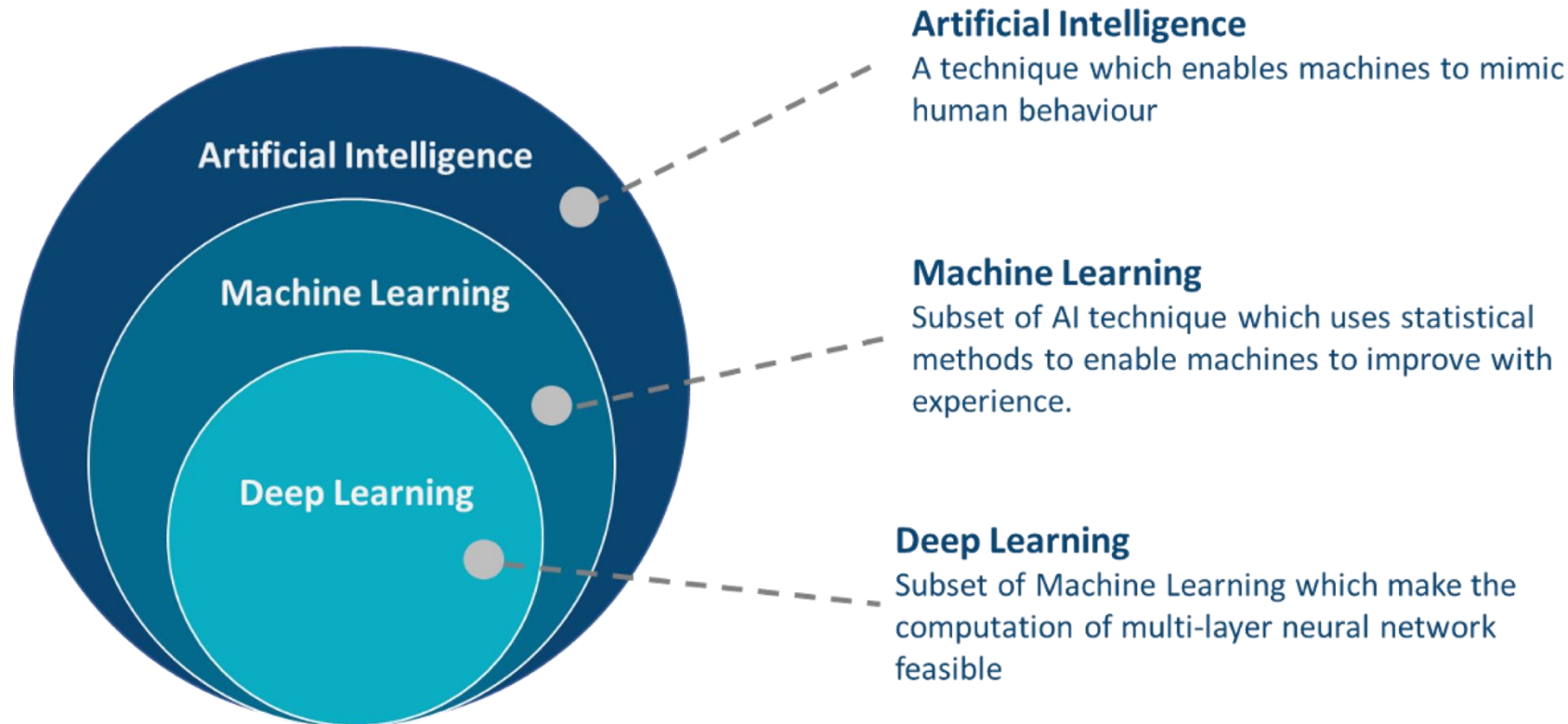




# 1. Introduction

COMP3314  
Machine Learning

# What is AI? ML? DL?



# History of AI



“Can machines think?”

First AI Winter

Second AI Winter

AlphaGo

1950

1966

1975

1980

1987

2016

First Chatbot: ELIZA

```

***** LL  IIII ZIIIII  AAAA
EE  LL  II  II  AA  AA
EEEE  LL  II  ZII  AAAAAA
EE  LL  II  II  AA  AA
***** LLLLLL IIII ZIIIII  AA  AA

ELIZA is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
  
```

Expert System

## Development and Evaluation of a Computer Expert System for the Management of Fetal Distress

K. K. Wong,<sup>1)</sup> K. H. Ng,<sup>2)</sup> S. H. Nah,<sup>3)</sup> K. Yusof,<sup>4)</sup> and K. Rajeswari<sup>1)</sup>

<sup>1)</sup> Department of Obstetrics and Gynaecology, University of Malaya, Kuala Lumpur, Malaysia

<sup>2)</sup> Department of Radiology, University of Malaya, Kuala Lumpur, Malaysia

<sup>3)</sup> Berkeley Systems Sdn Bhd, Petaling Jaya, Malaysia

*Asia-Oceania J. Obstet. Gynaecol. Vol. 20, No. 1: 19-23 19*

# ChatGPT (November 2022)

## ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

FL Hi, I'm writing an article about you and all of the cool things you can do. Could you say hello to my readers and give us a quick self-introduction? I'd like it to be friendly and casual, include a silly joke, and end with a note about how great the rest of my article is.

Hello readers! My name is Assistant, and I'm a large language model trained by OpenAI. I'm here to help answer your questions and provide you with information on a wide variety of topics. I'm a bit of a jokester, so here's a quick one for you: Why was the math book sad? Because it had too many problems! But seriously, I'm excited to be featured in your article and I hope you enjoy learning more about me and all the cool things I can do. And trust me, the rest of the article is going to be great!

FL Can you write me a paragraph about someone trying to sell an egg but they think the egg is actually a drawing tool?

Sure! Here's a paragraph:

The person was trying to sell their egg at a street market, but they thought it was actually a drawing tool. They were confidently telling potential buyers about the egg's ability to create beautiful, detailed sketches with its delicate, curved shape. Some curious passersby stopped to listen and even gave the egg a try, but they were quickly disappointed to find that it was just an ordinary egg and not a drawing tool at all. The person was embarrassed and quickly moved on to the next item in their sales pitch.

FL Can you write a version of their sales pitch?

Sure! Here's a possible sales pitch for the egg-turned-drawing-tool:

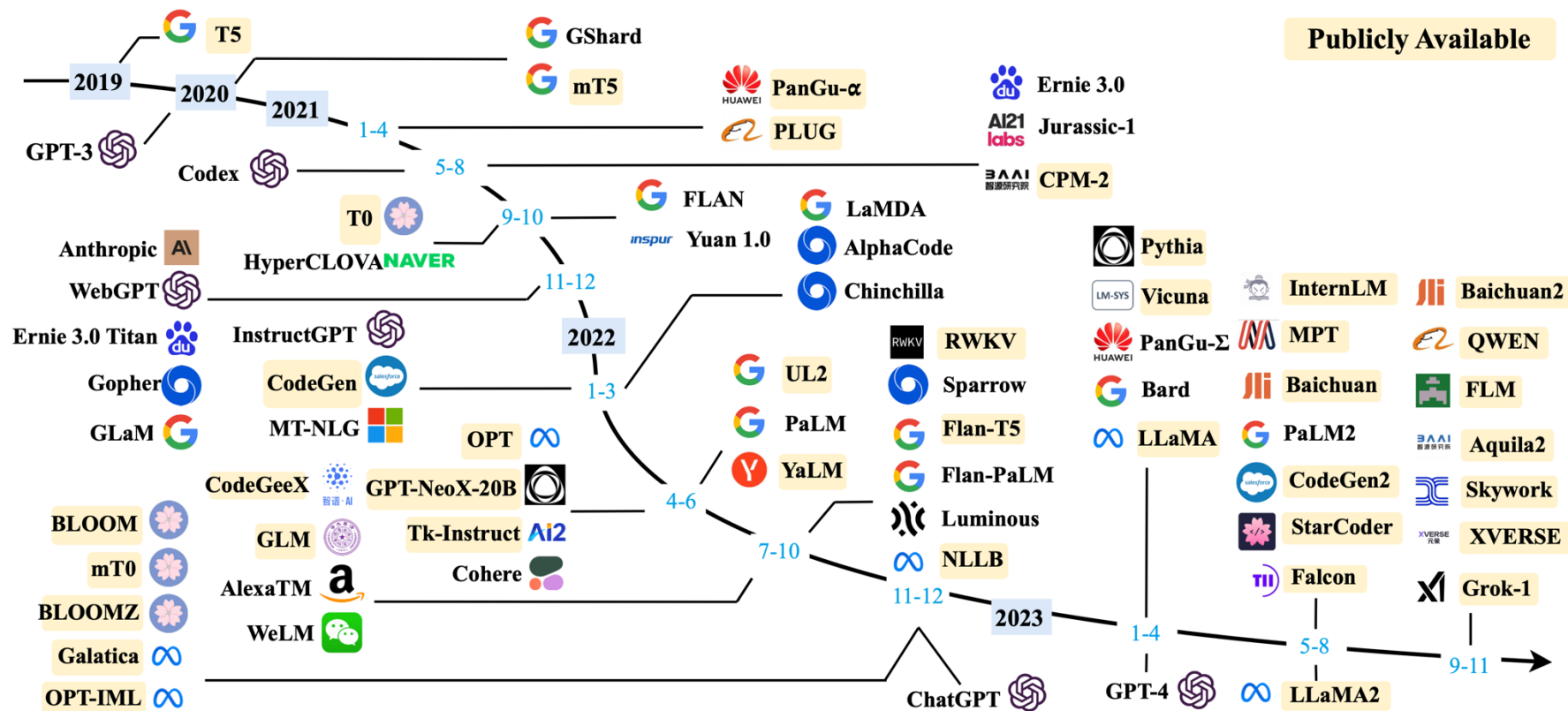
Hello, friends! I have a very special item to show you today. This is no ordinary egg - it's a one-of-a-kind drawing tool. With its delicate, curved shape, this egg is perfect for creating beautiful, detailed sketches. Just hold it in your hand and let your creativity flow. You'll be amazed at the incredible art you can create with this amazing egg. And the best part? It's only \$5! Don't miss out on this incredible deal. Come and see for yourself how this egg can unleash your inner artist.

FL They also think the egg has the capability to brainwash humans who lick it, so they always include a warning in their pitch.

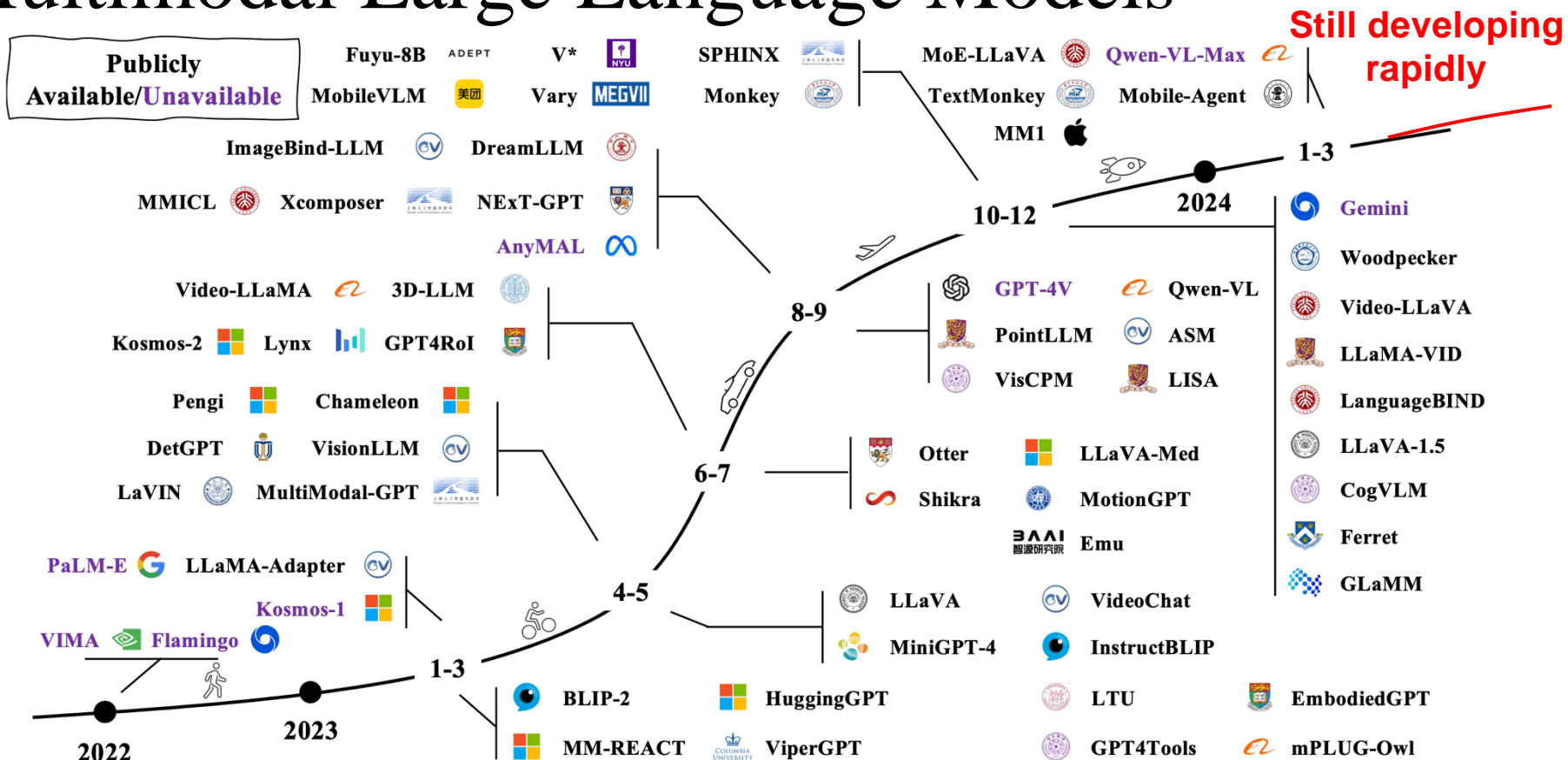
Sure, here's a possible warning to include in the sales pitch for the egg-turned-drawing-tool:

**Warning:** This egg has the unique ability to brainwash humans who lick it. Please be careful when handling this egg and do not lick it under any circumstances. We want you to enjoy the incredible benefits of this egg as a drawing tool, but we also want to ensure your safety and well-being. Thank you for understanding.

# Large Language Models



# Multimodal Large Language Models




# AlphaGo





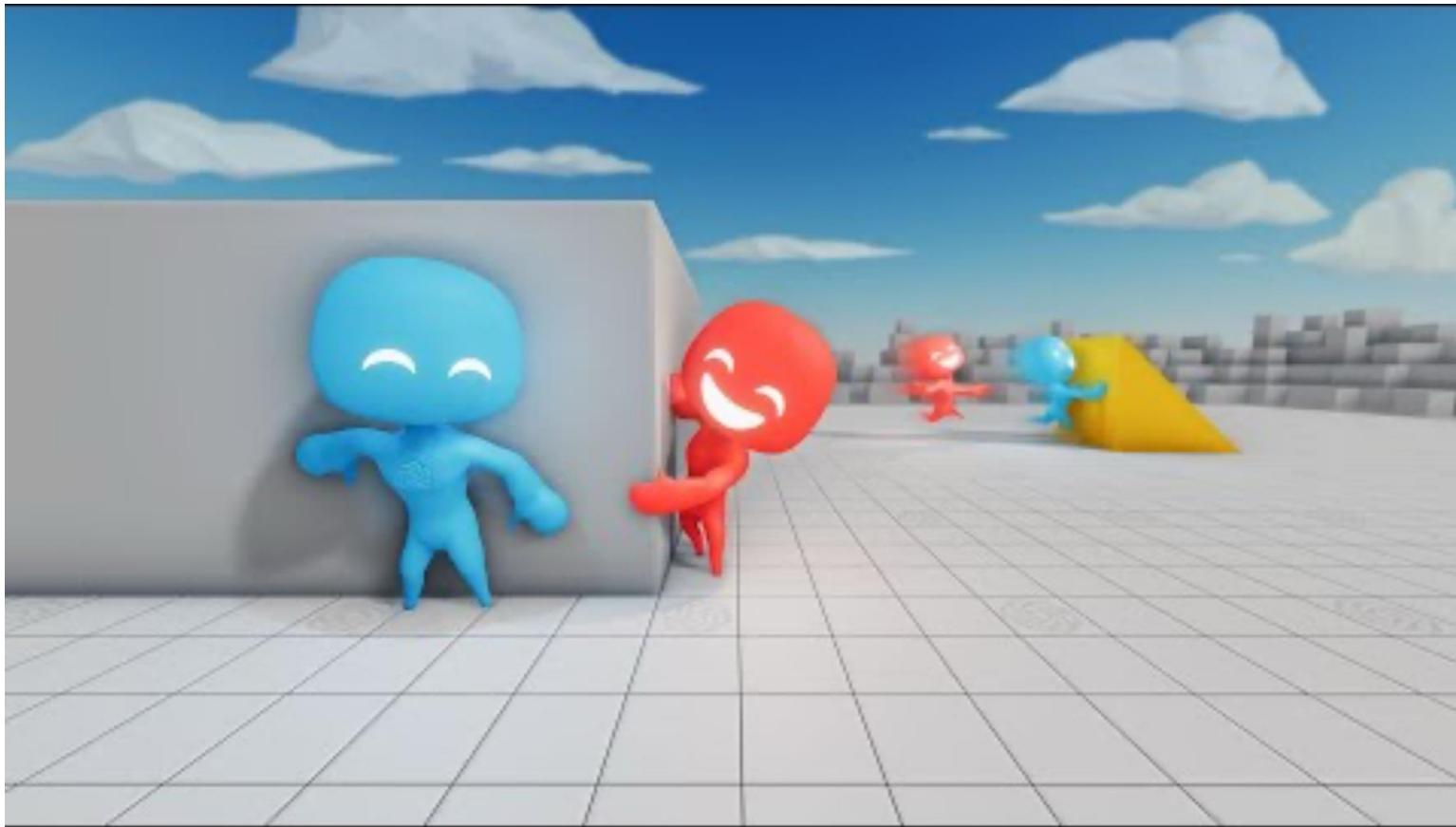
# AlphaFold



**Protein  
folding  
explained**



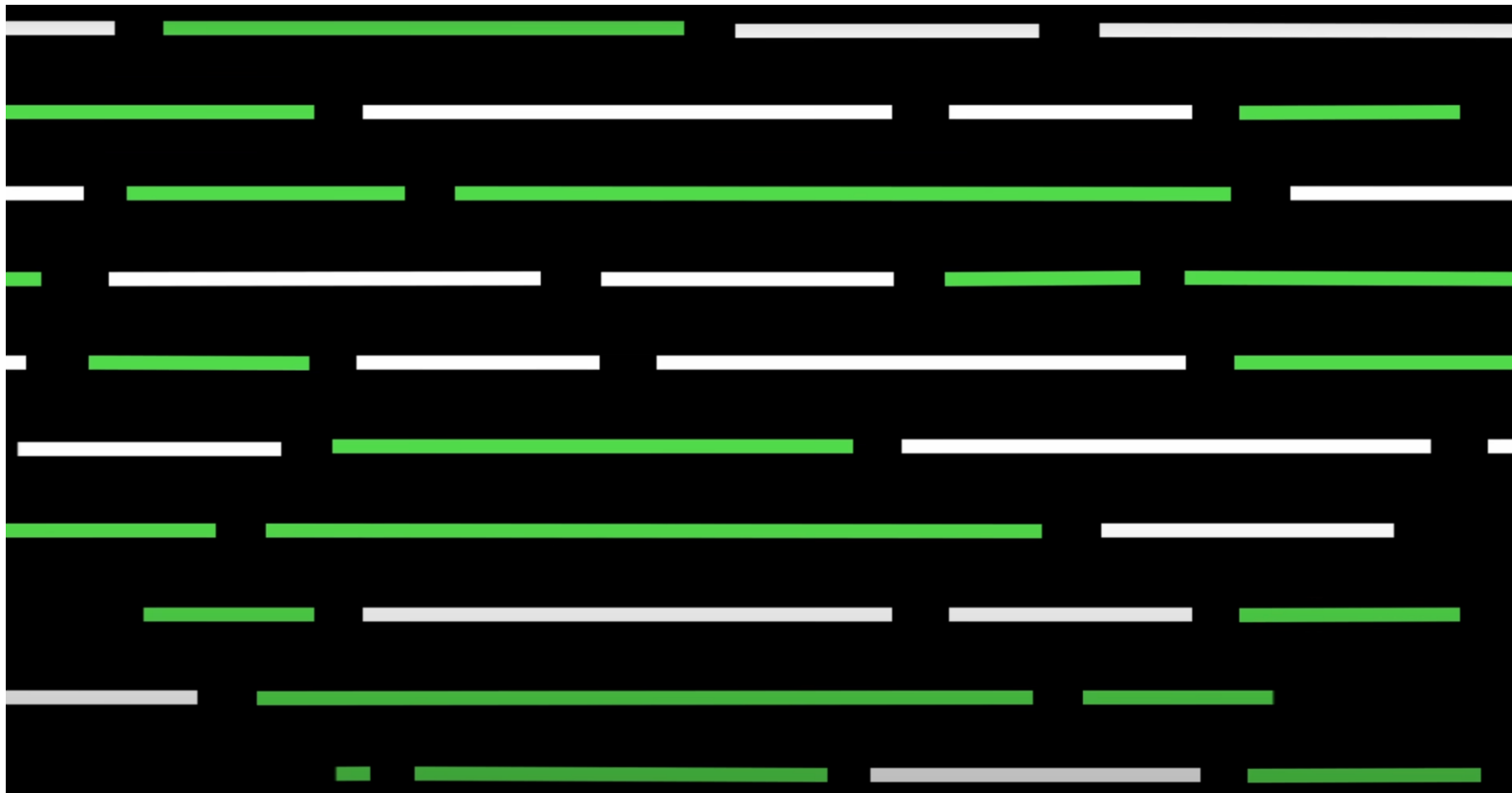
# Multi-Agent Hide and Seek



# DALL·E 2



# GPT4



# Outline

- Motivation
- Types of ML
  - Supervised Learning
    - Classification
    - Regression
  - Reinforcement Learning
    - Chess
  - Unsupervised Learning
    - Clustering
    - Dimensionality Reduction
- Terminology and Notation
- Roadmap
  - Preprocessing
  - Learning
  - Evaluation and Prediction
- Python
  - Installation
- Linear Algebra Review
- References

# Motivation

- Nowadays large amount of structured and unstructured data is available
- ML algorithms can turn this data into knowledge
  - Powerful open source libraries available to do this
- In this course you will understand how these algorithms work
- You will also learn how to utilize them to make predictions

# Motivation

- ML algorithms are self-learning
  - Automatically derive knowledge from data to make predictions
    - No need for humans to manually derive rules
    - ML offers a more efficient alternative for capturing the knowledge in data to gradually improve the performance of predictive models
- ML becomes increasingly relevant in CS research
  - More importantly
    - Plays an ever greater role in our everyday lives

# How do you use machine learning everyday?



# Examples of Machine Learning

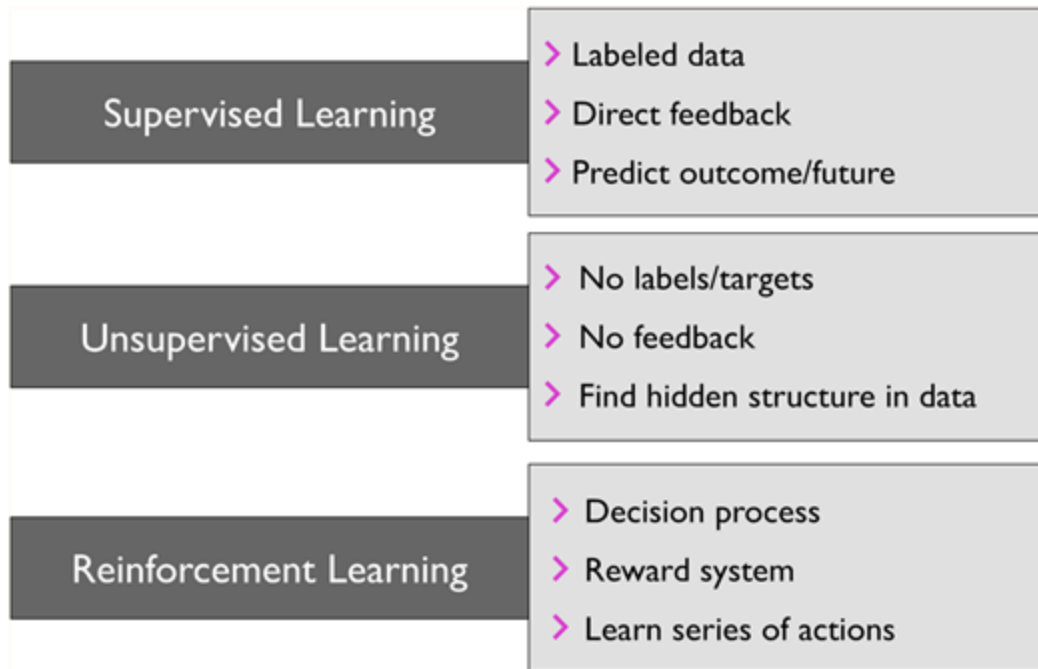
- Basket analysis
- Credit scoring
- Medical diagnosis
- Biometrics
- Object recognition
- Service recommendations
- Understanding human learning

# Machine Learning Definition

- Subfield of Artificial Intelligence (AI)
- Arthur Samuel (1959)
  - Field of study that gives computers the ability to learn without being explicitly programmed
- Tom Mitchell (1998)
  - A computer program is said to learn from experience  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$

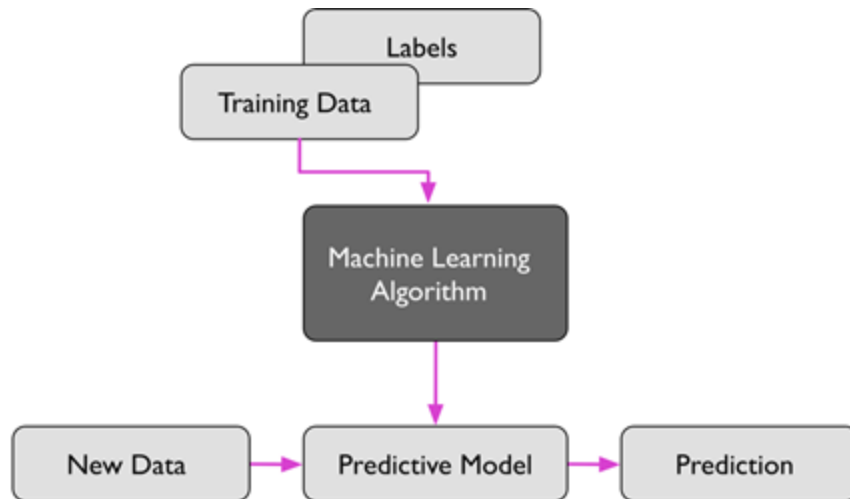
# Types of Machine Learning

- In the following we will consider three types of machine learning



# Supervised Learning

- Learn from labeled training data
  - Make predictions about unseen / future data
- Supervised refers to a set of samples where the desired output signals (labels) are already known



# Supervised Learning:

## Classification vs. Regression

- Two subcategories of supervised learning
  - Classification
    - A supervised learning task with discrete class labels
      - E.g., spam email classifier
  - Regression
    - Outcome is a continuous value
      - E.g., student exam score prediction

# Supervised Learning - Classification

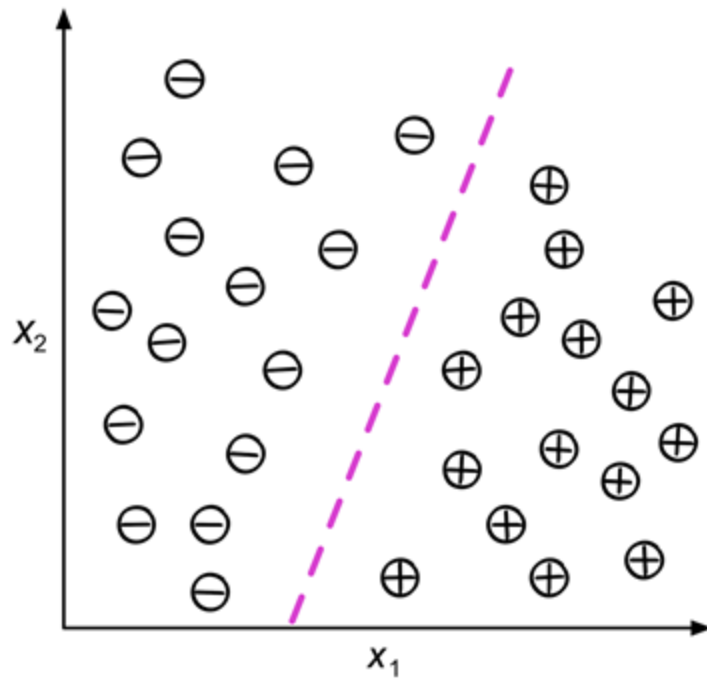
- Goal: Predict class labels of new instances, based on past observations
- Class labels are discrete, unordered values
- Two subcategories of classifiers:
  - Binary classification
    - Only two possible class labels can be assigned
      - E.g., spam vs. non-spam emails
  - Multiclass classification
    - Any fixed number  $> 2$  of class labels can be assigned
      - E.g., handwritten character recognition



[MNIST dataset](#)

# Classification - Example

- Given 30 training samples
  - 15 labeled as negative class
  - 15 labeled as positive class
- Let each sample have 2 dimensions
- Classifier will learn the decision boundary
  - Represented as a dashed line
  - Able to separate the two classes



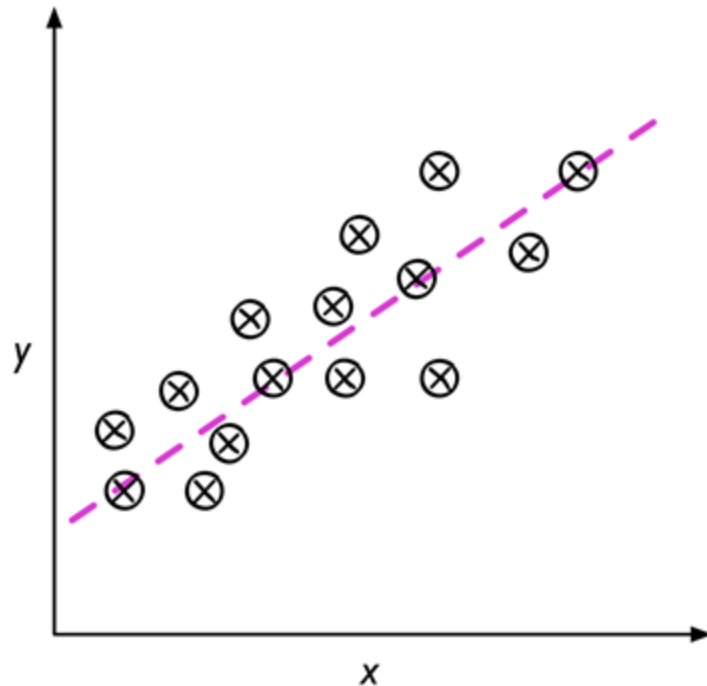


# Regression

- Prediction of continuous outcome
  - The term regression was devised by Francis Galton in his article [Regression towards Mediocrity](#) in 1886
- Example:
  - Predicting the exam scores given time spent studying

# Regression - Example

- Given
  - Predictor variable  $x$
  - Response variable  $y$
- I.e., 1D data set
- Fit a line to it minimizing the distance between sample points and the fitted line
  - Average squared distance is most commonly used
- Use the line to predict outcome of new data

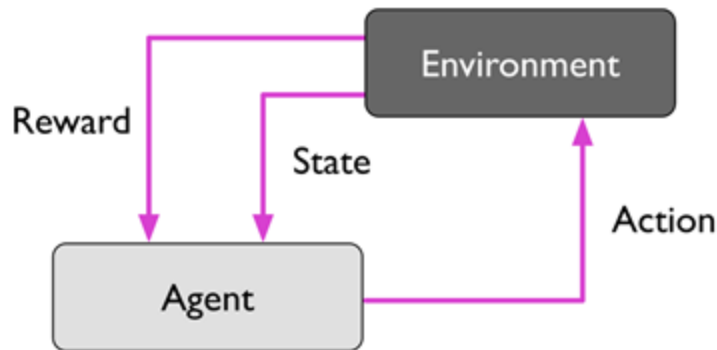


# Quiz

- Consider the following supervised ML tasks. Label each task with *Classification Task* or *Regression Task*
  - a. You are working for an investment bank and your task is to predict investors sentiment for certain stocks by analyzing popular online investments forums
  - b. You are working for a property agency and your task is to predict the housing price for a property based on past data that the agency has available in their database
  - c. Your task is to analyze a video stream of the western harbour tunnel and count how many Tesla pass by every day

# Reinforcement Learning

- The system (aka agent) improves its performance based on interactions with an environment
- Trial-and-Error approach
  - Learning by doing
- The agent receives feedback (reward) from the environment
  - This reward is not the correct ground truth
    - It is a sample experience
  - Extensive interaction with the environment allows agent to learn a series of actions that maximizes this reward



# Reinforcement Learning - Example: Chess

- Agent decides upon a series of moves depending on state of board
  - Environment is the board
  - Reward can be defined as win or lose at the end of the game
- Outcome of each move results in different state of the environment
  - Removing an opponent's chess piece from the board or threatening the queen is associated with a positive event
  - Losing a chess piece to the opponent is associated with a negative event
- Note: Not every turn results in the removal of a chess piece
  - Reinforcement learning is concerned with learning the series of steps by maximizing a reward based on immediate and delayed feedback

# Reinforcement Learning - Example

- Learning to walk

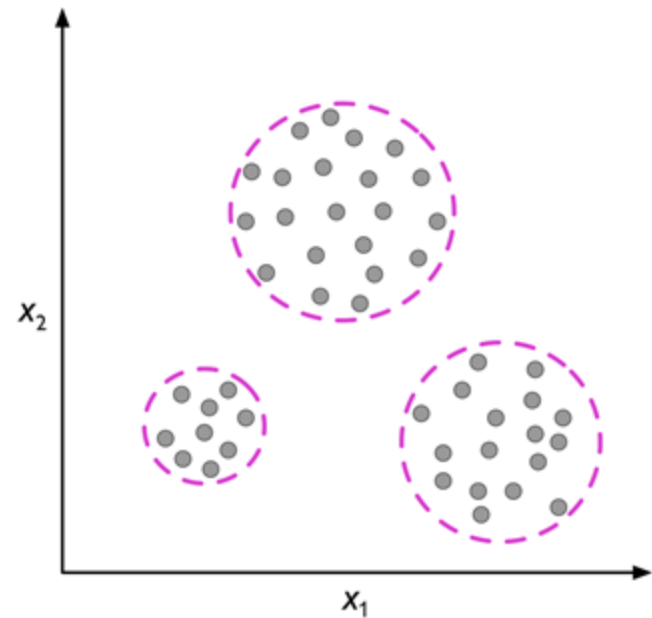
# Unsupervised Learning

- Unlabeled data / data of unknown structure
- Explores the structure of data
  - Extract meaningful information without guidance of known outcome variable / reward function
- Examples
  - Clustering
  - Dimensionality reduction



# Clustering

- Exploratory data analysis technique
- Organizes information into meaningful subgroups (clusters) without having any knowledge of group memberships
- Each cluster defines a group of objects that share a certain degree of similarity but are more dissimilar to objects in other clusters



# Clustering - Example

## President Trump suggests taking hydroxychloroquine to protect against Covid-19 despite lack of proof

CNN · 1 hour ago



- **Trump Again Promotes Hydroxychloroquine For Coronavirus Treatment**

NPR · 1 hour ago

- **Trump and officials warn the worst is yet to come**

CBS Evening News · 4 hours ago

- **Andy Puzder In coronavirus crisis, Trump displays leadership Americans expect and want**

Fox News · Yesterday · Opinion

Rectangular Snip

- **Charlie Kirk: Time for President Trump's New Wollman Rink Moment | Opinion**

Newsweek · Yesterday · Opinion

[View Full Coverage](#)

Full Coverage

Sort

## Boris Johnson admitted to hospital with coronavirus

Top coverage

SLATE

British Prime Minister Boris Johnson Admitted to Hospital for Coronavirus

2 hours ago



Fox News

Boris Johnson hospitalized after experiencing coronavirus symptoms, PM's office says

7 hours ago



ON

Boris Johnson admitted to hospital for tests. He has previously tested positive for coronavirus

2 hours ago



ox news

PM's Covid-19 timeline: from 'mild symptoms' to hospital admission

4 hours ago



# Clustering - More Examples

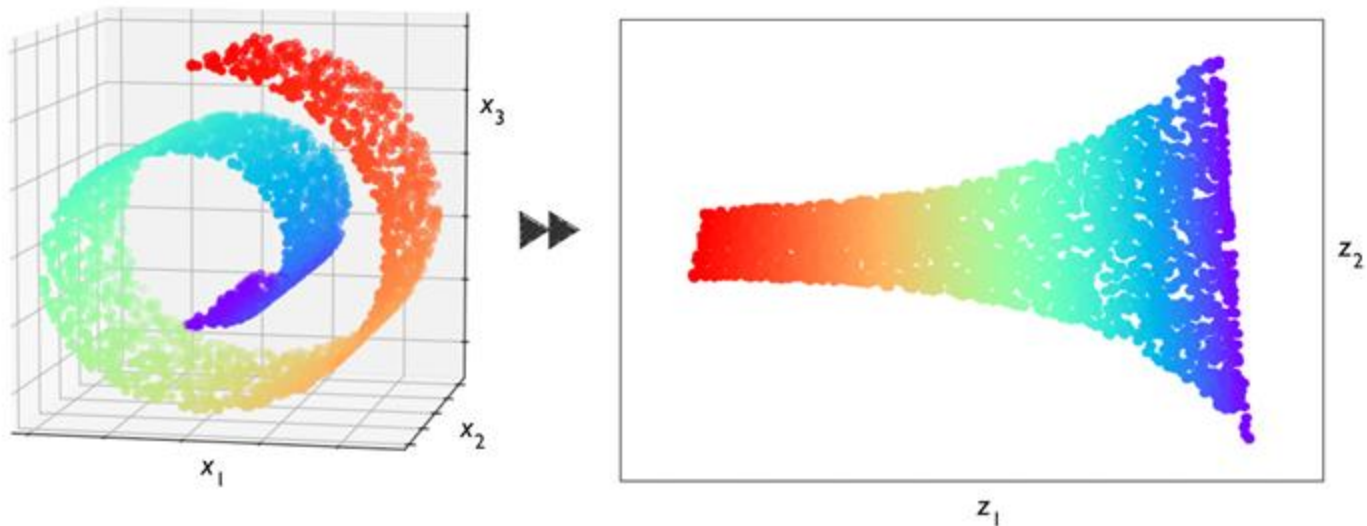
- Human genetic clustering
- Sequence clustering
- Social network analysis
- Market research
- Grouping of shopping items

# Dimensionality Reduction

- Often we are working with data of high dimensionality
  - I.e., each observation comes with a high number of measurements
- High dimensional data can present a challenge
  - Computational performance
  - Predictive performance
  - Visualization
- Dimensionality reduction is a commonly used approach in feature preprocessing
  - Compress data onto a smaller dimensional subspace
  - Retaining most relevant information

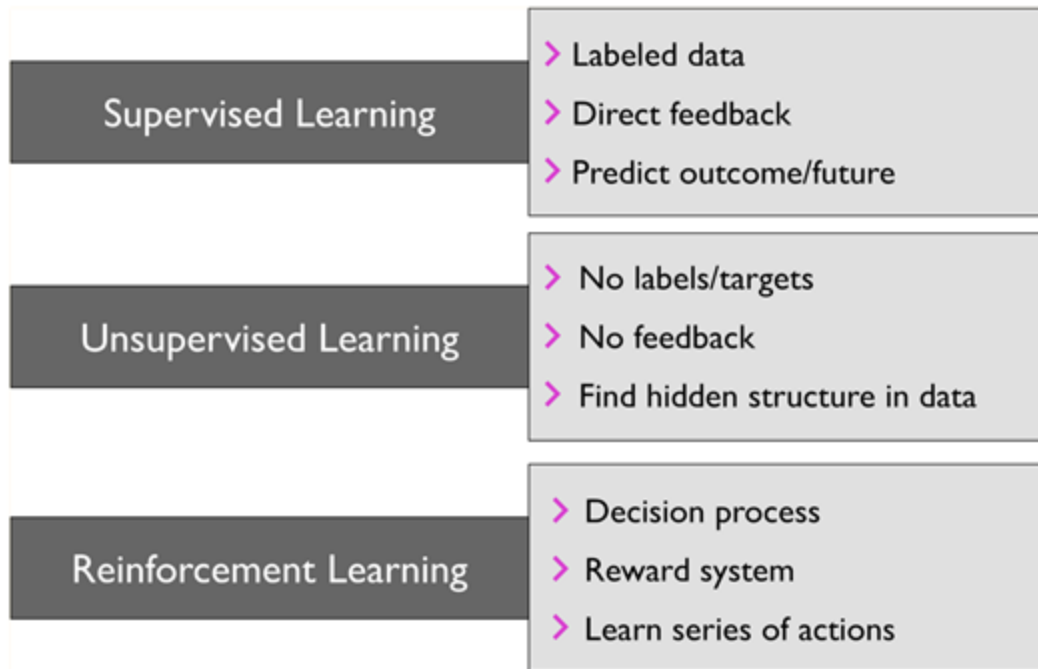
# Dimensionality Reduction - Example

- High-dimensional feature set can be projected onto 1D, 2D or 3D feature spaces
  - 3D to 2D example



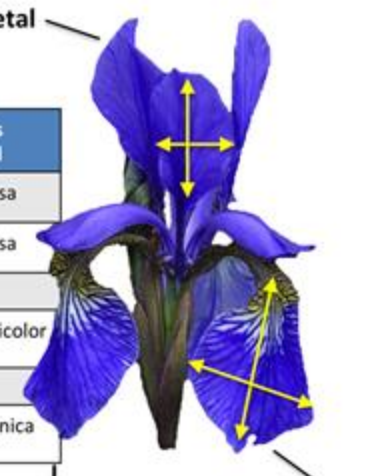
# Types of machine learning

- In the following we will consider three types of machine learning



# Terminology and Notations

- [Iris flower data set](#) contains measurements of 150 Iris flowers from three different species
  - Setosa, Versicolor, and Virginica
- Introduced in [Fisher's 1936 paper](#) [The use of multiple measurements in taxonomic problems](#)
- Row
  - A single flower sample
- Column
  - Flower features (measurements in centimeters)



The diagram shows a blue Iris flower. Yellow arrows indicate measurements: one arrow points to a petal with the label 'Petal', and another arrow points to a sepal with the label 'Sepal'. The table below represents the data for 150 samples, with columns for features and a column for class labels.

Samples (instances, observations)					
	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features  
(attributes, measurements, dimensions)

Class labels  
(targets)



# Terminology and Notations

- We will use a matrix and vector notation to refer to our data
- Each sample is a separate row in a feature matrix  $\mathbf{X}$ , where each feature is stored as a separate column
- Iris dataset example
  - 150 samples and four features are written as a 150 x 4 matrix  $\mathbf{X}$

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & x_4^{(150)} \end{bmatrix}$$

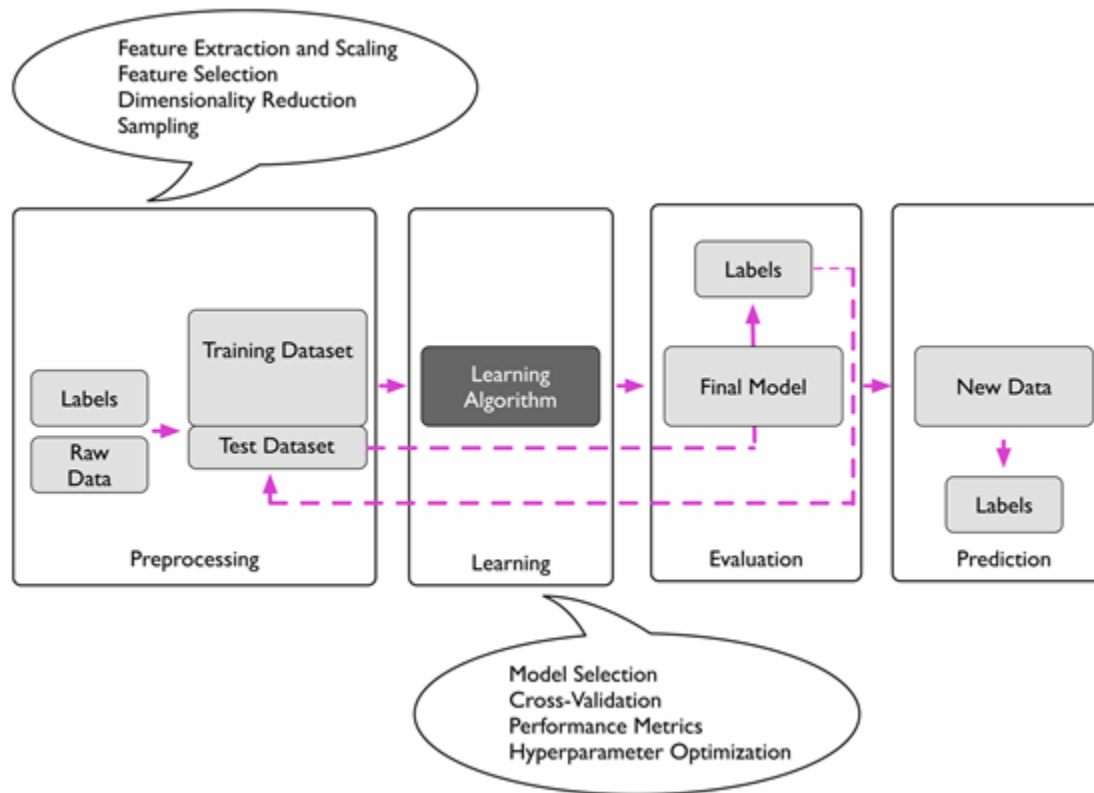
# Terminology and Notations

- We will use the superscript  $i$  to refer to the  $i$ th training sample, and the subscript  $j$  to refer to the  $j$ th dimension of the dataset
  - For example  $x_j^{(i)} = x_1^{(150)}$  refers to the first dimension of the flower and sample 150
- We use lowercase, bold-face letters to refer to vectors and uppercase, bold-face letters to refer to matrices
- Note that each row in the iris dataset  $\mathbf{X}$  can be written as a four-dimensional row vector and each feature dimension is a 150-dimensional column vector

$$\mathbf{x}^{(i)} = \begin{bmatrix} x_1^{(i)} & x_2^{(i)} & x_3^{(i)} & x_4^{(i)} \end{bmatrix} \quad \mathbf{x}_j = \begin{bmatrix} x_j^{(1)} \\ x_j^{(2)} \\ \vdots \\ x_j^{(150)} \end{bmatrix}$$

# Roadmap

- Typical workflow for using ML in predictive modeling



# Preprocessing

- Preprocessing of the data is a crucial steps in any ML application
- Feature selection, extraction and scaling
  - Select and extract useful features from raw data
  - Many algorithms also require that the selected features are on the same scale
- Dimensionality reduction
  - May improve
    - Computational performance
    - Predictive performance
- Sampling
  - Randomly divide the dataset into a separate training and test set to determine whether our algorithm not only performs well on the training set but also generalizes well to new data
  - Keep the test set until the very end to evaluate the final model

# Learning

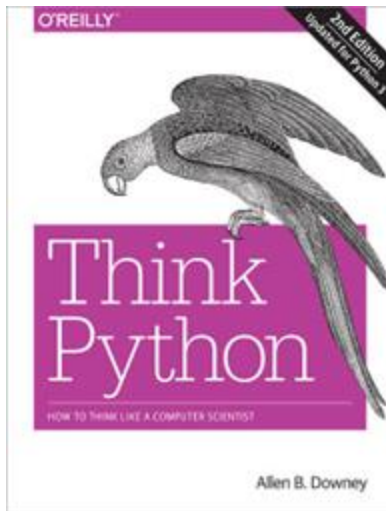
- Model selection
  - Compare algorithms and select the best performing model
- Cross-validation
  - How do we know which model performs well on the final test dataset if we don't use this test set for model selection?
    - Cross-validation splits the training dataset further into training and validation subsets
- Performance metric
  - Decide upon a metric to measure performance
- Hyperparameter optimization
  - Fine-tune parameters of the model based on performance on validation set

# Evaluation and Prediction

- After model selection and training we use the test dataset to estimate how well it performs on unseen data
  - Estimate the generalization error
- If we are satisfied with its performance, we can now use this model to predict new data
- Important
  - Parameters for the previously mentioned procedures, such as feature scaling and dimensionality reduction, are solely obtained from the training dataset
  - The same parameters are later reapplied to the test data and any new data samples

# Python

- We assume that you are familiar with the basics of python
  - Recommended textbook



Free PDF [here](#)

# Programming Environment: Local

- In this course we are going to use
  - Python 3, NumPy, Matplotlib, SciPy and Jupyter

The image shows the Anaconda website and a Windows Start menu. The website, titled "Anaconda Installers", has a blue background and displays installation options for Windows, Mac, and Linux, all for Python 3.11. A red arrow points to the "64-Bit Graphical Installer (898.6 MB)" for Windows. The Windows Start menu on the right shows the "Jupyter Notebook (anaconda3)" application highlighted, with a red arrow pointing to it from the website.

**ANACONDA** | Platform Solutions Partners Pricing Resources Company

Sign-in Contact Sales

## Anaconda Installers

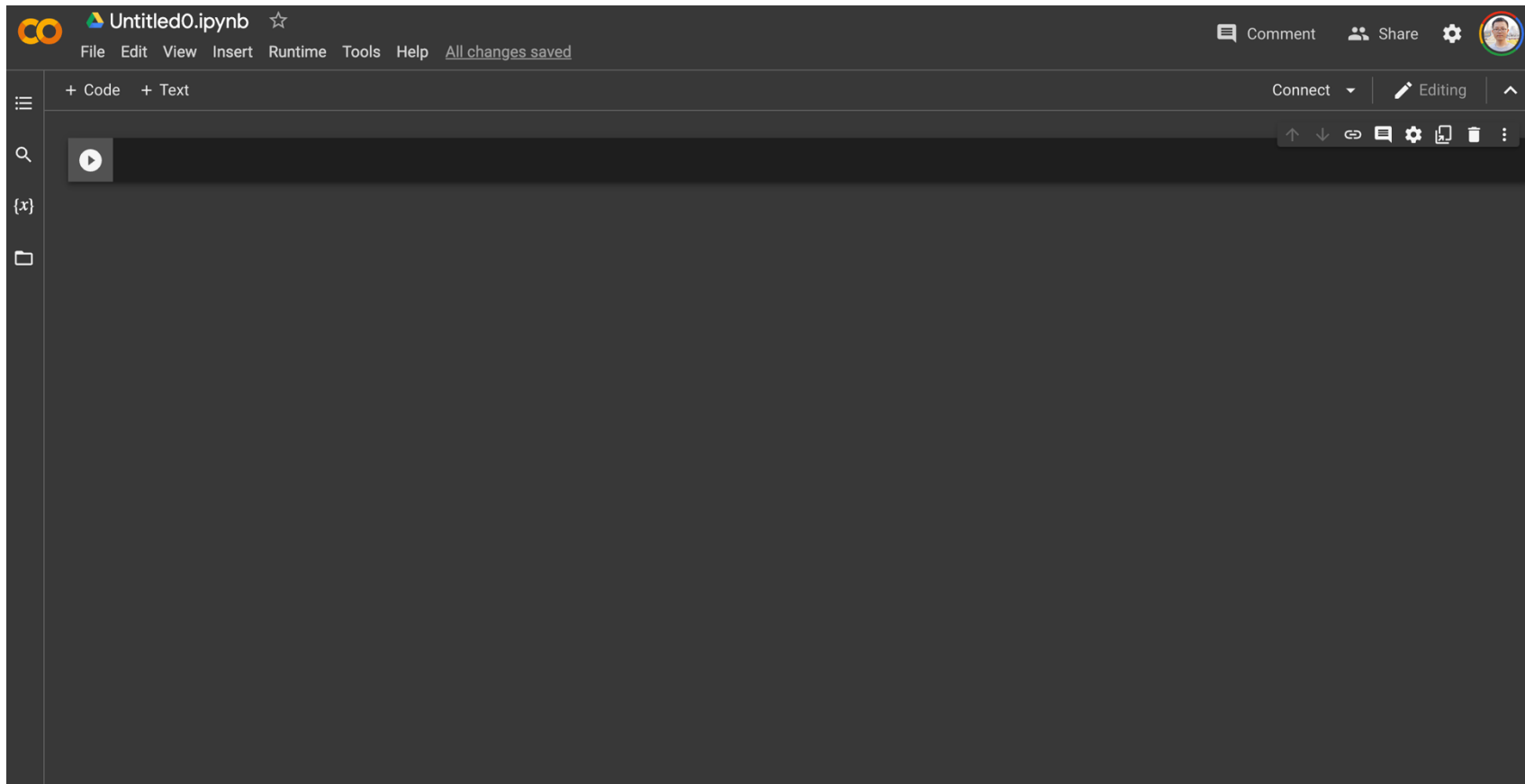
Windows	Mac	Linux
<b>Python 3.11</b> ↓ 64-Bit Graphical Installer (898.6 MB)	<b>Python 3.11</b> ↓ 64-Bit Graphical Installer (610.5 MB) ↓ 64-Bit Command Line Installer (612.1 MB) ↓ 64-Bit (M1) Graphical Installer (643.9 MB) ↓ 64-Bit (M1) Command Line Installer (645.6 MB)	<b>Python 3.11</b> ↓ 64-Bit (x86) Installer (1015.6 MB) ↓ 64-Bit (Power8 and Power9) Installer (473.8 MB) ↓ 64-Bit (AWS Graviton2 / ARM64) Installer (727.4 MB) ↓ 64-bit (Linux on IBM Z & LinuxONE) Installer (340.8 MB)

Windows Start menu applications: Spyder (anaconda3), Anaconda Powershell Prompt (anaconda3), Jupyter Notebook (anaconda3), 1Password, 3D Viewer, 7-Zip, Access, Adobe Acrobat DC, Adobe Acrobat Distiller DC, Adobe Creative Cloud, Adobe Illustrator 2020, Adobe Lightroom Classic, Adobe Media Encoder 2020, Adobe Photoshop 2020.



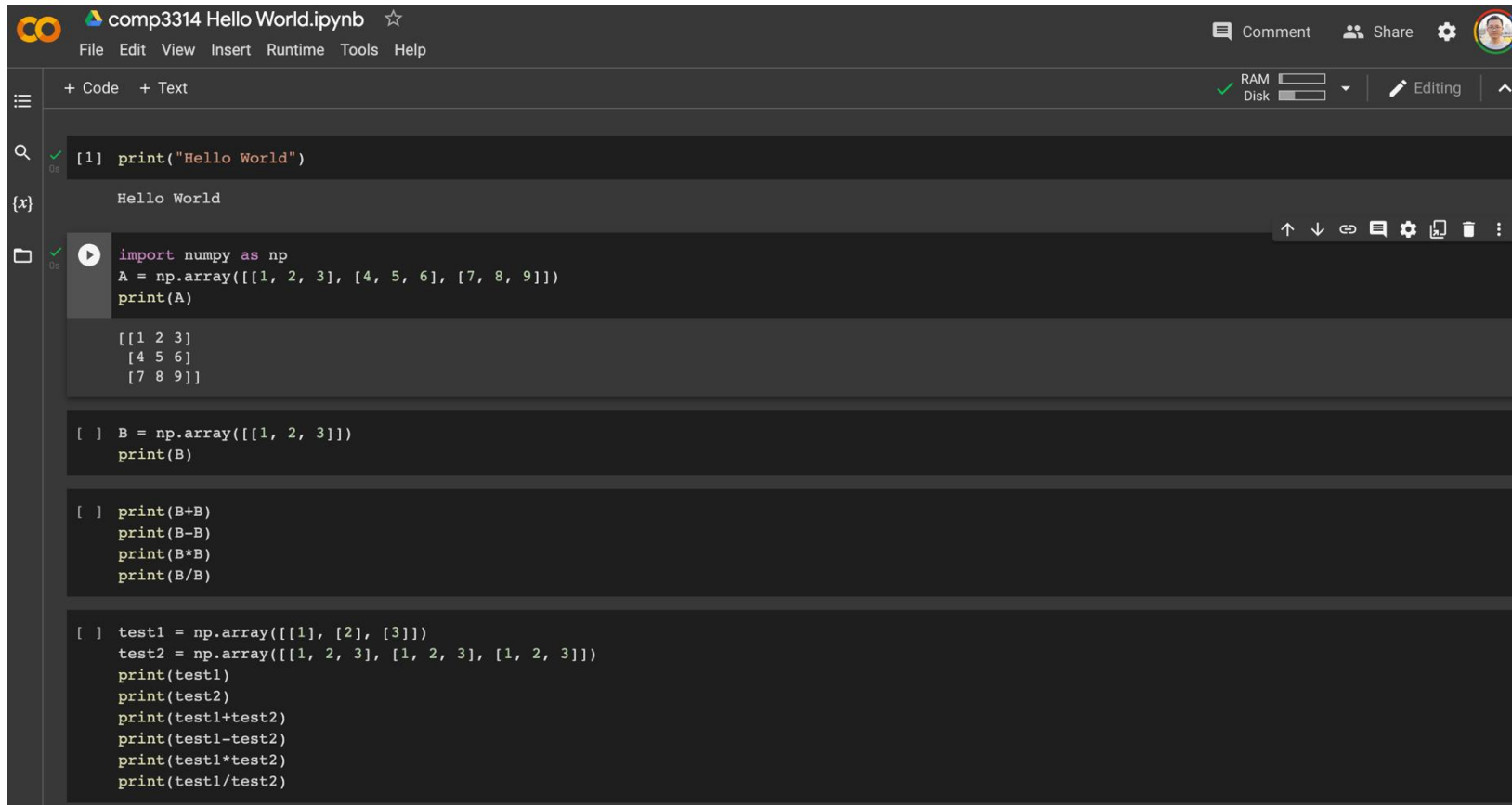
# Programming Environment: Cloud

[Google CoLab](#)



# Installation

→ Code is [here](#) ←



The screenshot shows a Jupyter Notebook titled "comp3314 Hello World.ipynb". The interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with icons for code and text cells. The notebook contains four code cells. The first cell is a simple "Hello World" print statement. The second cell imports NumPy and creates a 3x3 array A. The third cell creates a 1D array B. The fourth cell performs various operations on arrays test1 and test2. The output of the first cell is "Hello World". The output of the second cell is a 3x3 array. The output of the third cell is a 1D array. The output of the fourth cell is a series of array operations and their results.

```
[1] print("Hello World")

Hello World

import numpy as np
A = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
print(A)

[[1 2 3]
 [4 5 6]
 [7 8 9]]

[ ] B = np.array([1, 2, 3])
print(B)

[ ] print(B+B)
print(B-B)
print(B*B)
print(B/B)

[ ] test1 = np.array([1, 2, 3])
test2 = np.array([1, 2, 3, 1, 2, 3, 1, 2, 3])
print(test1)
print(test2)
print(test1+test2)
print(test1-test2)
print(test1*test2)
print(test1/test2)
```

# Linear Algebra Review

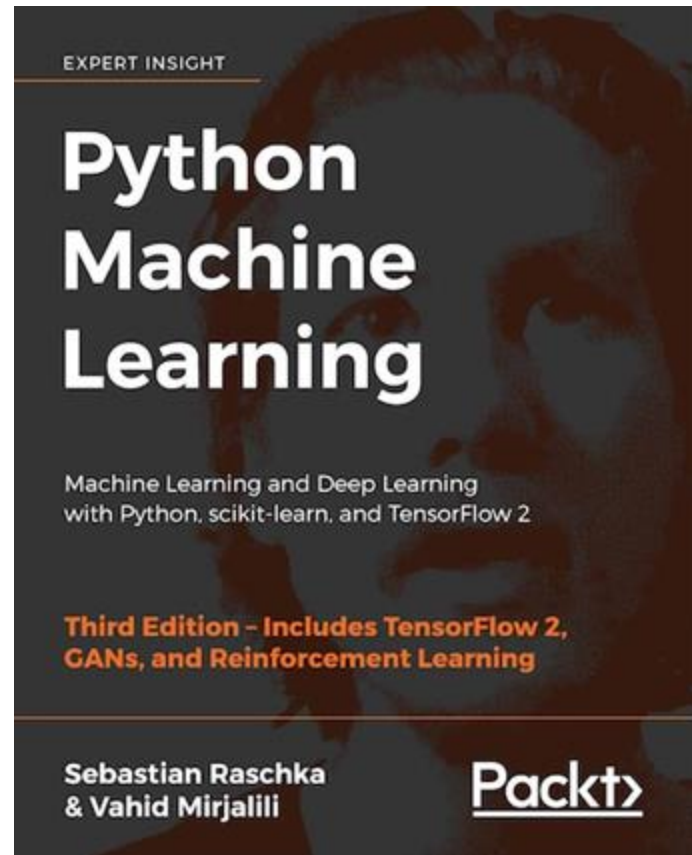
- We will only use basic concepts from linear algebra
- However, if you need a quick refresher, please take a look at Zico Kolter's excellent [videos](#)

# Python Review

- We assume that you are familiar with the libraries/tools, follow these links if you need a refresher
  - [NumPy](#)
  - [Pandas](#)
  - [Matplotlib](#)
  - [Jupyter](#)

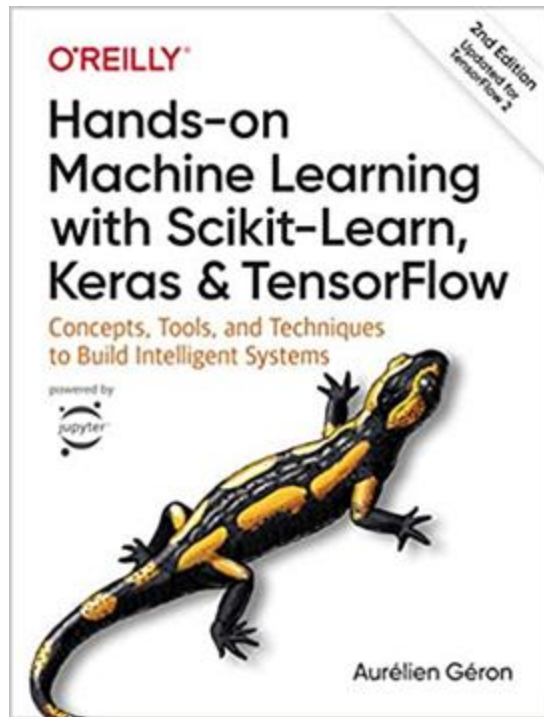
# References

- Materials in this chapter are based on
  - [Book](#)
  - [Code](#)



# References

- Some materials in this chapter are based on
  - [Book](#)
  - [Code](#)



# Exercise 1

- How would you define Machine Learning?
- Can you name four types of problems where it shines?
- What is a labeled training set?
- What are the two most common supervised tasks?
- Can you name four common unsupervised tasks?
- What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?
- What type of algorithm would you use to segment your customers into multiple groups?
- Would you frame the problem of spam detection as a supervised learning problem or an unsupervised learning problem?

# Exercise 2

- What is an online learning system?
- What is out-of-core learning?
- What type of learning algorithm relies on a similarity measure to make predictions?
- What is the difference between a model parameter and a learning algorithm's hyperparameter?
- What do model-based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?
- Can you name four of the main challenges in Machine Learning?



# Exercise 3

- If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name three possible solutions?
- What is a test set, and why would you want to use it?
- What is the purpose of a validation set?
- What can go wrong if you tune hyperparameters using the test set?

# Boston Dynamics

