HARLENE GILL

Machine Learning in the World

A presentation on the real world applications of Machine Learning

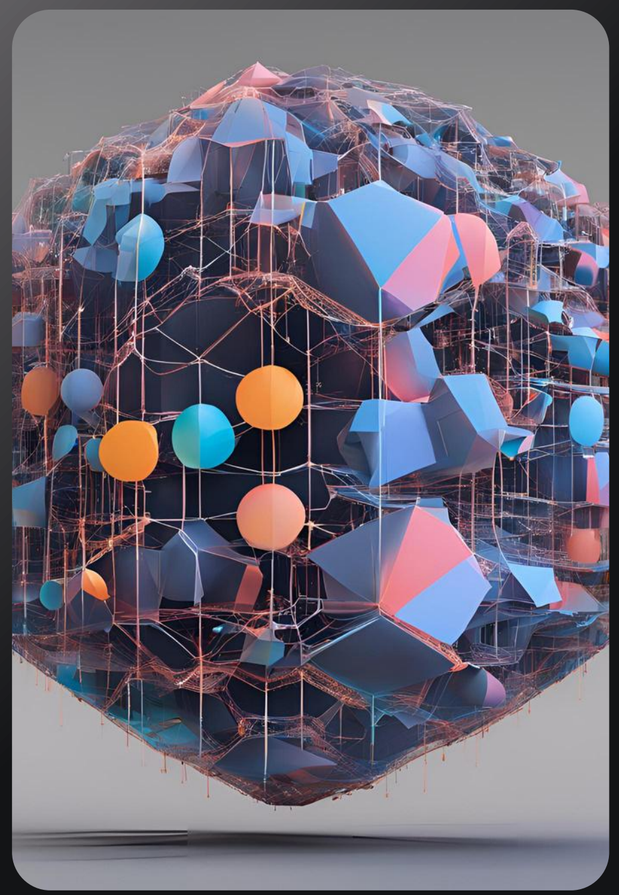
Canva Link:
https://www.canva.com/design/DAGkoHIRB7g/kgWPekyGnvZQO4RYk6ZCzQ/edit?utm_content=DAGkoHIRB7g&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

What is Machine Learning

We are now in “Industry 5.0” - the age of AI...

Machine learning is a branch of AI that enables computers to learn patterns from data and make decisions without being explicitly programmed.

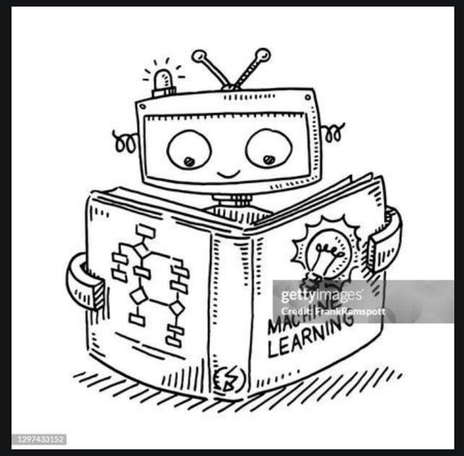
It uses algorithms to identify relationships, trends, or behaviors based on examples. As the model processes more data, it improves its accuracy and performance over time.



Types of Machine Learning

Supervised Learning

Learning to classify objects by training on labeled data that defines known categories.



Unsupervised Learning

Grouping and classifying objects based on similarities in their features without using labeled data.

Reinforcement Learning

Learning through a cycle of actions and feedback, where rewards guide behavior within an environment.

ML in Industry

1. Biology - AlphaFold (Unsupervised Learning)
2. Cosmology - Cosmoflow (Supervised Learning)
3. Robotics - Boston Dynamics (Reinforcement Learning)

Google

- **Company started 1998 Sergey Brin & Larry Page**

DeepMind - AlphaFold

An AI start up founded in 2010 as DeepMind. Merged with Google in 2014 to create Google DeepMind and created AlphaFold in 2020.

DeepMind has created other technologies such as AlphaGo (a program that plays the board game “Go”).

Build to advance research in the field of molecular biology but has implications in numerous industries.

TensorFlow - CosmoFlow

TensorFlow founded by Google on the Apache 2.0 license in 2015. CosmoFlow built in 2018.

CosmoFlow is built on TensorFlow to propel innovation within the field of Cosmology.

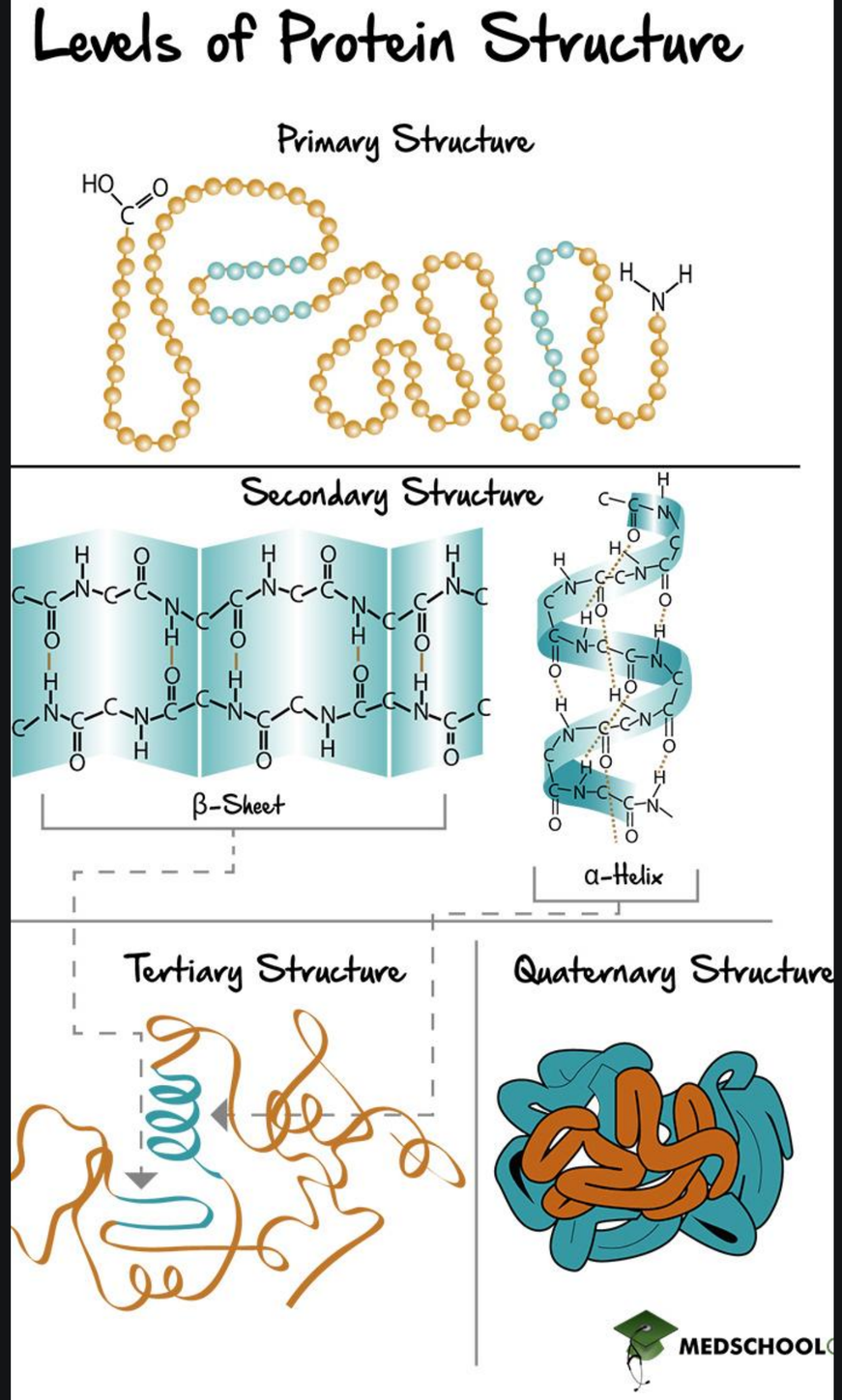
Although, created by academics its logic provides use for applications involving large amounts of 3D spatial data.

- CosmoFlow was the first large scale application of TensorFlow to efficiently run at supercomputer scale with fully synchronous training and to process 3D spatial volumes at this scale

Google DeepMind

TensorFlow

AlphaFold



1 Purpose

To predict protein structures (ie. solve the protein folding problem) using a Graph Neural Network (GNN).

2 Importance

Predicting protein structures is one of grand challenges of biology. Protein takes a unique 3D shape which takes an immense computational undertaking to map out. Scientists have only been able to map out structures for 180,000 proteins over 60 years. Each protein has a different function within an organism.

3 Achievment

AlphaFold has been able to map 200 million predictions to date.

Nobel Prize winner 2024 in chemistry.

AlphaFold

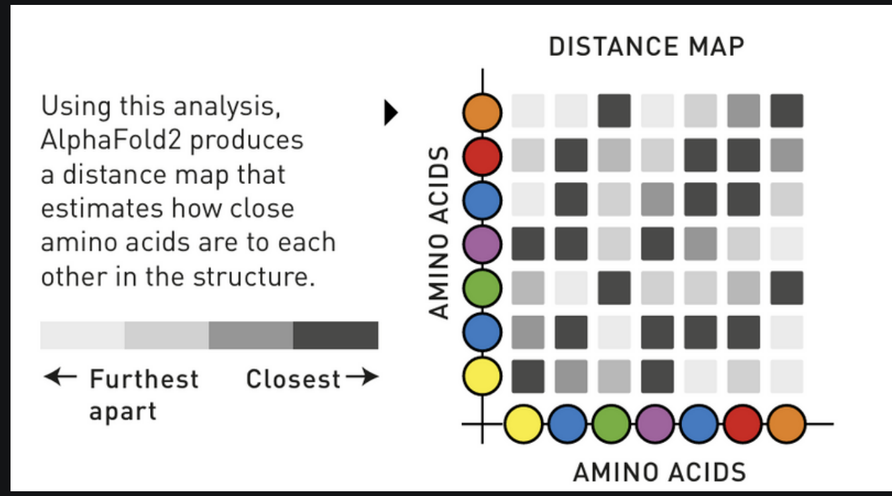
Interacting AA's co-evolve!

4 What is a Graph Neural Network (GNN)?

A deep learning algorithm that works on graph data structures:

- Vertices = Amino Acids (AA's) in a protein sequence
- Edges = potential interactions between the AA's such as spatial proximity, chemical bonds, or co-evolutionary relationships

The GNN learns patterns by passing messages along the edges, allowing each amino acid (node) to aggregate information from its neighbors and update its own representation accordingly.



5 AlphaFold Model

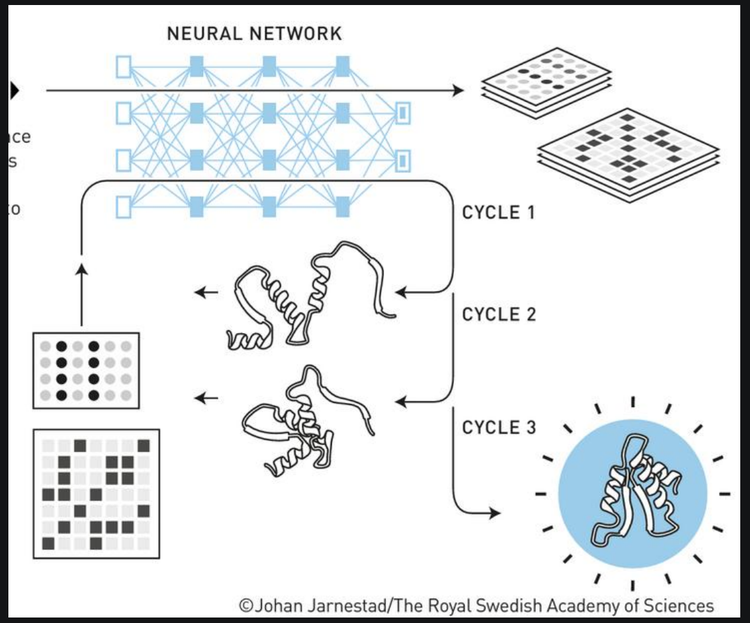
Model Flow

- Input: Amino acid (AA) sequence with unknown structure
- Supervised Learning: Aligns similar AA sequences from different species using known structures from the Protein Data Bank (PDB) to learn evolutionarily conserved relationships - uses Distance Map
- Unsupervised Learning: Processes Multiple Sequence Alignments (MSAs) to uncover co-evolutionary patterns (e.g., AA pairs that mutate together and likely interact in 3D space), without needing explicit structure labels

AlphaFold

5 AlphaFold Model Flow Cont.

- Transformers (type of neural network) identify which parts of the sequence may influence each other; used together with results from above in an iterative process
  - Evoformer: Specialized transformer that parallelly processes MSA and pairwise features, enabling message passing between AAs and their edges
  - Edges carry and update features, such as spatial closeness or evolutionary correlation.
- Building Structure: Iteratively (3 times) assembles a 3D structure of protein
- Output: The final predicted structure is compared to known data (more supervised training)

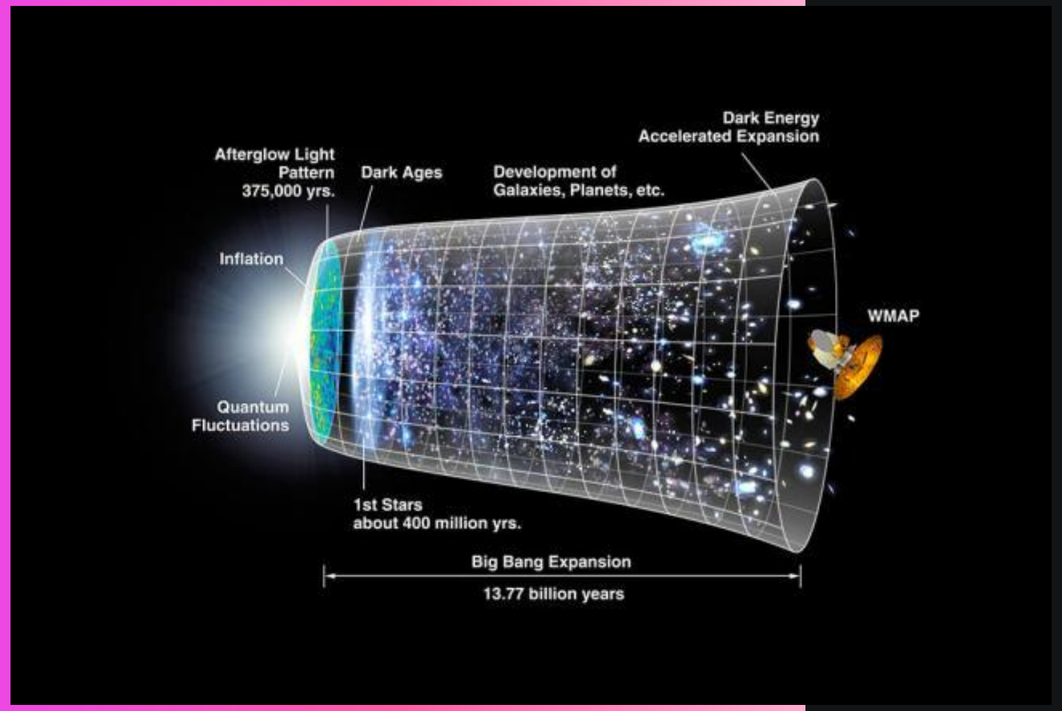


Open Source Dataset
https://alphafold.com/download#full-dataset-section

6 Industry

- Breaking down plastic enzymes to achieve 100% recyclability - Climate and Chemistry
- Tracing proteins back to the origins of life - Life Sciences
- Designing more efficient drugs - Medicine

https://deepmind.google/technologies/alphafold/impact-stories/



CosmoFlow

- **Purpose**

  To predict cosmological constants from 3D simulations of our universe using Convolutional Neural Networks (CNN).

  Constants predicted:

  - Matter density parameter : Represents the fraction of the universe’s total energy density made up of matter (dark matter + normal matter)
  - Amplitude of Density Fluctuations : Indicates how clustered matter is in the universe
  - Spectral Index : Provides info about the initial distribution of matter fluctuations from the early universe

- **Importance**

  Predicting cosmological constants helps us understand the nature of the universe and how features such as as these are evolving:

  - Galaxy formation
  - Dark matter distribution
  - Universal acceleration rates (which is propelled by dark matter)

CosmoFlow

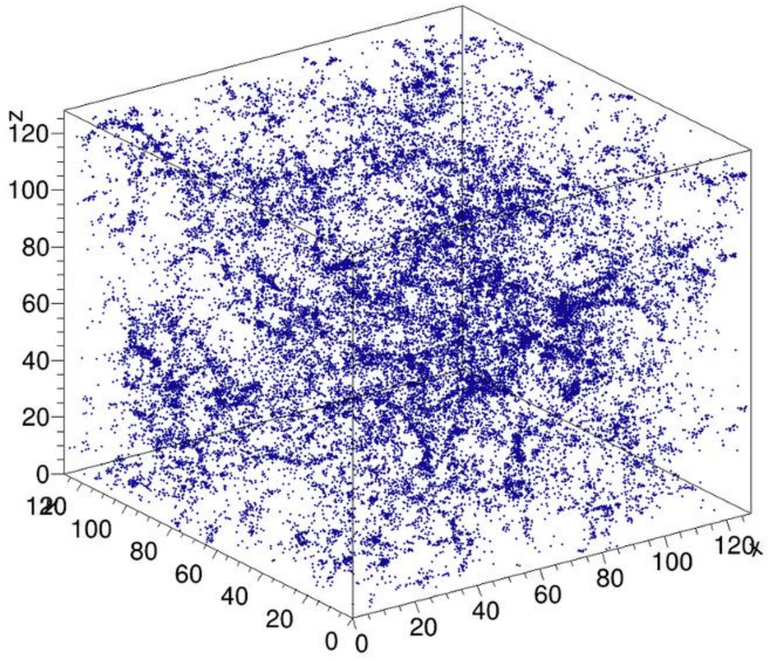


Fig. 1. Example simulation $256Mpc/h^3$, 128^3 voxel sub-volume, used as input to the CosmoFlow network. This sub-volume is taken from the full $512Mpc/h^3$ simulation of dark matter in the universe, evolved over 3 billion years to a redshift of 0 (i.e. today).

Open Source CosmoFlow Data: https://portal.nersc.gov/project/m3363/

What is a Convolutional Neural Network?

CNN is a type of deep learning neural network that recognizes patterns directly from the pixels, or in our case voxels, of an image/video/3D structure. There are 3 layers/steps that combine to make final predictions of the classification or regression results:

- Convolutional Layer
- Pooling Layer
- Fully Connected Layer

The layers/steps combine iteratively (ie. classifies smaller structures such as dark matter filaments and combines them into the bigger structures such as the cosmic web).

CosmoFlow Model

Model Input:

- 3D volumetric data of dark matter computer simulations
- Models are based off the rules of physics and predict how the universe naturally clusters and evolves given specific cosmological parameters

Supervised Leaning:

- Training simulation data is provided with labels for the cosmological parameters

Output:

- Accurate predictions on the cosmological parameters for given simulations

Boston Dynamics

Founded in 1992 by Mark Raibert as an extension from the MIT Leg Lab robotics group. Now owned by Hyundai (2021).

The vision was to develop robots that ran and maneuvered like animals and could “go where people go”. This opened doors for organizations worldwide such as the US army and NASA to work with BD to develop specialized robots that could traverse difficult environments.

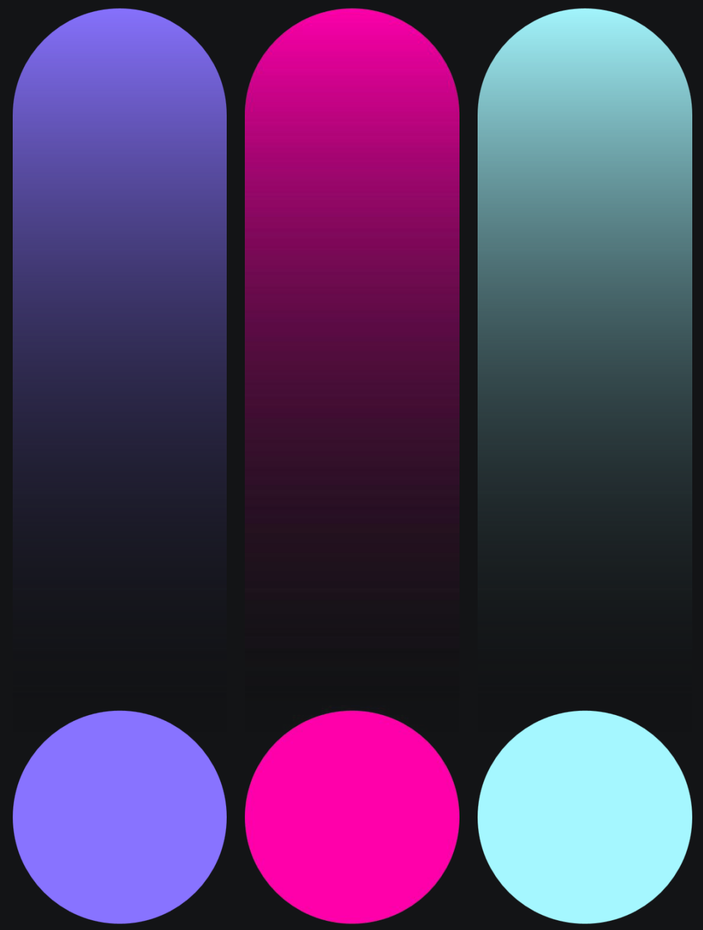
These robots have applications in

- Autonomous exploration
  - Geology - mining operations
  - NASA - planetary exploration
  - Factory exploration

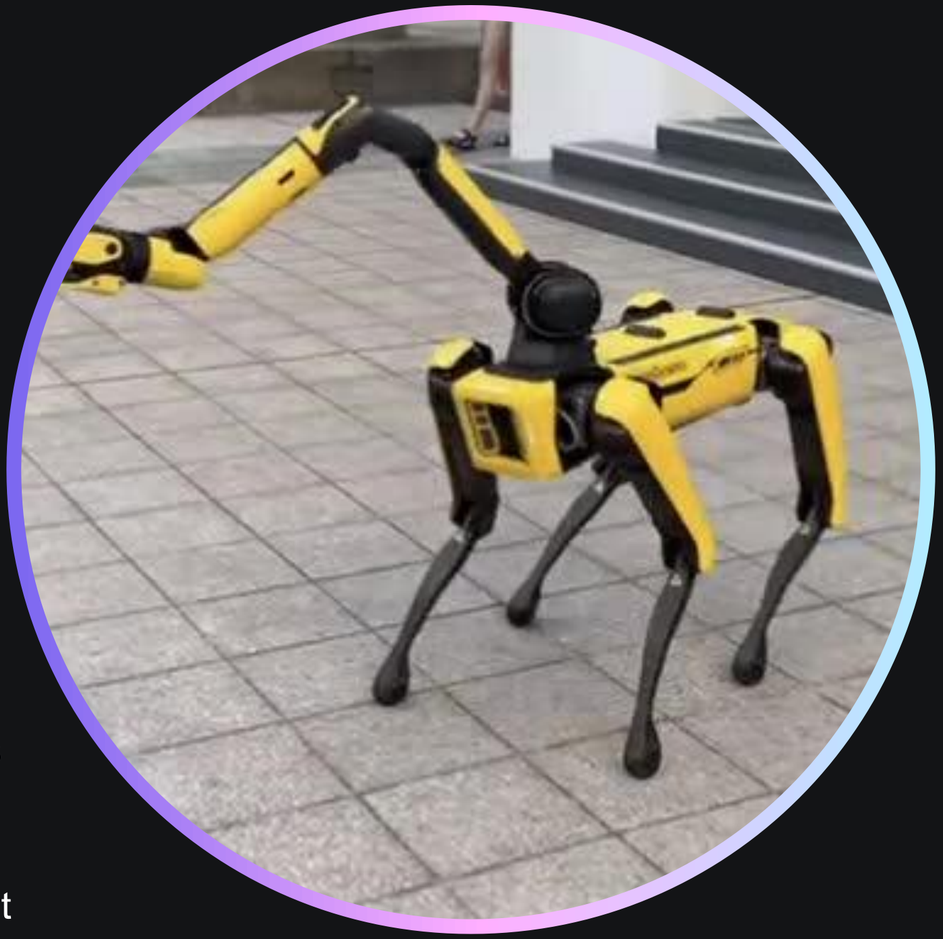
In 2020 SPOT become commercially available.

BD has pledged to safely use their robotics along with maintaining policies to ensure their company upholds these values.

Fun Fact: Google owned Boston Dynamics in 2013 and sold due to lack of ability to develop revenue streams.



SPOT

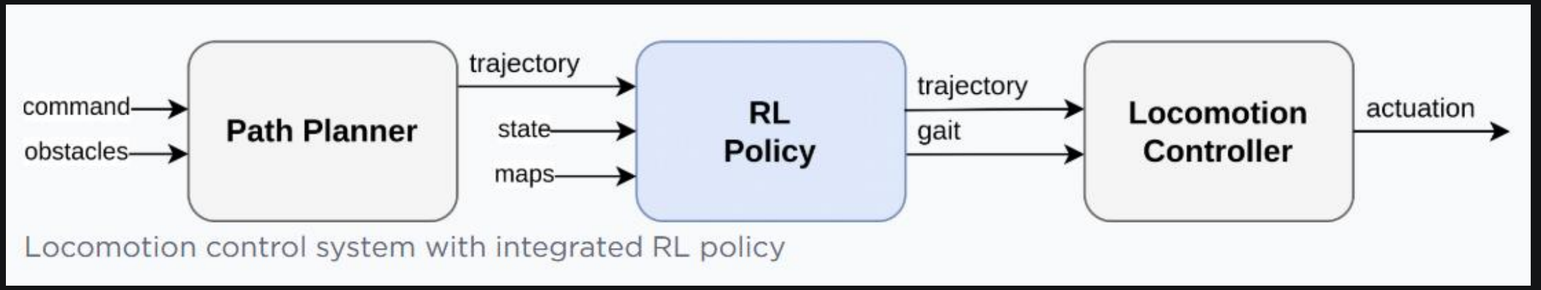


Traditional Control Approach (Robotic Engineering)

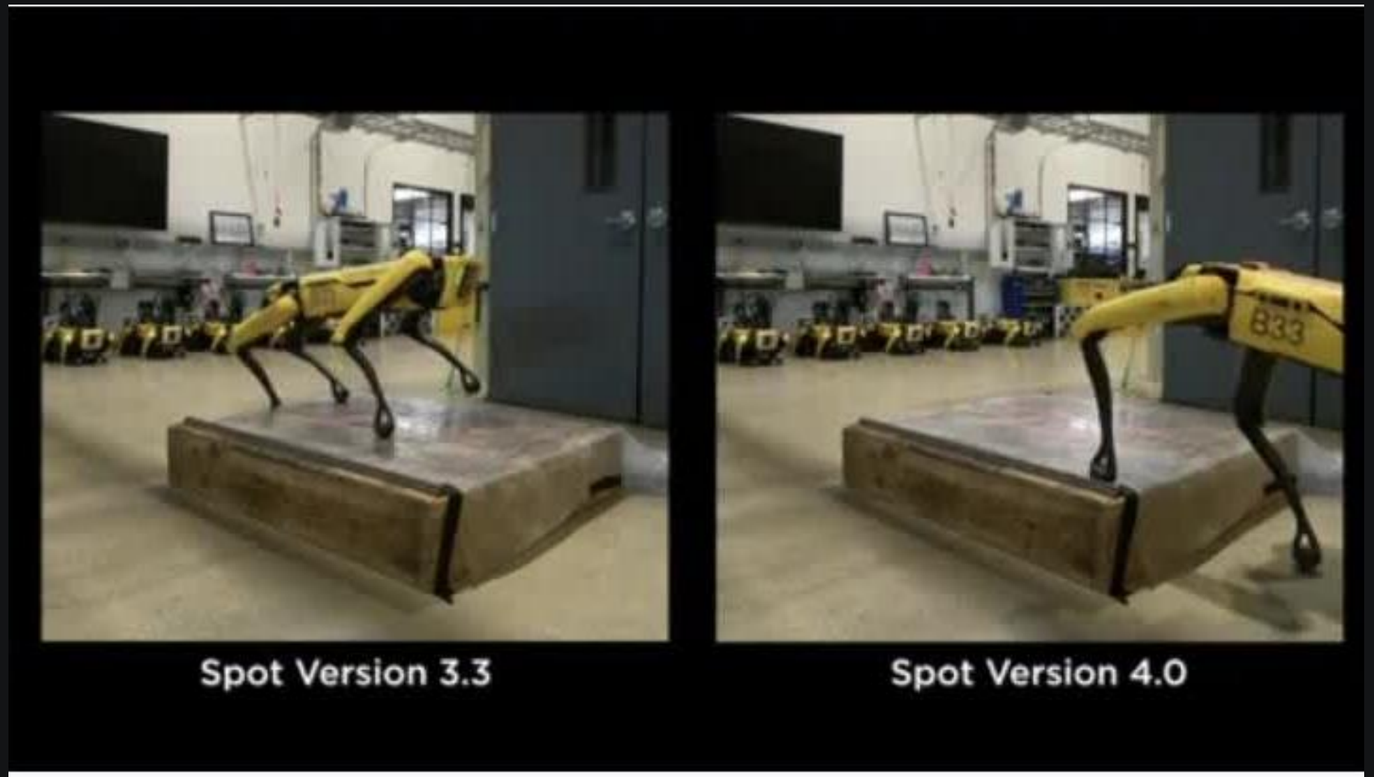
- Spot initially relied on Model Predictive Control (MPC), evaluating multiple predictive models in parallel to determine optimal movements
- While effective in controlled environments, MPC faced challenges with computational demands and adaptability to unpredictable real-world terrains

Reinforcement Learning (RL)

- Learns optimal behaviors through trial and error in simulated environments, improving its ability to navigate complex terrains
- Policies are trained in simulation through neural networks
- Defines reward functions that encourage desired behaviors, such as stability and efficient navigation.
- Trained policies are rigorously tested in both simulations and real-world scenarios to ensure reliability



Locomotion control system with integrated RL policy



Before and after performance using Spot Version 4.0

Thank You

References

- https://www.databricks.com/glossary/tensorflow-guide
- https://www.gizmochina.com/2023/01/19/boston-dynamics-atlas-robot-demo/
- https://bostondynamics.com/case-studies/nasa-jpl-search-for-life/
- https://rwisynthetics.com/
- https://deepmind.google/technologies/alphafold/
- https://www.maginative.com/article/google-deepmind-and-isomorphic-labs-announce-alphafold-3/
- https://www.ibm.com/think/topics/convolutional-neural-networks
- https://www.cfa.harvard.edu/research/topic/dark-energy-and-dark-matter
- https://lambda.gsfc.nasa.gov/education/graphic_history/fluctsize.html
- image--- https://webbtelescope.org/contents/media/images/2022/035/01G7DCWB7137MYJ05CSH1Q5Z1Z
- https://arxiv.org/abs/1808.04728
- https://www.dreamstime.com/chemical-molecules-bonds-dna-chain-white-flat-organic-inorganic-isolated-compounds-particles-research-chemistry-image189411737
- https://www.nobelprize.org/uploads/2024/10/fig2_ke_en_24.pdf
- https://www.nature.com/articles/s41586-021-03819-2
- https://seon.io/resources/dictionary/graph-neural-network-gnn/#:~:text=By%20analyzing%20nodes%20and%20edges,industry%20sectors%20and%20online%20systems.
- https://www.theverge.com/2020/2/19/21144648/boston-dynamics-spot-robot-mass-state-police-trial-issues
- https://www.crunchbase.com/organization/boston-dynamics
- https://www.latimes.com/business/technology/la-fi-softbank-20170609-story.html
- https://www.youtube.com/shorts/bH49PyGmIDo
- https://bostondynamics.com/video/reinforcement-learning-with-spot/
- https://bostondynamics.com/blog/starting-on-the-right-foot-with-reinforcement-learning/
- https://medium.com/geeks-for-tech/artificial-intelligence-machine-learning-in-industry-5-0-ba2945e5d54e