



By Tanujit Chakraborty



# Overview



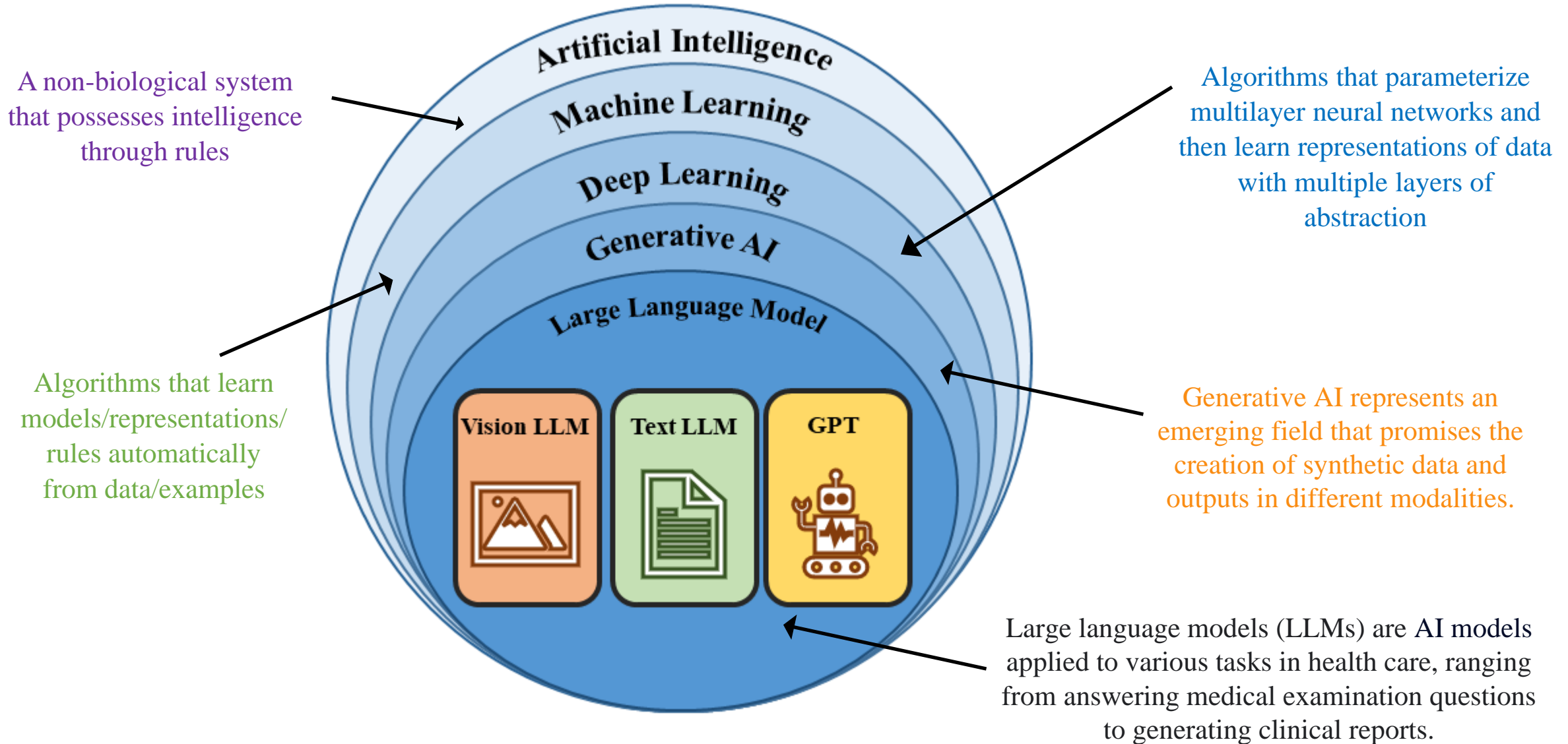
- Artificial Intelligence vs. Machine Learning vs. Data Science
- How to develop intelligent machines?
- AI timeline
  - DL success in computer vision
  - DL success in natural language processing
  - Generative text-to-image models
  - Foundation models
- AI limitations and challenges
- Prospective Trends in AI

# Artificial Intelligence



- **Artificial Intelligence (AI)** is a scientific field concerned with the development of algorithms that allow computers to reason or learn without being explicitly programmed
  - AI is opposite to **natural intelligence** displayed by humans and animals
- AI as an academic discipline was founded in 1956
- AI studies theories and technologies related to:
  - Planning and reasoning
  - Knowledge representation
  - Machine learning
  - Natural language processing
  - Computer Vision
  - Large Language Models
  - Perception
  - Motion and manipulation

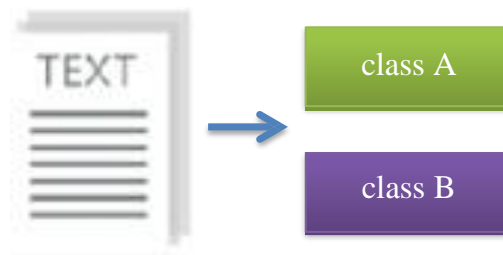
# Family of AI



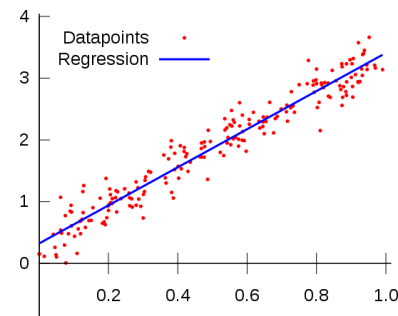
# Machine Learning



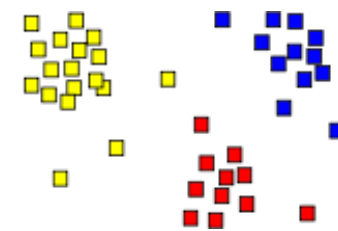
- **Machine Learning** is a subfield of Artificial Intelligence, that focuses on methods that learn from data and make predictions on unseen data
- Categories of ML approaches
  - **Supervised learning**: learning with **labeled data**
    - Example: image classification, email classification
    - Example: regression for predicting real-valued outputs
  - **Unsupervised learning**: discover patterns in **unlabeled data**
    - Example: cluster similar data points
  - **Reinforcement learning**: learn to act based on **feedback/reward**
    - Example: learn to play Go



Classification

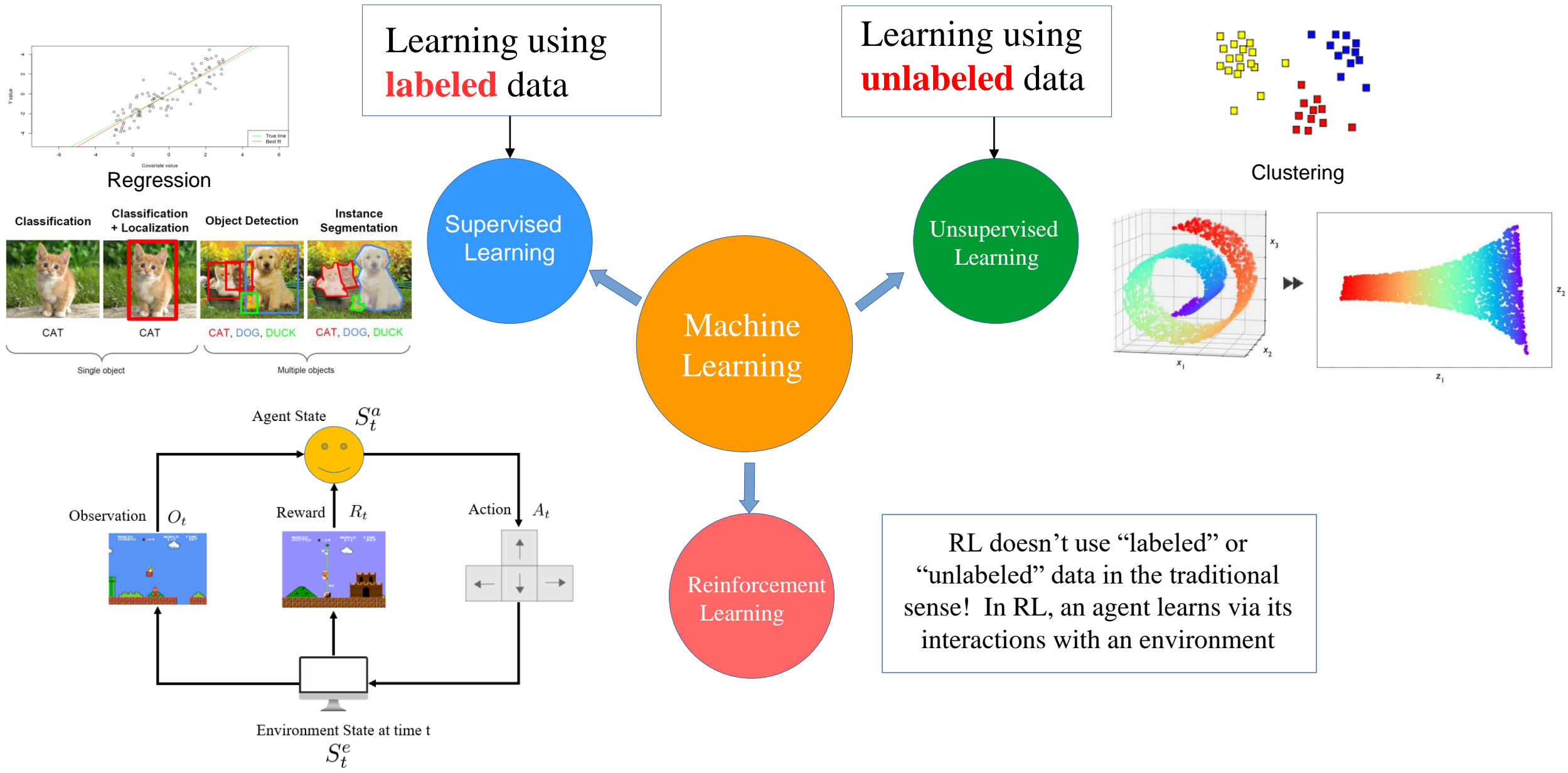


Regression



Clustering

# A Loose Taxonomy of ML

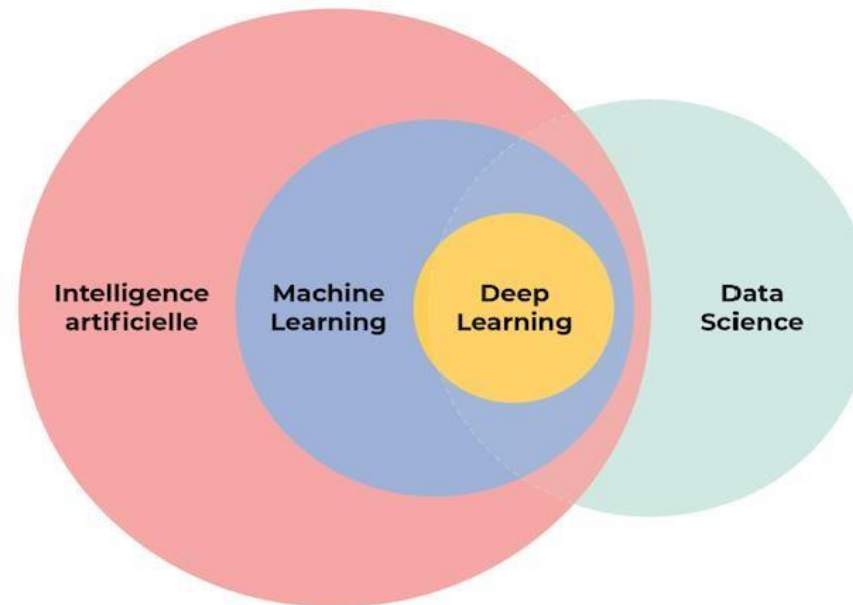


# Data Science



- **Data Science (DS)** is an interdisciplinary field that uses scientific methods and algorithms to extract knowledge from data, and applies the insights to application domains, such as to make business decisions
- Data Science versus Machine Learning
  - DS focuses on extracting knowledge and insights from data
  - DS can rely on ML approaches, but it can also obtain insights via statistical analysis, data cleaning, data visualization, exploratory data analysis, feature engineering

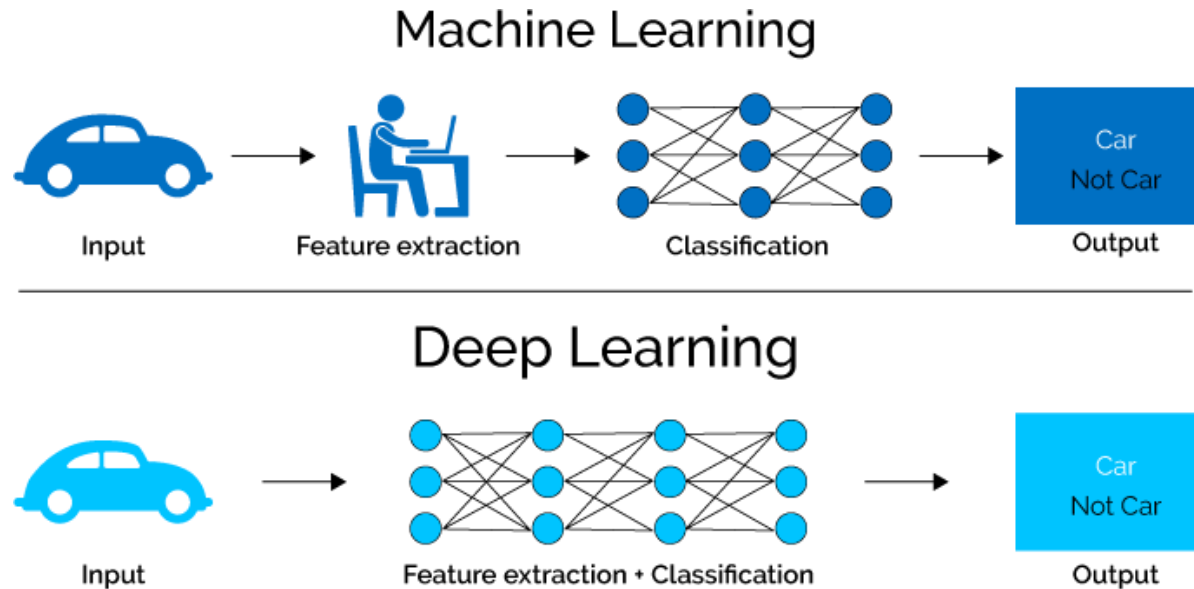
## AI vs. ML vs. DS



# Deep Learning



- **Deep Learning** (DL) is a subarea in machine learning that uses **artificial neural networks** (ANNs) with multiple layers for learning data representations
  - A major advantage of DL is the ability to automatically extract features in data
  - The most common architectures in deep ANNs are: multi-layer perceptron NNs, convolutional NNs, recurrent NNs (LSTM, GRU), graph NNs, transformer NNs





# What is Intelligence?



- An **intelligent agent** is any system that perceives the environment and takes actions to maximize the chances of achieving its goals
  - Goals can vary, e.g., human goals can be to make a coffee, build a wall, solve a math problem, drive a car, cook a meal, etc.
- Definition: *Intelligence* is an agent's ability to achieve goals in a wide range of environments
- Intelligent agents should be able to acquire and retain knowledge, and use it to respond effectively to new tasks or act in new situations and environments
  - E.g., more intelligent humans should be able to solve many physics problems that they haven't seen before (e.g., think Einstein)
  - Intelligence encompasses many related abilities for:
    - Reasoning and rational thinking, comprehend ideas, apply planning, problem-solving
    - Learning and adaptation, deal with unexpected situations and uncertainties
    - Interacting with the real world to perceive, understand, and act

# How to Develop Intelligent Machines?

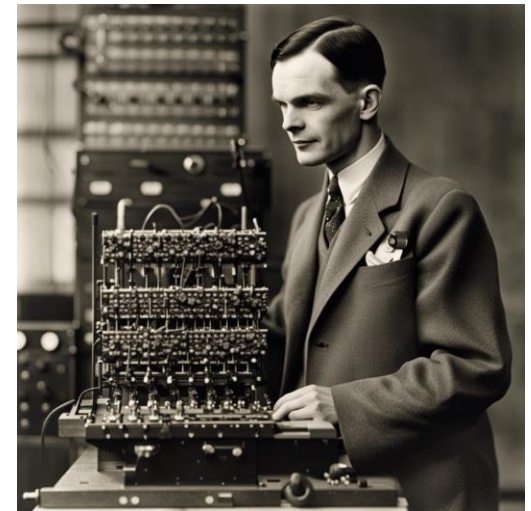
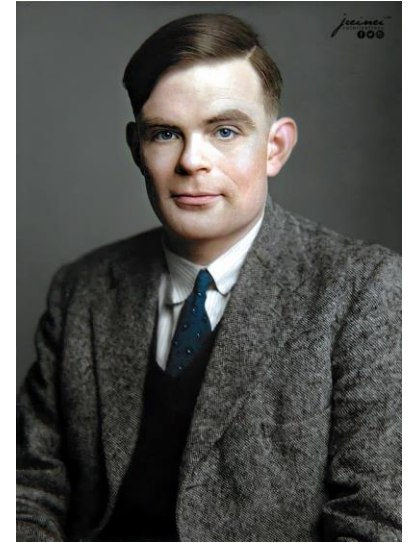


- AI scientists in 1950s believed that machines with human-level intelligence can be achieved within 10 to 20 years
- Initial AI approaches
  - Imitate step-by-step reasoning that humans use to solve a problem
  - Create a knowledge database based on human domain knowledge about a task, and develop an inference engine to process the states and make decisions
  - Challenges: handling uncertainties, combinatorial explosion (the space of solutions quickly becomes too large for most problems)
- These approaches failed to deliver, as the scientists underestimated the complexity of human intelligence
- Various **misconceptions** about intelligence has perpetuated in the AI field
  - E.g., computers can process information -> human thinking is similar to logic processing -> encoding human thinking into a program can lead to intelligent machines
  - E.g., chess is a game of intellect and chess players are very intelligent people -> developing computers that can reason and play chess at a human expert level can lead to machines with human-level intelligence

# Weak vs. Strong AI



- AI systems can be classified into weak AI and strong AI systems
- **Weak AI**, or **narrow AI**: can solve one specific task
  - E.g., image classification ML models
  - E.g., Deep Blue computer that defeated the world chess champion
- **Strong AI**, of **artificial general intelligence (AGI)**: can solve a variety of tasks
  - AGI is the ability to understand or learn any intellectual task that a human being can
    - AGI performance would be indistinguishable from that of humans
  - At present, AGI systems do not exist
- How to evaluate AI?
  - **Turing test**, proposed by Alan Turing in 1950
    - “A computer would deserve to be called intelligent if it could deceive a human into believing that it was human”
    - Test: a human interacts with other humans and an AI agent; the test is passed if the human cannot distinguish the AI agent from the humans
    - Turing called the test “Imitation Game”
    - The test has not been passed yet by an AI system

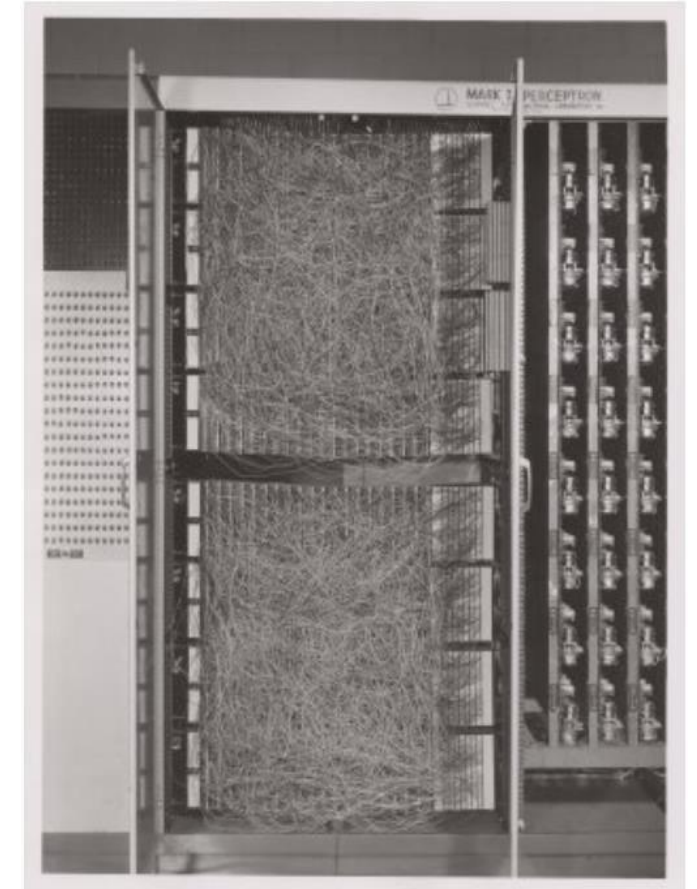


[Alan Turing with an enigma machine - AI Generated Artwork](#)

# AI Timeline



- 1943 – The first model of a simple artificial neuron proposed
- 1950 – Alan Turing introduced the **Turing test**
- 1955 – The **term Artificial Intelligence** used for the first time
- **1956 – Workshop on AI held in Dartmouth College**, New Hampshire, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, Claude Shannon
  - “Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.”
  - Official beginning of AI as academic discipline
- 1958 – **Perceptron** algorithm proposed by Rosenblatt
  - Shown is the Mark I Perceptron computer, used for implementing the algorithm



# AI Timeline



- 1966 – **Eliza**, a chatbot that simulates conversations with a psychotherapist
- 1970-1980 – First AI winter, agencies reduced funding for AI projects due to unsatisfactory progress
- 1982 – An expert systems deployed for configuring computer orders
- 1987-1992 – Second **AI winter**, DARPA cut AI funding for expert systems
- 1995 – The advent of machine learning and statistical methods
- 1997 – IBM's supercomputer **Deep Blue** won against world chess champion Gary Kasparov
- 2011 – IBM's supercomputer **Watson** won against two human rivals in the quiz show Jeopardy
- 2012 – **Deep NN model AlexNet** won image classification contest - *beginning of the era of deep learning*
- 2014: China's **Tianhe-2** – fastest system, a 3.86-petaflops supercomputer located in the National Supercomputer Center in Guangzhou, China
- 2015 – **GAN** (Generative Adversarial Network) introduced

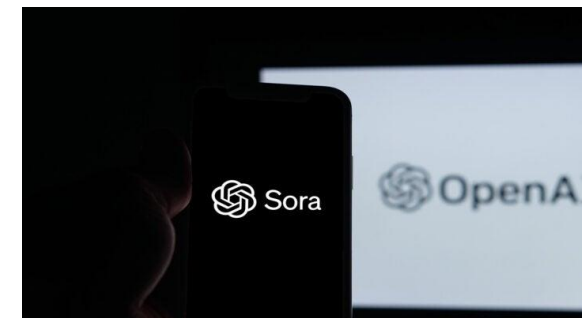




# AI Timeline



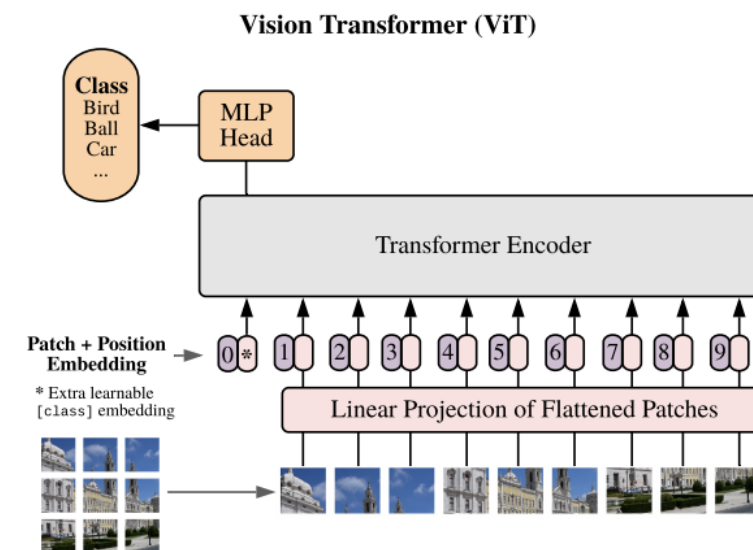
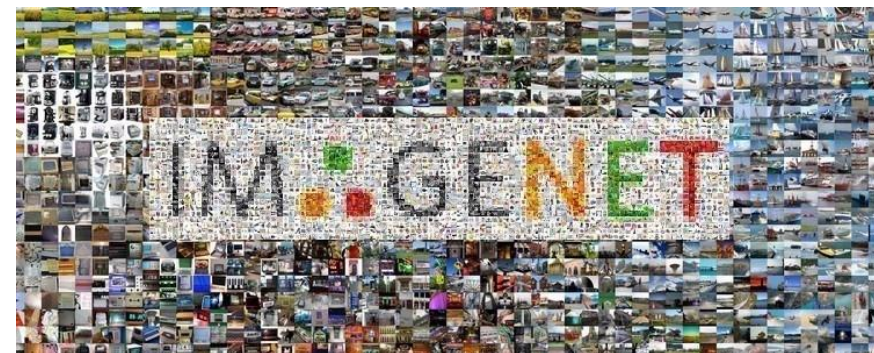
- 2016 – Google’s DeepMind program **AlphaGo** defeated the Go grandmaster Lee Sedol
  - The game of Go is more difficult than chess, because the number of possible moves is much greater
- 2017 – **Transformer** network architecture was introduced in the paper by Vaswani et al. “Attention Is All You Need”
- 2020 – The University of Oxford developed an AI test called **Curial** to rapidly identify COVID-19 in emergency room patients.
- 2020 – OpenAI’s **GPT-3** is the first large language model with 175B parameters, performed well on many NLP tasks
- 2021 – DeepMind’s **AlphaFold** achieved high accuracy in predicting the 3- dimensional shape of proteins
- 2022 – OpenAI’s **DALL·E 2** generated photorealistic images with remarkable quality
- 2022 – Facebook’s **NLLB** (No Language Left Behind) model for machine translation between 200 languages
- 2022 – Deep Minds’ **Gato** model was trained to perform over 450 tasks
- 2023 – OpenAI announced the GPT-4 multimodal LLM (**ChatGPT**) that receives both text and image prompts.
- 2024 – **Sora** is an AI model that can create realistic and imaginative scenes from text instructions.



# DL Success in Computer Vision



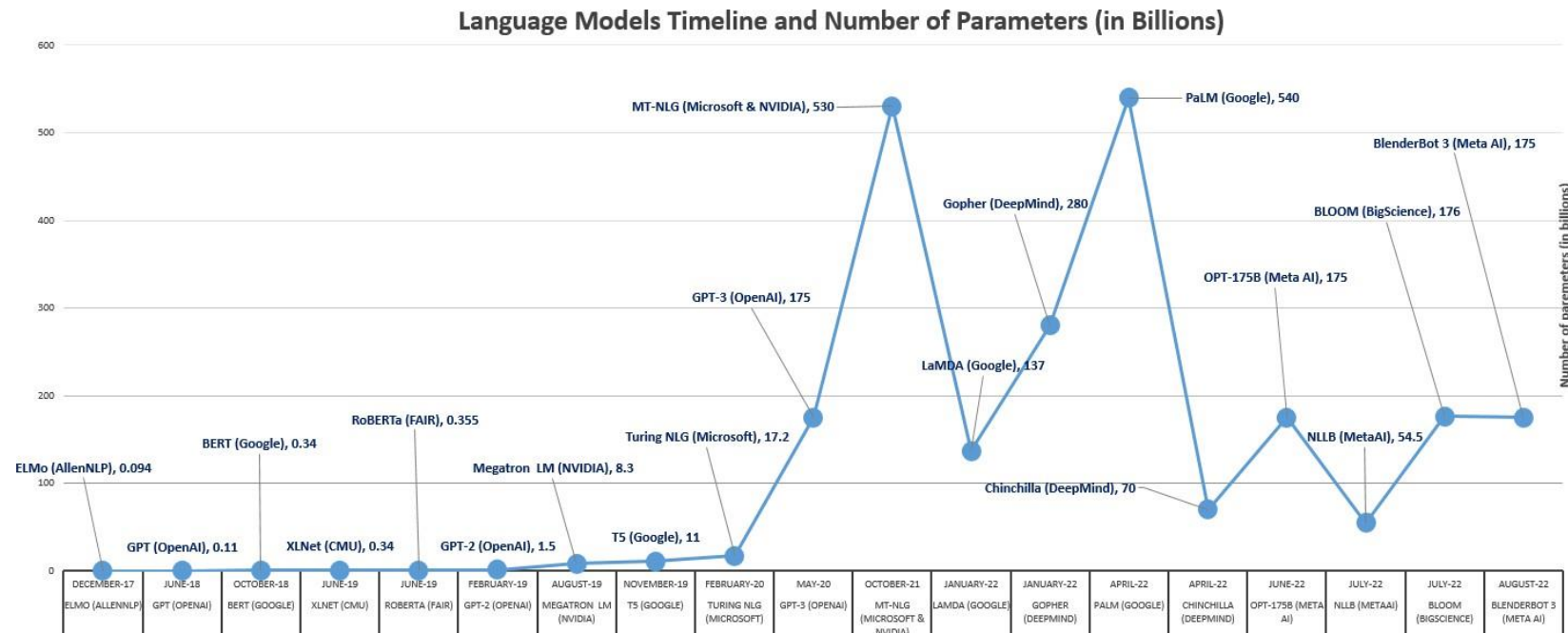
- *Computer vision* tasks
  - Image and video recognition/classification, segmentation, object detection, image synthesis
- Important architectures
  - AlexNet – 2012
    - Convolutional NNs for image recognition, 5 layers, GPU for parallel processing
    - ImageNet Large Scale Visual Recognition Challenge (ILSVRC): AlexNet reduced the error on ImageNet from 26% by traditional ML approaches to 15%
  - VGG 2014
    - 16 layers CNN architecture
  - Inception – 2015
    - Stacked 1x1 convolutions, 22 convolutional layers
  - ResNet - 2015
    - Introduced residual connections, it is a family of networks with 18, 34, 50, 101, and 152 layers
    - Several related models were proposed afterwards, e.g., ResNeXt (2017), EfficientNet (2019)
  - Vision Transformers – 2020
    - Employ attention layers, inspired by the transformer models used in NLP



# DL Success in Natural Language Processing



- *Natural Language Processing* (NLP) tasks
  - Text classification, text summarization, speech recognition, machine translation, dialog generation, part-of-speech tagging
- In the last 4 years, *Language models (LMs)* powered by deep NNs achieved unprecedented success in NLP tasks
  - Compared to the human brain having between 100 and 500 trillion synaptic connections, these models are still fairly small





# GPT-3: DL Success in NLP



- **GPT-3** (Generative Pretrained Transformer 3)
  - Number of parameters: 175 billion
  - Training dataset: 45 TB of text (= a large portion of all text available on the web)
  - Training time and GPUs: 36 days with 1,024 NVIDIA A100 GPUs
  - Training cost: \$US 12 million
- GPT-3 training
  - Self-supervised learning – it is a form of unsupervised learning (from unlabeled data)
  - Very simple approach: predict (assign probability to) the next word in a given sequence of words

Input: A quick brown

Input: Marry had a little

Input: Nothing is

Output: fox Output:

lamb Output:

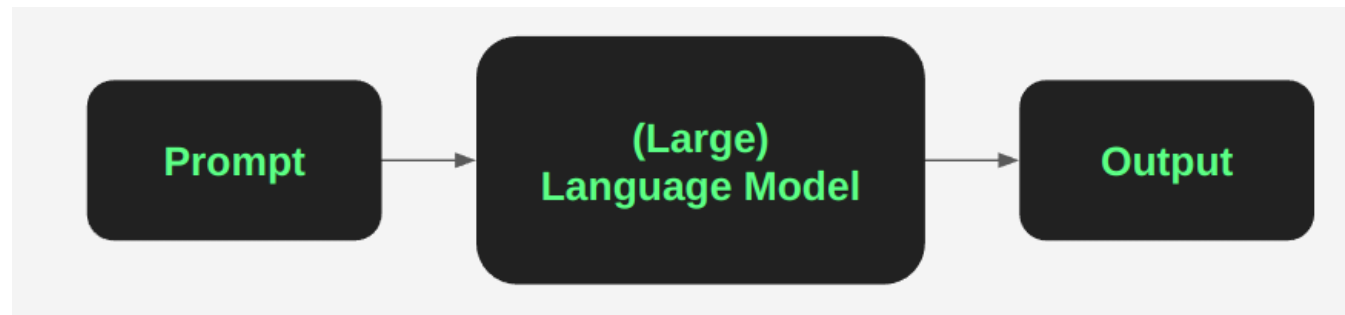
impossible

- Finite, discrete solution space: the next word must be from a finite dictionary
  - There are about 170,000 words in the English language
  - A person on average uses 20,000 to 30,000 words
  - About 3,000 words cover 95% of all written text

# Large Language Models



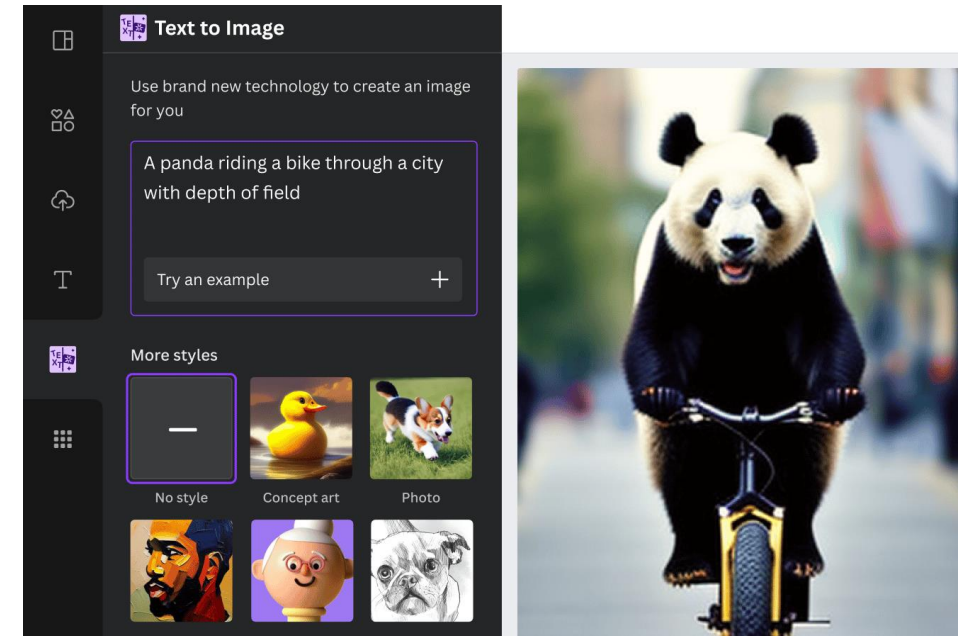
- *Large language models* (LLMs)
  - The architecture of all LLMs is based on the **transformer** networks
  - Transformers employ the **attention mechanism** to identify the words in a sentence that impact the meaning of other words
  - I.e., important characteristics is the ability for modeling words based on the context
- LLMs work with projected words into an embeddings space, where each word is replaced with a numerical **token**
  - Given a sequence of tokens from a dictionary, the training objective is to estimate the probability of the next token
- The quality of generated text by recent large LMs is often undistinguishable from human-written text
- Concerns regarding LLMs:
  - Misuse and unethical use of AI, amplifying disinformation, environmental impact (high carbon emissions), increasing economic inequalities, centralization of power (e.g., affordable only by the largest corporations)



# Generative Text-to-Image Models



- **Generative models** learn to generate new data instances, given a training set
  - The family of GAN models (StyleGAN, CycleGAN, SRGAN) were the most important generative models in recent years
- Latest **text-to-image models** released this year include:
  - [DALL·E 2](#) by OpenAI
  - [Imagen](#) by Google
  - [Stable Diffusion](#) by Stability.ai
- **Remarks:**
  - Significant progress has been made since 2014 when GAN was introduced
  - The above text-to-image models employ text embeddings from pretrained LLMs (e.g., GPT-3 used with DALL·E 2)
  - Produce images with remarkable photorealism, accurate fine details, compositionally, spatial relations of the objects in images, and even with creativity in image synthesis
  - They employ **diffusion probabilistic models**, which outperformed GANs
    - Diffusion models use NNs to learn the steps of adding and removing noise to images
  - Can create new images which are unlikely to have been seen in the training data



# Images Generated by DALL·E 2



- These are a few (cherry-picked) examples of images generated by DALL·E 2



A photo of a quaint flower shop storefront with a pastel green and clean white facade and open door and big window



Cat sipping tea and posting to twitter while sitting on a couch



A rabbit detective sitting on a park bench and reading a newspaper in a victorian setting



A lion in a hoodie hacking on a laptop



Teddy bears shopping for groceries in ancient Egypt



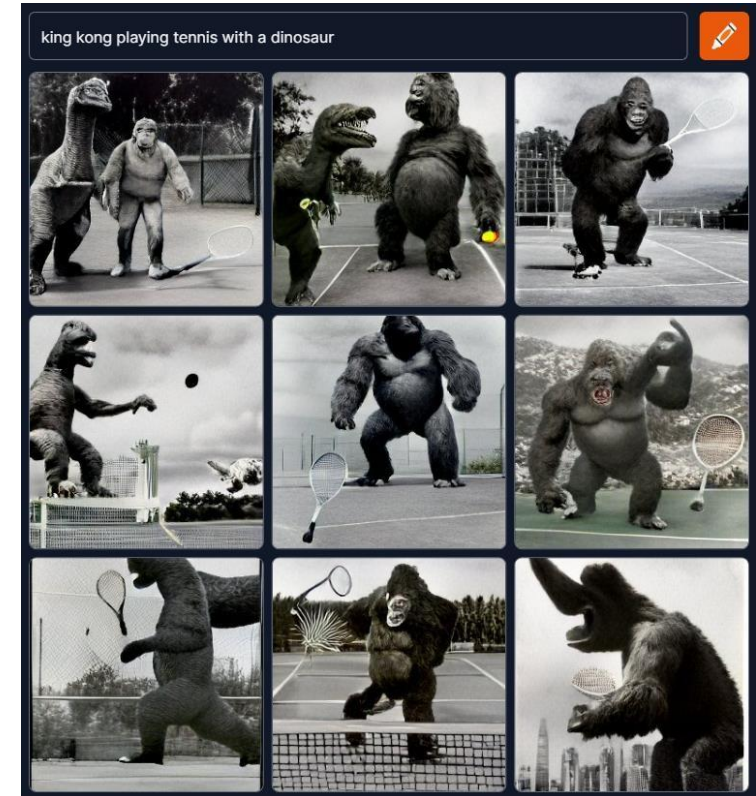
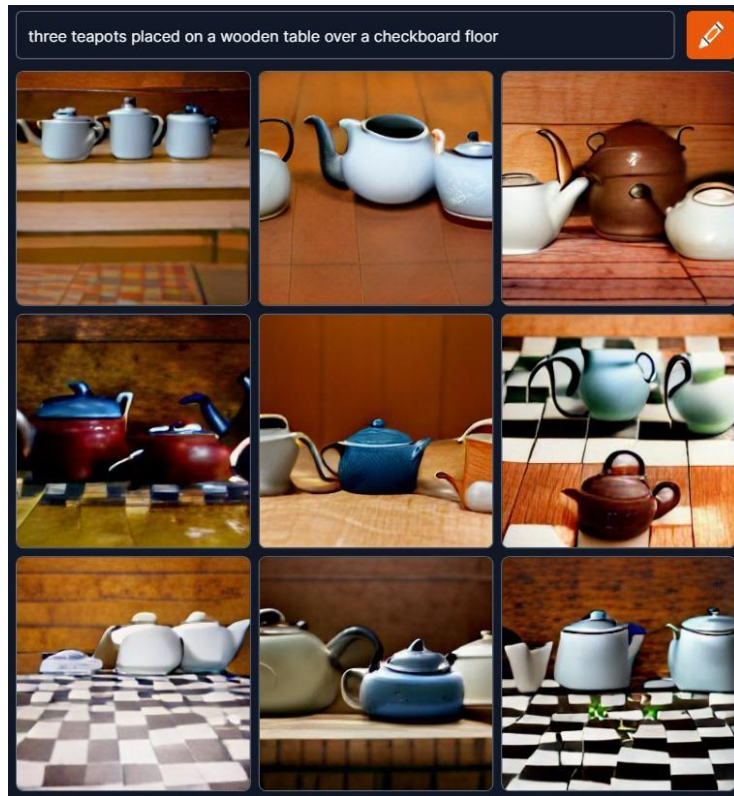
Teddy bears working on new AI research on the moon in the 1980s



# Open-Source Text-to-Image Models



- A smaller model called *Crayon* (a.k.a. *DALL·E Mini*) is freely available [here](#)
  - It takes about 2 minutes to create 9 images to a text prompt
  - However, it is less powerful, and the results are less impressive
- *Stable Diffusion* was released recently, can be accessed [here](#)



# Foundation Models



- *Foundation models* are large NN models trained at **scale** with high capabilities for **transfer learning** to many other applications
  - Early examples of foundation models are the LLMs, such as GPT-3 and PaLM
- The scale of these models results in new **emergent capabilities** – e.g., perform well on tasks on which they were not explicitly trained to do
  - “**Emergence** is when quantitative changes in a system result in qualitative changes in behavior”
  - This allow fine-tuning to new tasks with small number of training data instances
    - E.g., **few-shot learning** refers to fine-tuning with only a few instances
- Notable applications of pretrained LLMs include:
  - Programming code completion models: CoPilot, AlphaCode, Codex, Codegen
  - Text-to-image generative models: DALL·E 2, Imagen, Stable Diffusion
  - Protein sequence prediction, solving math problems, preparing legal documents  
(other task examples are listed on the next page)
- Transfer learning is what makes foundation models possible, but scale is what makes them powerful

# Foundation Models



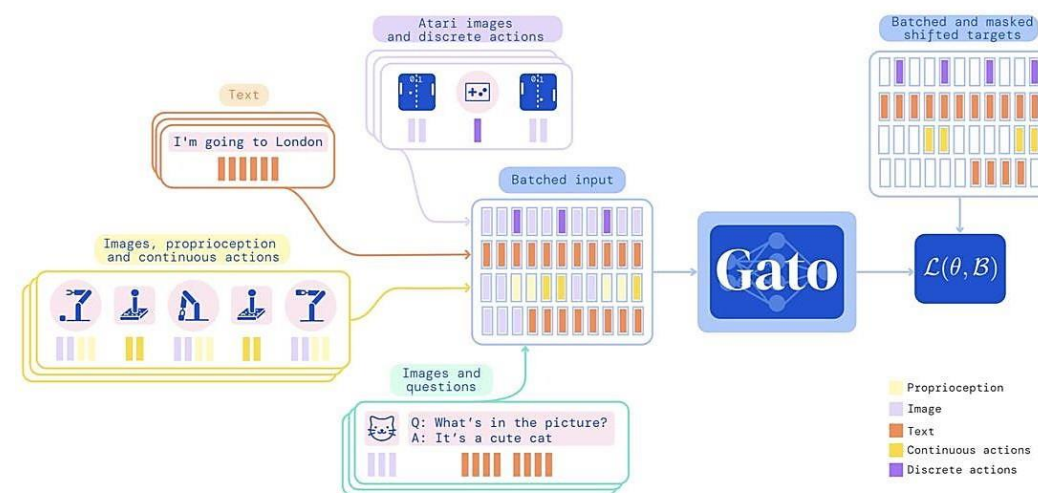
- Examples of applications and downstream tasks in which foundations models are being used

<b>Program writing</b>	<b>Image captioning</b>	<b>Generate images</b>	<b>Parse data</b>	<b>Classify text</b>
Use natural language to generate SQL/Python/Java code	Describe and classify images	Create images based on natural language	Extract data from images	Identify entities, parts-of-speech, and other text categories
<b>Q&amp;A</b>	<b>Writing assistant</b>	<b>Summarize</b>	<b>Solve homework</b>	<b>Translate</b>
Answer natural language questions based on knowledge base	Correct your writing	Summarize text to key concepts	Solve basic math and programming problems	Translate text from one language to another
<b>Code explanation</b>	<b>Copywriting</b>	<b>Sentiment rating</b>	<b>Recipe creation</b>	<b>Chat</b>
Writes the description of code functionality in natural language	Generate ad/product/job descriptions based on short prompts	Rates the sentiment, toxicity, warmth, etc. of text	Use at your own risk	Talks like a human

# Gato – A Generalist Agent



- **Gato** by Deep Mind is a multi-modal, multi-task, multi-environment network
- The same model with the same weights can: play games, manipulate a robot, caption images, generate dialog, navigate in 3D, and many other tasks
  - Inputs: text, images, robotic joint torques (proprioception), button presses (for games)
  - Outputs are based on context: text (dialog, translate, summarize), torque powers (for the actuators of a robotic arm), button presses (to play games), etc.
- Gato demonstrates versatility and adaptability to many tasks (over 450 tasks)
  - The model has “only” 1.2 billion parameters





# AI Limitations and Challenges



- Despite excellent pattern recognition abilities, current DL models are **unable to reason about the objects** in images or take **context into consideration**
- E.g., predictions by a DL model on images of randomly positioned parts
  - The model assigns weights to different features in images, and outputs a category based on the sum of weights for all features
  - It does not take into account the spatial relations between the features in making the prediction

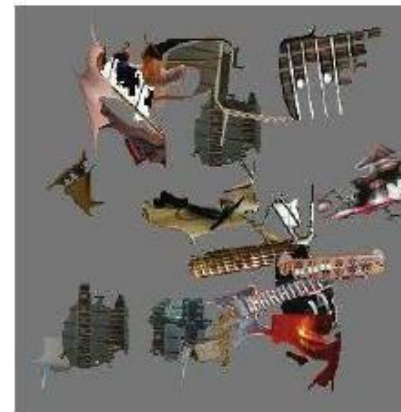
Basketball



Zebra



Electric Guitar



# Trustworthy AI



- *Trustworthy AI* – efforts to address the limitations to ensure that end-users can trust the predictions by AI models
- Topics in trustworthy AI include:
  - **Robustness**
    - Even unnoticeably small perturbations can impact the model predictions
  - **Generalization**
    - OOD (out-of-distribution) inputs; e.g., a model trained on medical images in one hospital performs poorly on images in another hospital (due to different equipment or settings used)
  - **Explainability**
    - The decision-making process of large models is non-transparent and difficult to understand
  - **Fairness**
    - Predictions can show bias against demographic groups, based on gender, age, culture
  - **Privacy protection**
    - Models can memorize and reveal input data; e.g., a model can reveal sensitive private information in medical records used for training
  - **Ethics**
    - The models should produce ethical decisions that are aligned with our human values (also referred to as **AI Alignment**)

# Engineering vs Science Phase of Technology



- Theoretical guarantees about the AI performance are currently lacking
  - Currently, AI is in *Engineering phase*: models are designed to solve tasks, are integrated into new products, add value to companies, have economic impact
  - *Science phase* of AI is to follow: Develop theory to guarantee convergence, prove stability, interpret the decisions, explain successes and failures of models
- Various technologies historically began with an engineering phase (inventions made, products built) to be later followed by a science phase (theory developed)
  - Steam engines were used in paper mills and factories since 1776; the theory of Thermodynamics was developed between 1820s and 1850s
  - Airplanes were constructed and flown since 1904-1905; the modern theory of Aerodynamics was developed in 1930s
  - Electric circuits were discovered around 1800; the theory of Electromagnetism was founded between 1820s and 1830s

# The Bitter Lesson



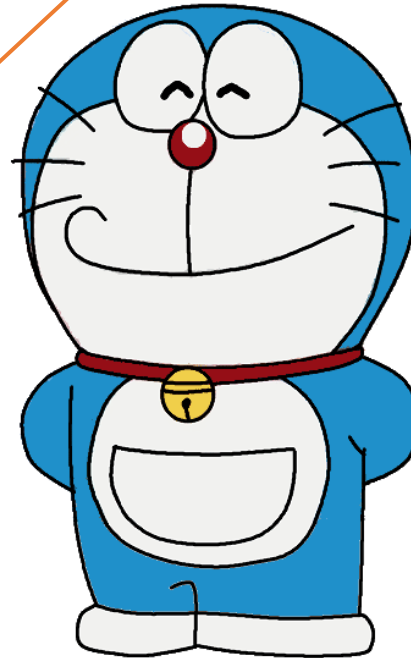
- *The Bitter Lesson* (2019) is a short paper by Rich Sutton
  - <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>
- The Bitter Lesson is based on his observations regarding the historical development of AI methods, which can be characterized with three phases:
  - Phase 1 - AI researchers incorporate human domain knowledge into their AI methods, which helps in short term
  - Phase 2 - In the long term, the performance of such models plateaus without further progress
  - Phase 3 - Progress is eventually achieved by general methods that scale computation with search and learning
- In conclusion:
  - AI methods that **leverage computation and search at scale** are the most effective
  - Human-centric approaches complicate methods and make them less suited to leveraging computation and search at scale
  - The search for solutions should be done by our methods, not by us
  - We want AI methods that can discover like us, and not based on our discoveries

# Prospective Trends in AI



- *Unsupervised/self-supervised learning*
  - Increased use of raw data without annotations or labels
- *Homogenization*
  - Convergence of architectures and methodologies in building AI systems across different applications
  - E.g., transformers are replacing convolutional, recurrent networks, and are increasingly being used in computer vision, NLP, time-series, tabular data tasks
- *Training at scale*
  - We can expect to see further scaling along the three main factors: amount of computation, number of model parameters, and training dataset size
- *Multi-modal learning*
  - Capacity to learn from multiple simultaneous sources of information (like humans)
  - Task-specific models being replaced with general models that can solve multiple tasks
- *Causal learning*
  - Can new learning algorithms be developed that are capable of learning cause and effect, semantic relationships?

# Happy Learning



## References and Acknowledgements

[Machine learning Overview: Science Paper](#) (2015)

[Deep Learning Overview: Nature Paper](#) (2015)

[Python Programming for Data Science by Alex](#) (2023)