

# Machine, Data and Learning

# Machine Learning

- Scientific study of algorithms and statistical models that computer systems use
  - To perform a specific task effectively without using explicit instructions
  - Rely on patterns and inference instead.
- Involves
  - Building a **mathematical model** based on sample data, known as "training data" to make predictions or decisions
  - No explicit programming done to perform the task

# Machine Learning

- Term coined around 1960
- Why learn ? Why not just hire enough programmers and code in rules ?
  - Lots of patterns for an activity/event
  - Events can be dynamic
  - **Data** is increasing exponentially
  - **Data** is also in various formats [Text, Audio, Video]
  - Higher quality **data** due to cheaper storage
- Can be broadly classified into three categories
  - Unsupervised, Supervised and Reinforcement learning

# Unsupervised Learning

- Takes a set of data that contains only inputs and finds structure in data E.g., Grouping or Clustering of data points
- **Marketing:** Finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records.
- **Biology:** Classification of plants and animals given their features.
- **Earthquake studies:** Clustering observed earthquake epicenters to identify dangerous zones.
- **World Wide Web:** Clustering weblog data to discover groups of similar access patterns.

# Supervised Learning

- Builds mathematical model using data set that has both inputs and desired outputs E.g., Classification and Regression tasks

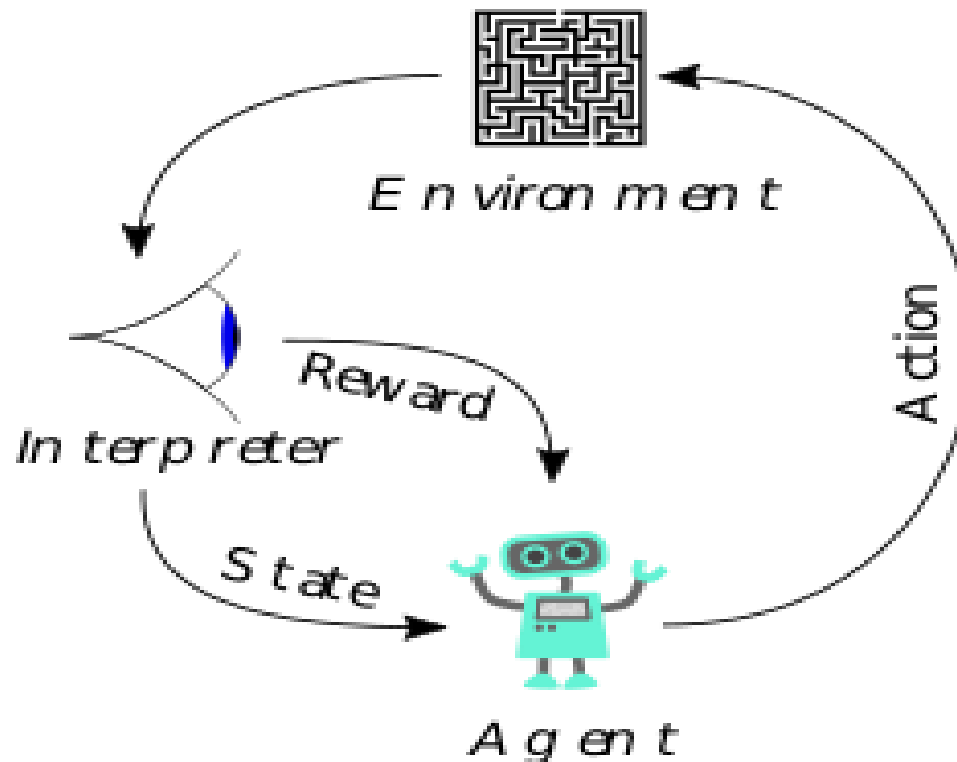
User ID	Gender	Age	Salary	Purchased	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
15624510	Male	19	19000	0	10.69261758	986.882019	54.19337313	195.7150879	3.278597116
15810944	Male	35	20000	1	13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
15668575	Female	26	43000	0	17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
15603246	Female	27	57000	0	20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
15804002	Male	19	76000	1	22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
15728773	Male	27	58000	1	24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
15598044	Female	27	84000	0	24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
15694829	Female	32	150000	1	23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
15600575	Male	25	33000	1	22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
15727311	Female	35	65000	0	20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
15570769	Female	26	80000	1	17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
15606274	Female	26	52000	0	11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
15746139	Male	20	86000	1	14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
15704987	Male	32	18000	0	18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
15628972	Male	18	82000	0	22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
15697686	Male	29	80000	0	24.23155922	988.796875	19.74790765	318.3214111	0.329656571
15733883	Male	47	25000	1					

Figure A: CLASSIFICATION

Figure B: REGRESSION

# Reinforcement Learning

- Concerned with how software agents should take actions in an environment to maximize cumulative reward E.g. Autonomous vehicles, Computer games



# Some Applications

- Search engines
- Information retrieval
- Recommendation systems
- Credit card fraud detection
- Disease diagnosis
- Election prediction
- Image processing
- Speech translation
- ...

# AlphaGo

- First computer Go program to defeat a 9-dan professional player
- Uses Monte Carlo Tree search algorithm based on knowledge learned by a deep learning method
- Beat World No. 1 ranked player in 2017
  - Retired after this match
- <https://deepmind.google/research/breakthroughs/alphago/>
- <https://www.youtube.com/watch?v=WXuK6gekU1Y>
- AlphaGo Zero – Version without human data and stronger than AlphaGo [defeated 100-0]



# AlphaZero & MuZero

- AlphaZero, a generalized version of AlphaGo Zero  
Took 4 hours to learn Chess and defeat reigning world computer chess champion 28 to 0 in 100 matches
- [https://www.youtube.com/watch?time\\_continue=7&v=tXIM99xPQC8](https://www.youtube.com/watch?time_continue=7&v=tXIM99xPQC8)
- MuZero: Master games without knowing rules
- Uses approach similar to AlphaZero, developed in 2019
- Trained via self-play and play against AlphaZero with no access to rules, opening books or endgame tables
- Viewed as significant advancement over AlphaZero

# AlphaFold: solution to a 50 year old grand challenge in biology

- <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>
- Figuring out what shapes proteins fold into is known as the “protein folding problem” - grand challenge in biology for the past 50 years
- Focus of intensive scientific research for many years, using a variety of experimental techniques such as nuclear magnetic resonance and X-ray crystallography.

# AlphaFold

- Number of ways a protein could theoretically fold before settling into its final 3D structure is astronomical.
- Cyrus Levinthal estimated  $10^{300}$  possible conformations for a typical protein.
- Estimated would take longer than the age of universe to enumerate all possible configurations. Yet in nature, proteins fold spontaneously, some within milliseconds - referred to as Levinthal's paradox.

# Nobel Prize in Chemistry 2024 !!!

- The Nobel Prize in Chemistry 2024 is about proteins, life's ingenious chemical tools.
- David Baker has succeeded with the almost impossible feat of building entirely new kinds of proteins.
- Demis Hassabis and John Jumper have developed an **AI model** to solve a 50-year-old problem: predicting proteins' complex structures. These discoveries hold enormous potential.

# Nobel Prize in Physics 2024 !!!

- They used physics to find patterns in information
- John J Hopfield and Geoffrey Hinton
- “for foundational discoveries and inventions that enable machine learning with artificial neural networks”

# Year 2023 in review for AI

- Article titled 2023: The Crazy AI Year by Nisha Arya, KDnuggets
- Many media sources claim year 2023 can be considered the year of AI
- Jan:
  - With huge buzz around ChatGPT Microsoft announced \$10 billion funding in OpenAI
- Feb:
  - Google came up with BARD. Microsoft came up with its Bing chatbot

# Year 2023 in review for AI

- Mar:
  - Access to Bard was given to a limited number of people to kickstart the Google GenAI journey.
  - Initiated a domino effect with Adobe introducing Firefly and Canva introducing their virtual design assistant.
  - OpenAI also launched APIs for ChatGPT, as well as their text-to-speech model called Whisper. On the 14th of March, OpenAI released its most advanced model GPT-4.

# Year 2023 in review for AI

- Apr:
  - Announcement of Google DeepMind - a combination of Google Research and DeepMind.
  - Russia's Sberbank released ChatGPT rival GigaChat
  - HuggingFace also entering the market with the release of an AI chatbot to rival ChatGPT called HuggingChat
- May:
  - Google announced the Bard chatbot to the public - added some fuel to the GenAI fire with Microsoft revealing its debut AI assistant for Windows 11.



# Year 2023 in review for AI

- Market capitalization of NVIDIA topped \$1 trillion for the first time, holding its status as the AI chip leader.
- Elon Musk's new brain implant startup, called Neuralink, in which the company aims to create and implant AI-powered chips in people's brains. This was approved by the FDA for human trials.
- Jun:
  - Apple's Vision Pro, the AI-powered augmented reality headset was developed to take immersive experiences to the next level.
  - European Parliament made some negotiations about the EU AI Act, with 499 votes in favor, 28 against, and 93 abstentions.

# Year 2023 in review for AI

- McKinsey predicted that GenAI has the potential to add up to \$4.4 trillion in value to the global economy.
- July:
  - Meta introduced Llama 2, an open-source Large Language Model (LLM) which was trained on a mix of publicly available data, and designed to drive applications such as OpenAI's ChatGPT, Bing Chat, and other modern chatbots.
  - Anthropic also released Claude 2, which dethroned ChatGPT and has it shaking in its boots.
  - The safety around AI is becoming a popular topic as LLMs are becoming a part of our day-to-day lives.

# Year 2023 in review for AI

- Microsoft announced that it will charge customers \$30 per month to use Microsoft 365 Copilot.
- Aug:
  - Google said that it would also be charging \$30 per month for users to make use of their GenAI tools in their Duet AI for Workspace.
  - OpenAI introduced custom instructions to get the most out of ChatGPT. Poe a chatbot service that allows you to use state-of-the-art models such as Claude +, GPT-3.5-Turbo, and GPT-4.

# Year 2023 in review for AI

- Sept:
  - Amazon announced a \$4 billion investment in OpenAI competitor Anthropic.
  - OpenAI continues with its quest to visualize content with a Canva plugin for ChatGPT.
- Oct:
  - We experienced the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. This also was shaking up the AI world, with CEOs, leaders, and others having contradicting opinions about the implementation of AI systems into society.

# Year 2023 in review for AI

- Nov:
  - Elon Musk's AI startup, xAI, unveiled the AI chatbot "Grok", AWS with the release of Amazon Q, and Pika 1.0 from StabilityAI.
  - OpenAI also held its first developer event in November, where it delved into GPT-4 Turbo and the GPT Store.
  - OpenAI's CEO Sam Altman getting fired by the board out of nowhere. He was immediately offered a job by Microsoft with OpenAI employees threatening to resign if Sam Altman did not come back and claim his position as CEO. So now he is back, with some new board members as well as a new "observer" role for Microsoft.

