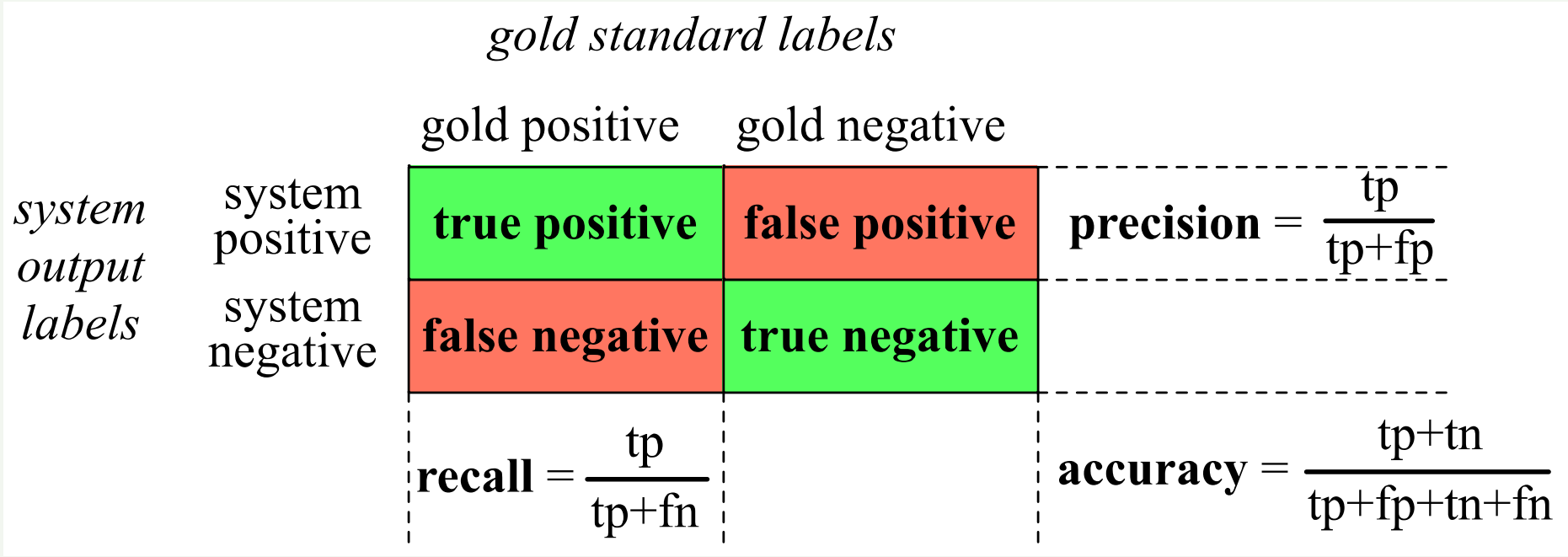
Intro on AI, ML and NLP

Evaluation

Let's consider just binary text classification tasks
Imagine you're the CEO of Delicious Pie Company
You want to know what people are saying about your pies
So you build a "Delicious Pie" tweet detector

- Positive class: tweets about Delicious Pie Co
- Negative class: all other tweets

The 2-by-2 confusion matrix



Evaluation: Accuracy

Why don't we use **accuracy** as our metric?

Imagine we saw 1 million tweets

- 100 of them talked about Delicious Pie Co.
- 999,900 talked about something else

We could build a dumb classifier that just labels every tweet "not about pie"

- It would get 99.99% accuracy!!! Wow!!!!
- But useless! Doesn't return the comments we are looking for!
- That's why we use **precision** and **recall** instead

Evaluation: Precision

% of items the system detected (i.e., items the system labeled as positive) that are in fact positive (according to the human gold labels)

$$\textbf{Precision} = \frac{\text{true positives}}{\text{true positives + false positives}}$$

Evaluation: Recall

% of items actually present in the input that were correctly identified by the system.

$$\textbf{Recall} = \frac{\text{true positives}}{\text{true positives + false negatives}}$$

Why Precision and recall

Our dumb pie-classifier

- Just label nothing as "about pie"

Accuracy=99.99%

but

Recall = 0

- (it doesn't get any of the 100 Pie tweets)

Precision and recall, unlike accuracy, emphasize true positives:

- finding the things that we are supposed to be looking for.

A combined measure: F

F measure: a single number that combines P and R:

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

We almost always use balanced F_1 (i.e., $\beta = 1$)

$$\mathrm{F}_1 = \frac{2PR}{P + R}$$

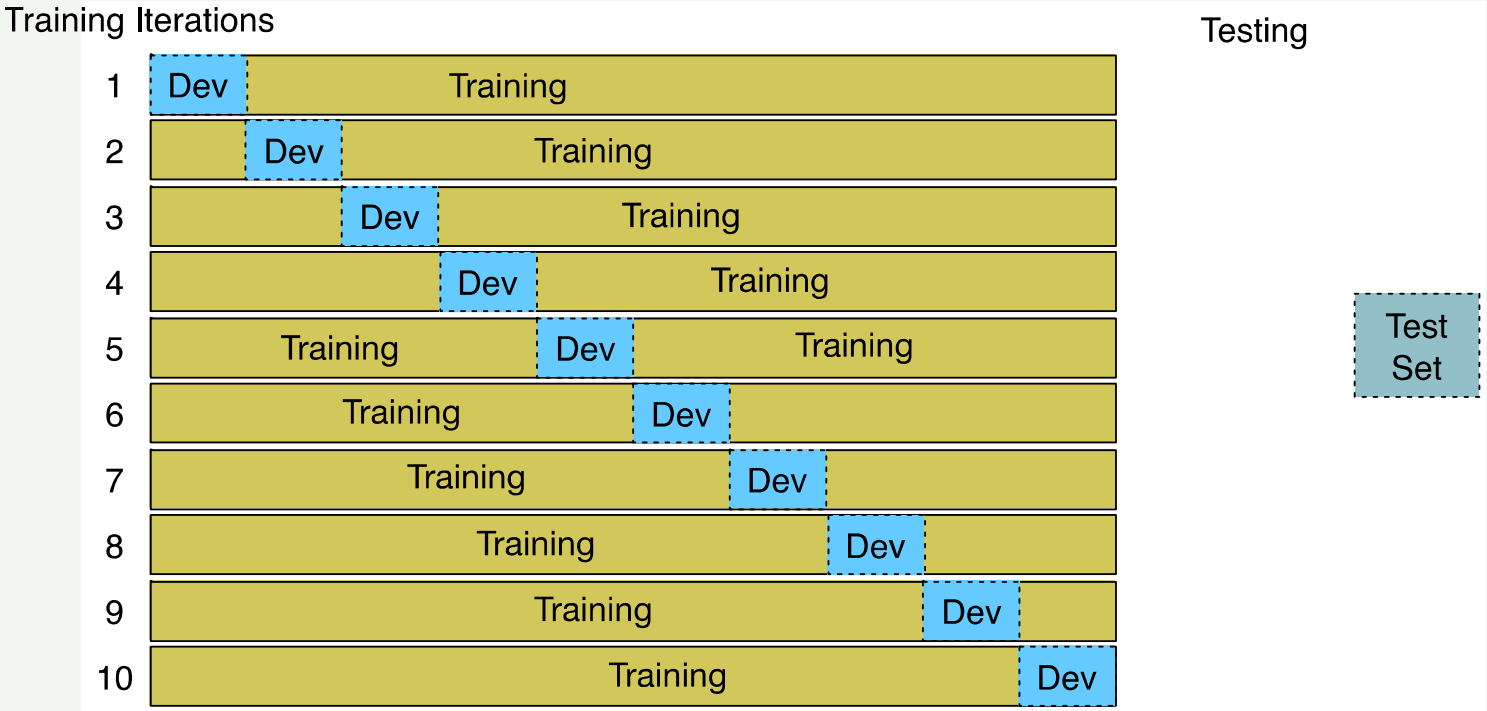
Development Test Sets ("Devsets") and Cross-validation

Train on training set, tune on devset, report on testset

- This avoids overfitting (‘tuning to the test set’)
- More conservative estimate of performance
- But paradox: want as much data as possible for training, and as much for dev; how to split?

Cross-validation: multiple splits

Pool results over splits, Compute pooled dev performance



Exercise

Precision: percentage of correct results among those found

Recall: percentage of correct results among all that are known to be correct

On Google Colaboratory

Performance Evaluation

Intro on AI, ML and NLP