Post Read : Gen-Ai Training

Contents

1. Introduction to AI

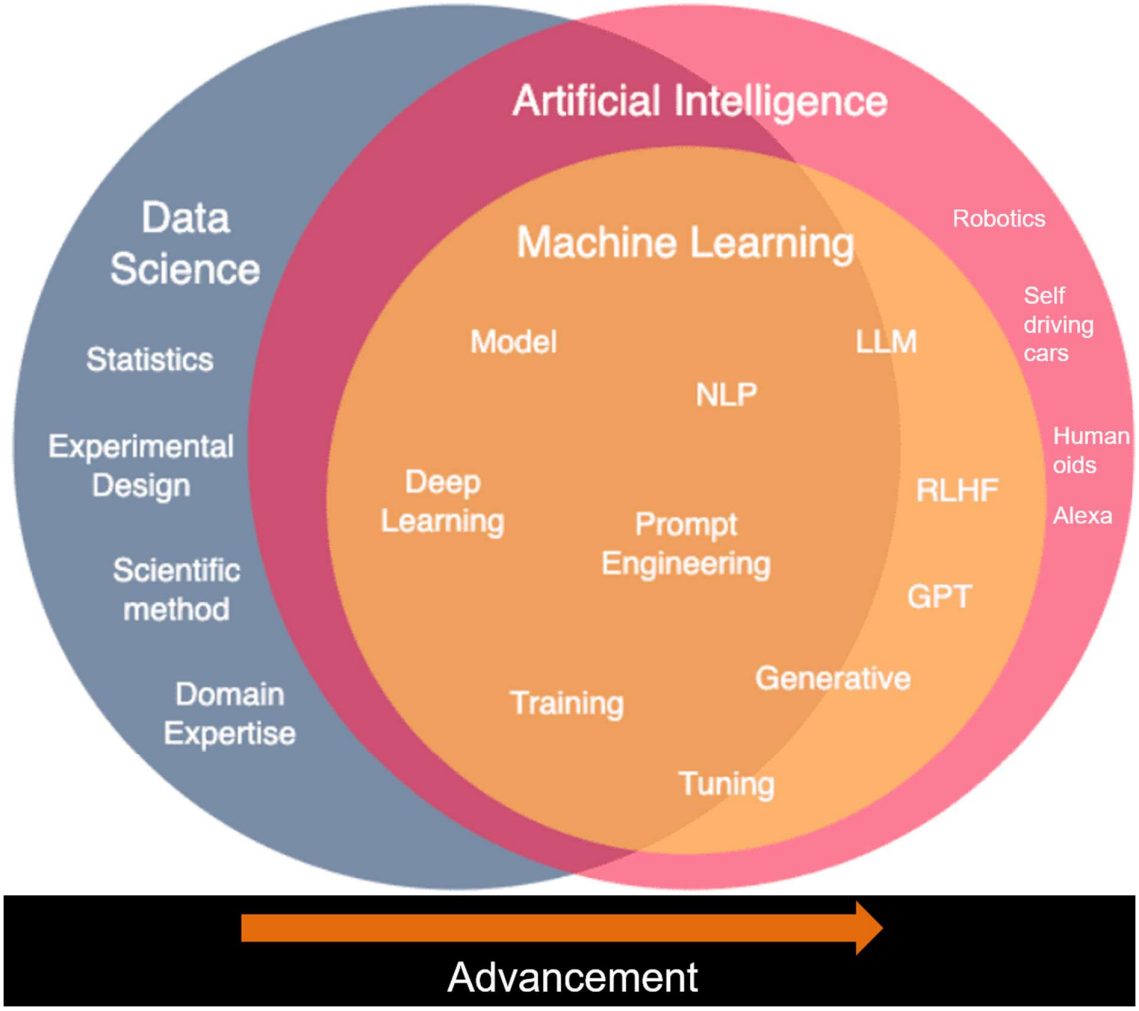
1.1 What is AI?

Artificial Intelligence (AI) is the field of computer science focused on creating machines or systems that can perform tasks that would typically require human intelligence.

These tasks often include:

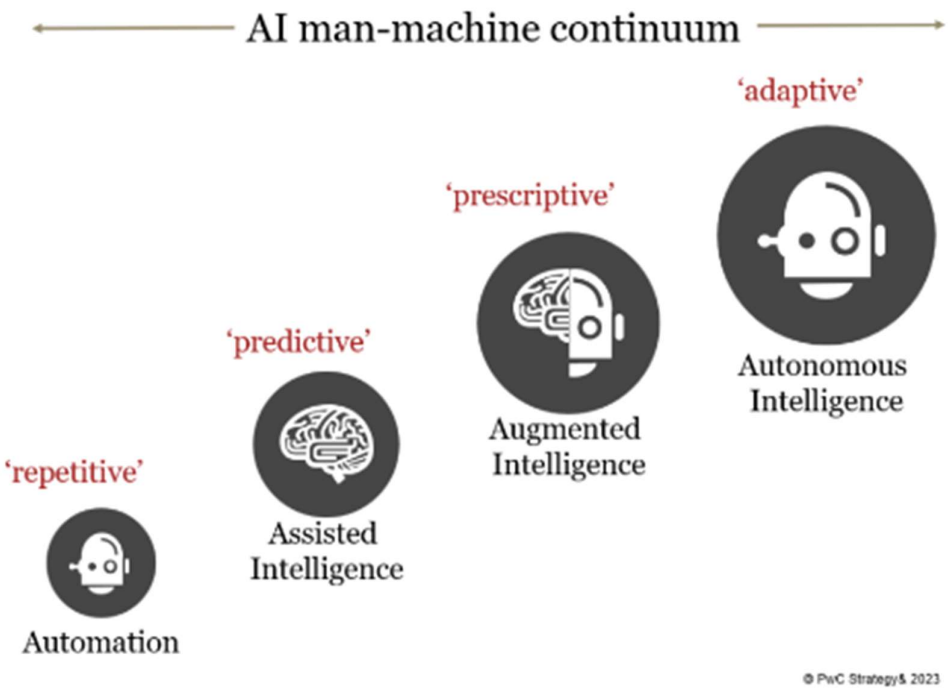
- Reasoning
- Learning
- Problem-solving
- Understanding natural language
- Recognizing patterns
- Making decisions
- Generating new stuff based on it's learning

AI systems can adapt and improve over time, especially those utilizing advanced techniques like machine learning (ML) and deep learning.



The AI Story: From Turing to GPT-4 Every great story has a beginning, and AI's starts with the Turing Test in 1950, a brainchild of Alan Turing, to test a machine's ability to imitate human intelligence. Fast forward to 1956, and we have the term "artificial intelligence" coined by John McCarthy. From there, the AI journey takes us to milestones like Shakey (1969), the first-ever general-purpose mobile robot, and Deep Blue's triumph over the world chess champion (1997), which illustrated AI's potential.

As we venture into the 21st century, AI continues to push boundaries. From robotic process automation (RPA) and smart homes to self-driving cars, and more recently, OpenAI's GPT-4 (2022), showcasing high-quality text generation—a leap in natural language processing capabilities.



1.2 Roadmap

A good roadmap at:

ℏεsam on X: "if you are completely new to LLMs and NLP, here is a solid roadmap to build your knowledge from the foundations: (+ link to resources) 1️⃣ **Foundations** ↳ Mathematics for Machine Learning [Book] ↳ Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow https://t.co/trhbboVoey" / X

Foundations ↳ Mathematics for Machine Learning [Book] ↳ Hands-On Machine

Learning with Scikit-Learn, Keras, and TensorFlow [Book] ↳ (Optional) Introduction to Statistical Learning [Book] — for statistics understanding

Basic NLP Concepts ↳ Stopwords, Stemming, Lemmatization ↳ Bag of Words ↳ TF-IDF ↳ BM25 ↳ Word2Vec ↳ N-grams ↳ Named Entity Recognition (NER) ↳ Tokenization ↳ Part-of-Speech (POS) Tagging

Get Familiar with Frameworks ↳ NLTK ↳ SpaCy ↳ PyTorch (important, more popular now) ↳ (Optional) Hugging Face basics

Understand Embeddings ↳ Word2Vec ↳ GloVe ↳ FastText (Understand pre-transformer embeddings properly)

Dive into Transformers & **Models** ↳ Understand how Transformers work ↳ Explore BERT, ModernBERT, DistilBERT, etc. ↳ Learn Sentence-Transformers & Hugging Face
Dive into Evaluations ↳ BLEU Score, ROUGE Score, Perplexity, etc. ↳ Accuracy, Precision, Recall, F1 Score, etc. ↳ Semantic Similarity evaluations (Cosine similarity, embedding-based metrics)

Explore Fine-tuning of Small Language Models (**e.g.**, **BERT**) ↳ Fine-tune for classification ↳ Fine-tune for question-answering ↳ Fine-tune for domain-specific embeddings ↳ Understand task-specific heads (classification/QA) ↳ Learn fine-tuning best practices: LR schedules, early stopping, regularization credits to the amazing Shantanu Ladhwe for his amazing solid roadmaps

NLP Roadmap (pre-LLM)

1. **Foundations**
   - ↳ Mathematics for Machine Learning [Book]
   - ↳ Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow [Book]

2. **Basic NLP concepts**
   - ↳ Bag of words
   - ↳ TD-IDF
   - ↳ BM25
   - ↳ word2vec
   - ↳ n-grams
   - ↳ Named Entity recognition
   - ↳ Tokenization
   - ↳ Stop words, Stemming, etc.

3. **Get familiar with frameworks**
   - ↳ NLTK
   - ↳ SpaCy
   - ↳ PyTorch → Important {more popular}

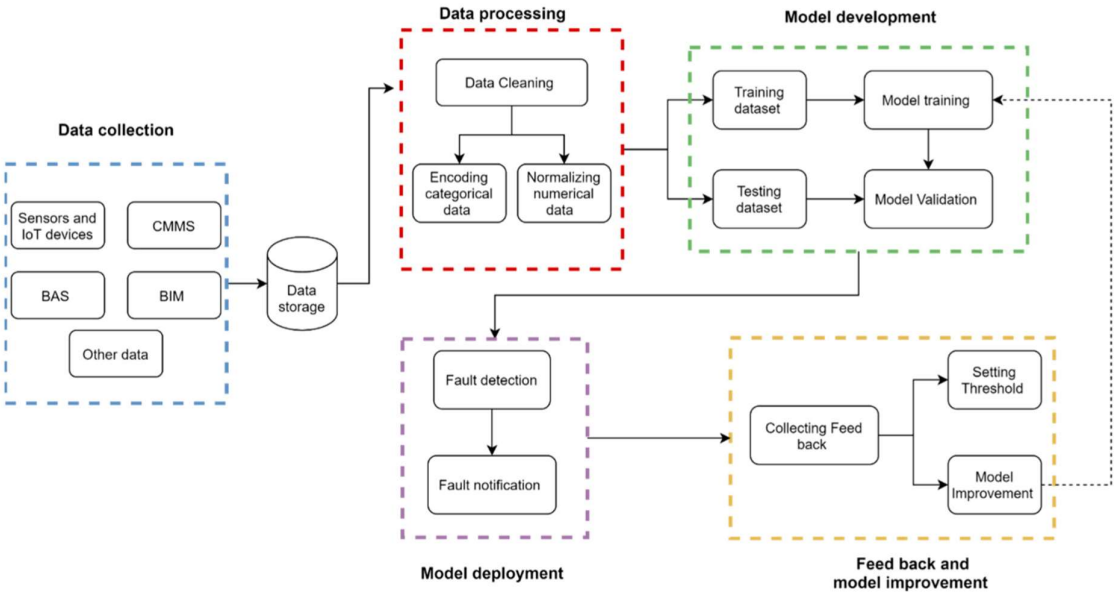
4. **Understand Embeddings**
   - ↳ word2vec, GloVe, FastText (pre-transformer era)

5. **Dive into Transformers & models** {skipping RNN, seq2seq}
   - ↳ Understand Transformer
   - ↳ BERT, ModernBERT, DistilBERT, ...
   - ↳ Explore sentence-transformer & HuggingFace

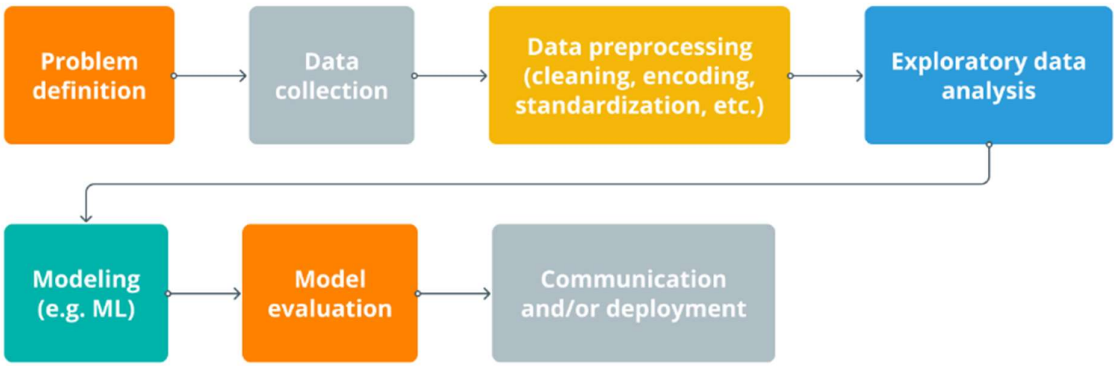
6. **Dive into Evals:⇒**
   - ↳ BLEUScore, ROUGE score, Perplexity... (more)

7. **Explore fine-tuning of small language models eg. BERT**
   - ↳ For classification
   - ↳ For question-answering
   - ↳ Generate fine-tuned embeddings

AI Project Lifecycle



2. Exploratory Data Analytics (EDA)



Data Cleaning

Handling missing values, outliers, duplicates, and inconsistencies.

Descriptive Statistics

Summarizing data using mean, median, mode, standard deviation, etc.

Data Visualization:

Using plots like histograms, scatter plots, box plots, and correlation matrices to visualize relationships between variables.

Feature Engineering

Creating new features or transforming existing ones to make data more meaningful.

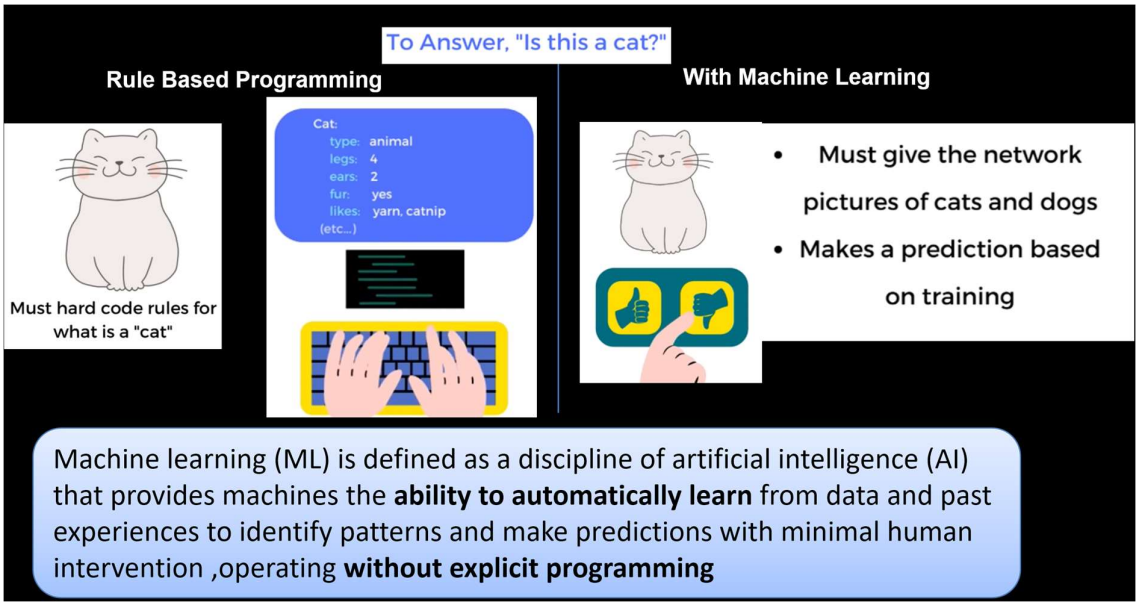
- Dimensionality Reduction
- Encoding (convert categorical features into numerical, for e.g. One Hot Encoding)
- Normalization (Min-max scaling, Log transformation)
- Binning
- Adding new columns

2.1 EDA Example

https://www.kaggle.com/code/abhishek14398/loan-defaulters-prediction-eda-lending-club-study

3. Machine Learning

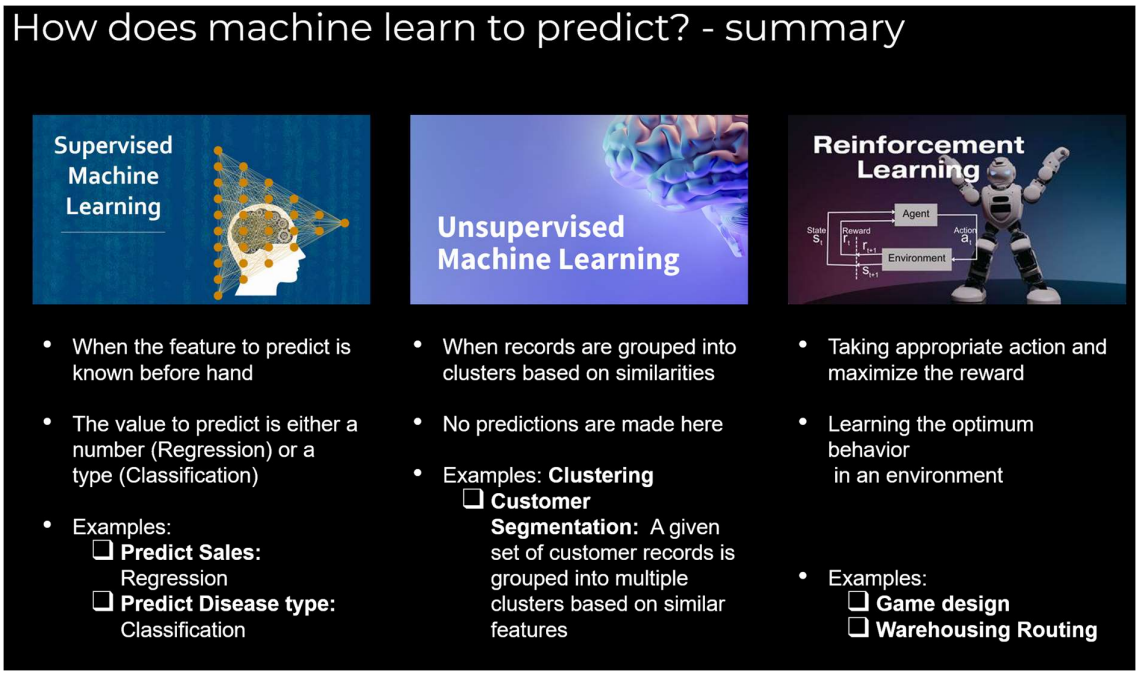
3.1 Machine Learning Vs Rule Based Programming



3.2 Data Science Problem Type with examples

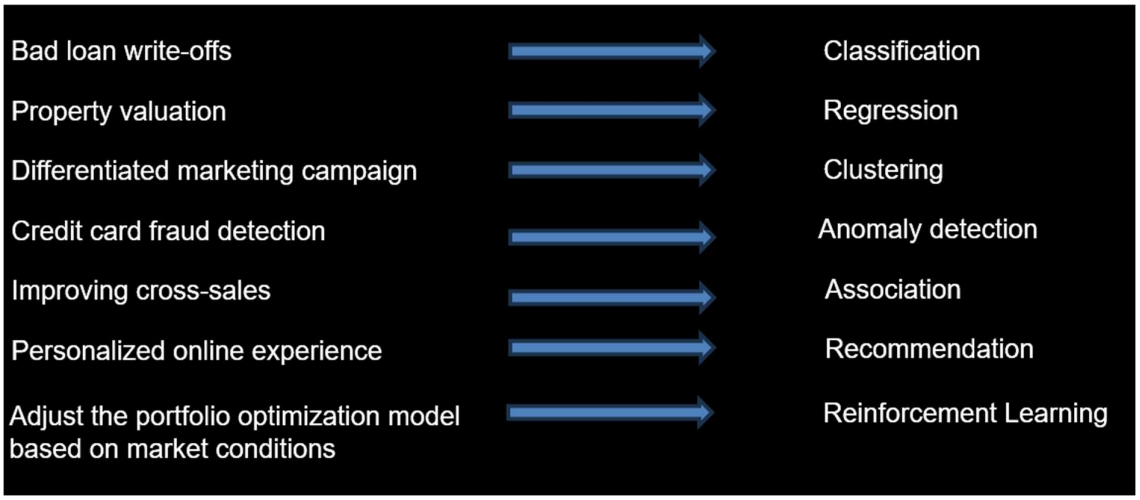
Problem	Description	Examples
Classification	Predicting whether a data point belongs to one of the predefined classes	Customer Churn Prediction, Spam Detection
Regression	Predicting the numerical value of the target variable	Agricultural Yield, Inflation Rate Prediction
Clustering	Identifying natural groups within the dataset based on similar inherent property of the data points	Social Network Analysis, Crime Incidence Analysis, Search Result Grouping
Anomaly Detection	Predicting whether a data point is an outlier in comparison with other points in the data set	Detecting Network Intrusions, Predicting Machine Failures, Detecting Gene Mutations
Association	Discovering rules that govern frequent simultaneous occurrence of certain items or phenomena	Medical Diagnosis, Protein Sequencing, Building Intelligent Transportation Systems
Recommendation	Suggesting items for users based on the past preferences of theirs and similar users	Recommendation of Movies, Books, Restaurants, Holiday Destinations

3.3 Common training methods and algorithms



Problem Type	Training Method	Common Algorithms
Classification	Supervised	Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Neural Networks
Regression	Supervised	Linear Regression, Ridge/Lasso Regression, Polynomial Regression, Random Forest, Gradient Boosting Machines (GBMs), Neural Networks
Clustering	Unsupervised	K-Means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Models (GMMs), Spectral Clustering
Anomaly Detection	Semi-supervised, Unsupervised, or Supervised	Isolation Forest, One-Class SVM, Autoencoders, DBSCAN, Gaussian Mixture Models, LOF (Local Outlier Factor)
Association	Unsupervised	Apriori, Eclat, FP-Growth (Frequent Pattern Growth)
Recommendation Engine	Supervised, Unsupervised, or Reinforcement Learning	Collaborative Filtering (Matrix Factorization, Singular Value Decomposition), Content-Based Filtering, Neural Collaborative Filtering, Reinforcement Learning (Bandit algorithms)

Examples:



3.3.1 Linear Regression

How it works:

Linear regression models the relationship between a dependent variable y and one or more independent variables X by fitting a linear equation:

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$

where β_0 is the intercept, $\beta_1 \ldots \beta_n$ are coefficients, and ε is the error term.

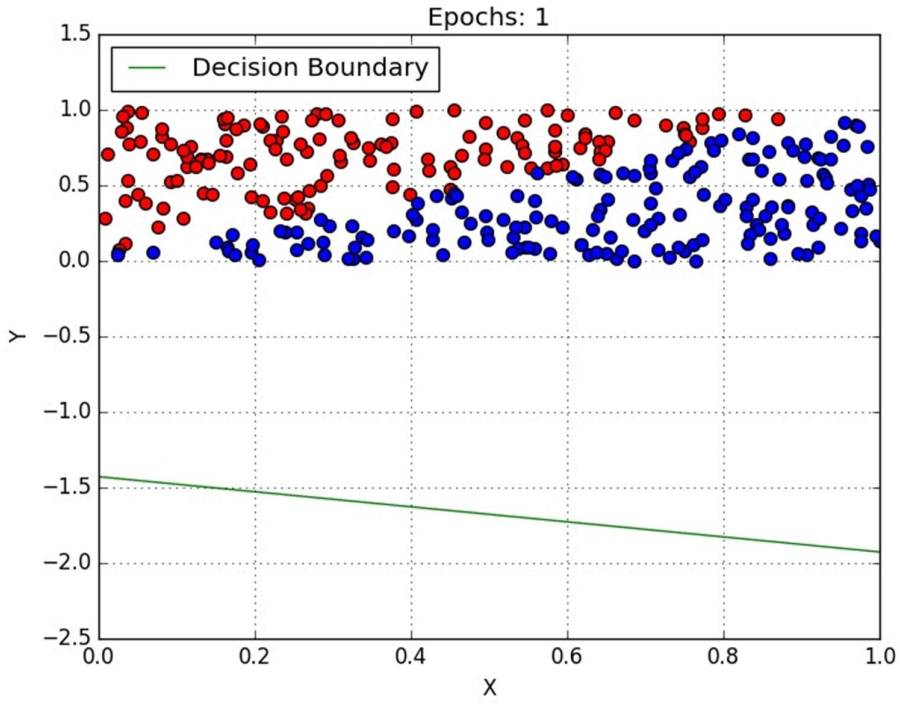
The model estimates these coefficients to minimize the sum of squared residuals (differences between predicted and actual values).

Evaluation Metrics:

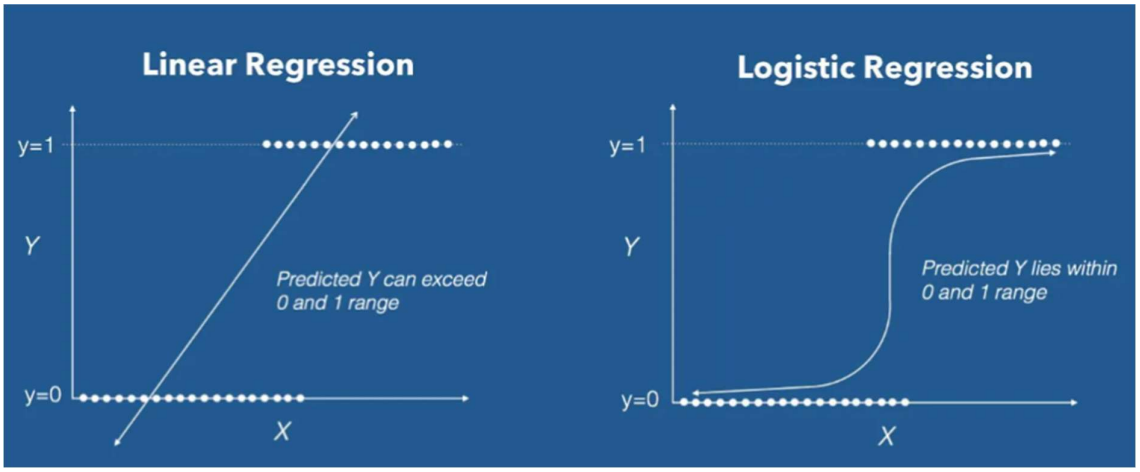
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (coefficient of determination)
- Adjusted R-squared

3.3.2 Logistic Regression

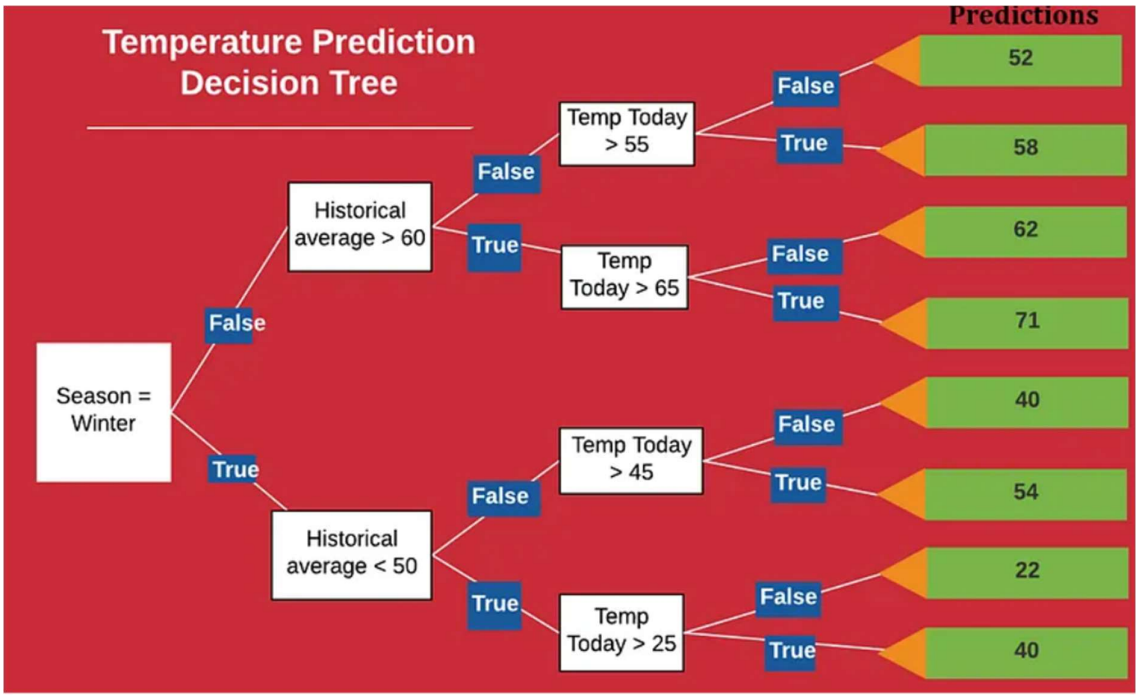
Used for classification tasks:



Logistic regression uses Sigmoid function to suit the classification task:



3.3.3 Decision Tree



Advantages

- No need of feature scaling
- Handles both categorical and numeric data
- No assumptions about distribution
- Explainability, easy to understand

Limitations

- Overfitting
- Biased when data is imbalanced
- Unstable (High variance)
- Inefficient with high dimensional data

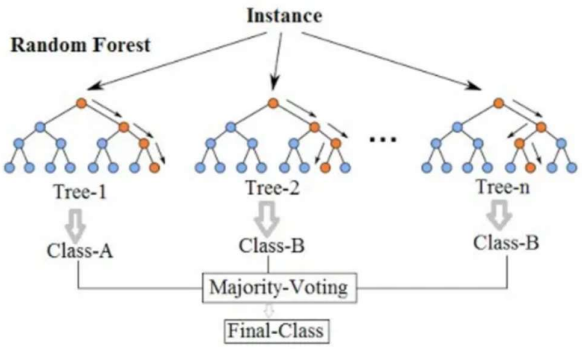
Decision trees have high variance

Models such as Random forest Gradient Boosting helps minimize the variance

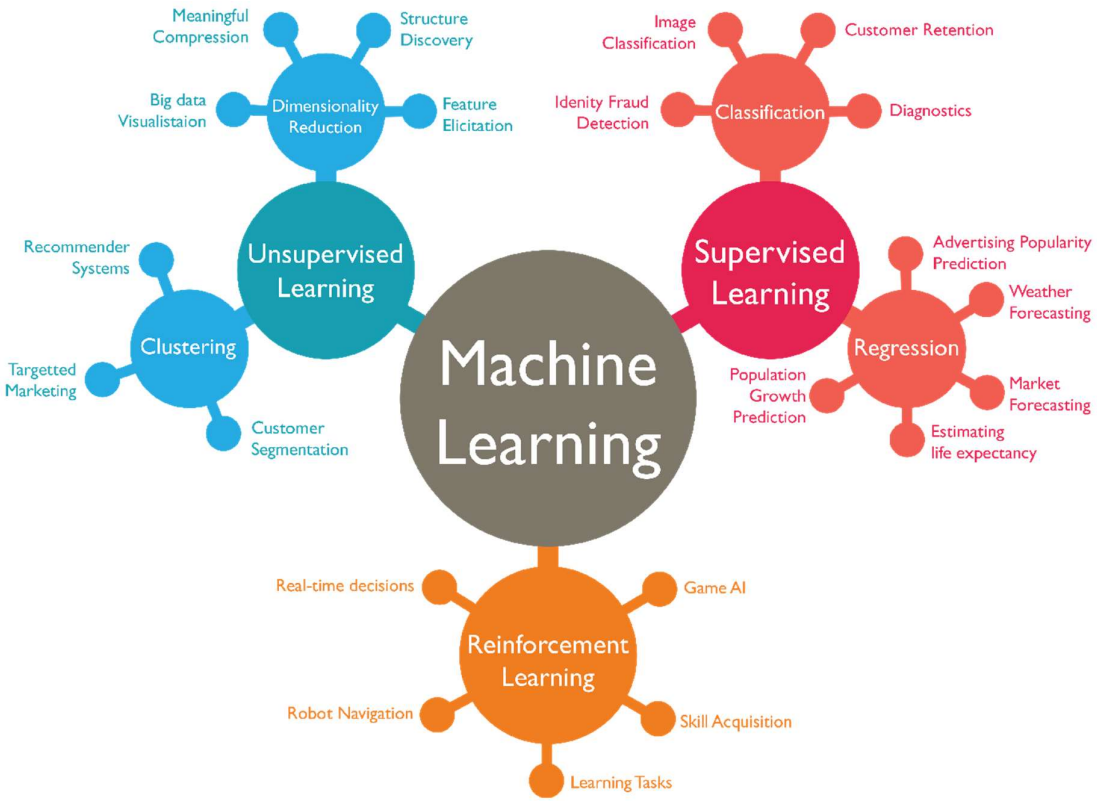
Random Forest Model

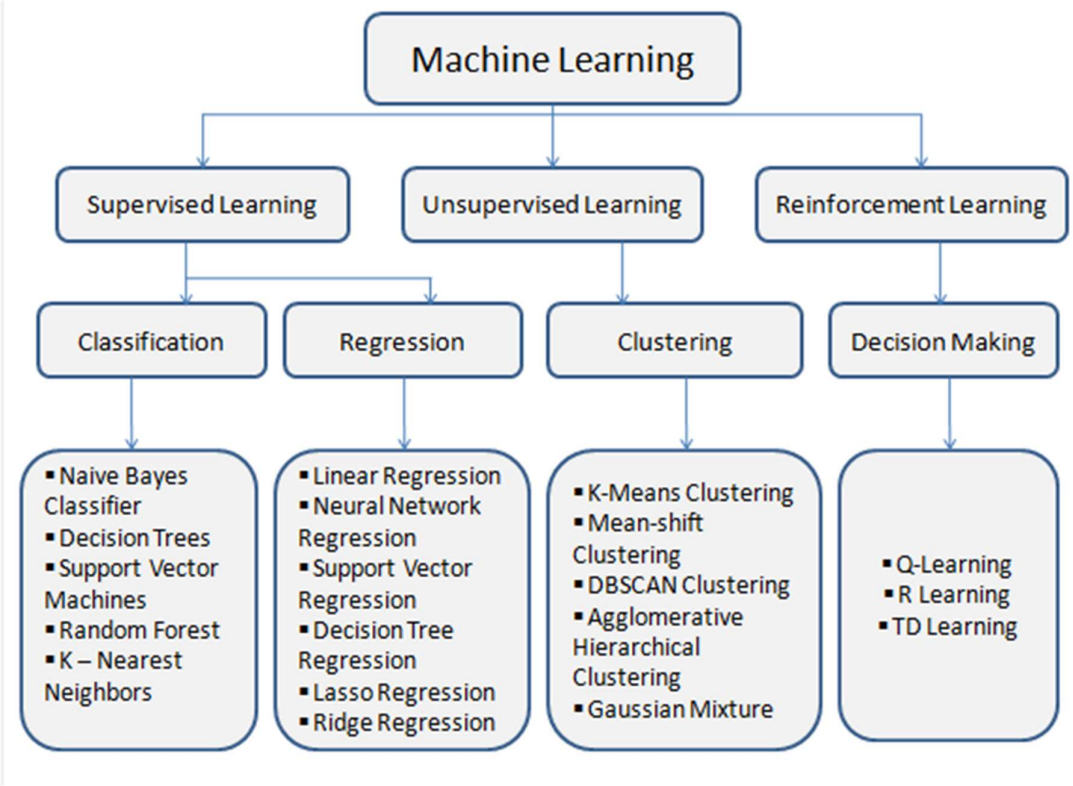
Uses **Bagging** technique to reduce variance.

Bagging is an ensemble learning technique that builds multiple independent models (in the case of Random Forest, these are decision trees) using different random subsets of the training data.

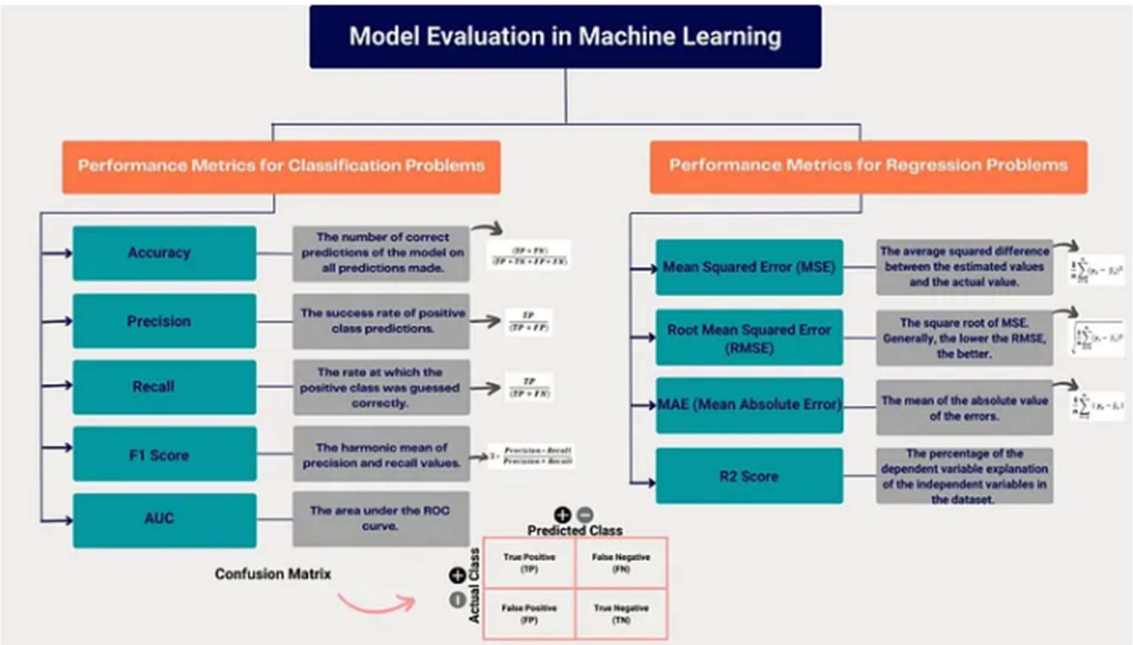


Machine Learning Map

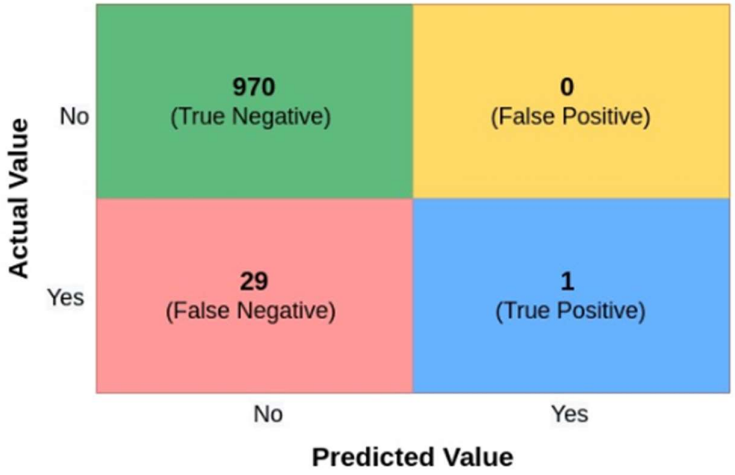

Machine Learning
Unsupervised Learning
Meaningful Compression
Structure Discovery
Dimensionality Reduction
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Image Classification
Customer Retention
Idenity Fraud Detection
Classification
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Decision Making
▪ Naive Bayes Classifier
▪ Decision Trees
▪ Support Vector Machines
▪ Random Forest
▪ K – Nearest Neighbors
▪ Linear Regression
▪ Neural Network Regression
▪ Support Vector Regression
▪ Decision Tree Regression
▪ Lasso Regression
▪ Ridge Regression
▪ K-Means Clustering
▪ Mean-shift Clustering
▪ DBSCAN Clustering
▪ Agglomerative Hierarchical Clustering
▪ Gaussian Mixture
▪ Q-Learning
▪ R Learning
▪ TD Learning

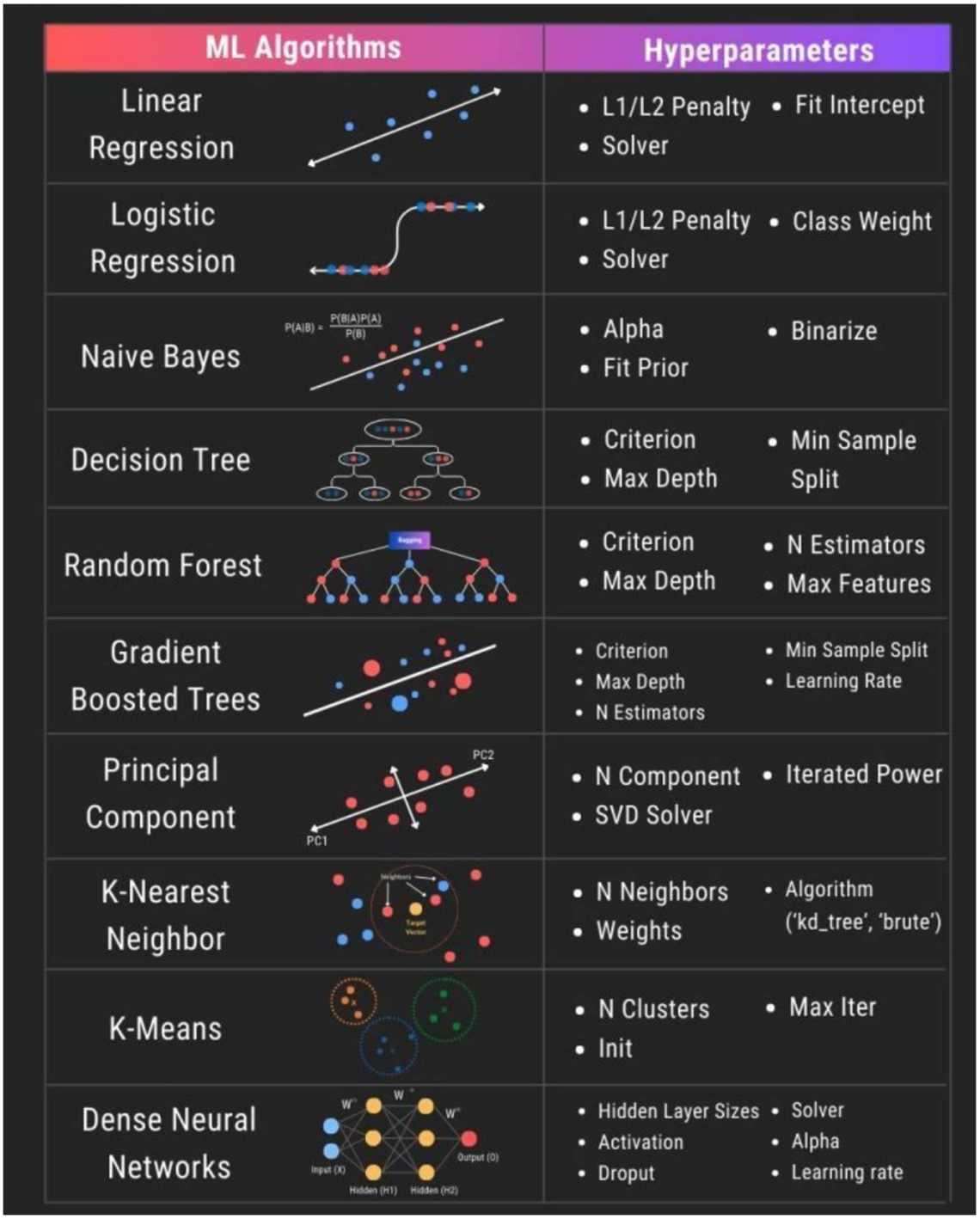
Model Evaluation



Confusion Matrix, F1 score, Precision, Recall



Hyperparameter tuning map



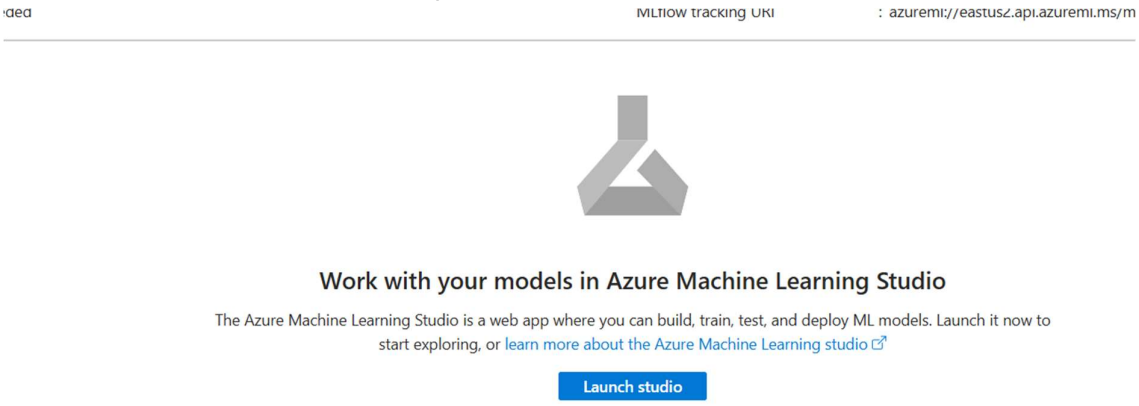
Ref: datainterview.com

Machine Learning with Azure ML Studio And Azure Cloud

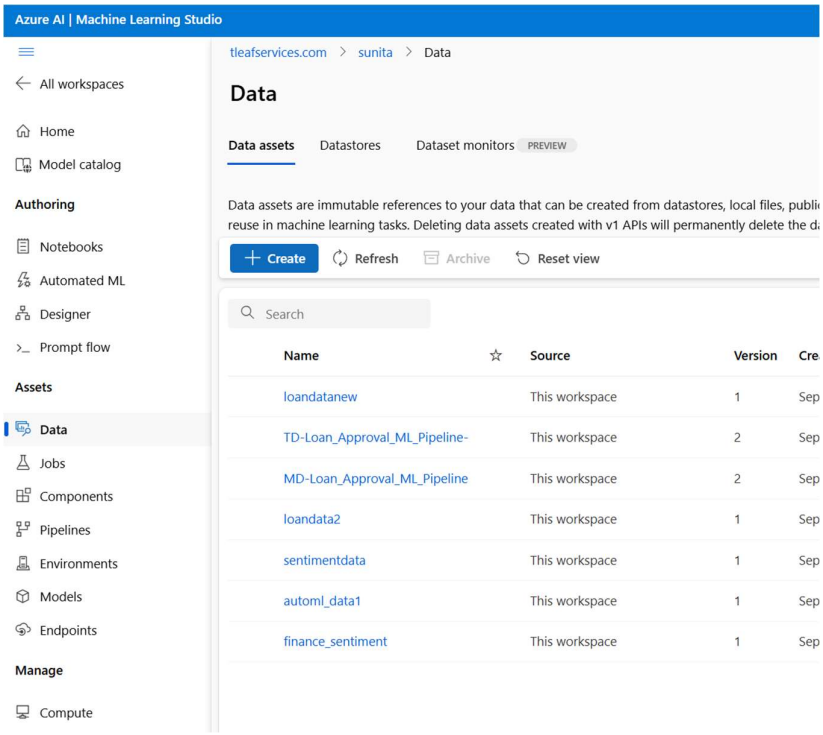
Azure ML Workspace and ML Studio can be used to model a ML pipeline with no-code capabilities, including Auto-ML

Steps involved

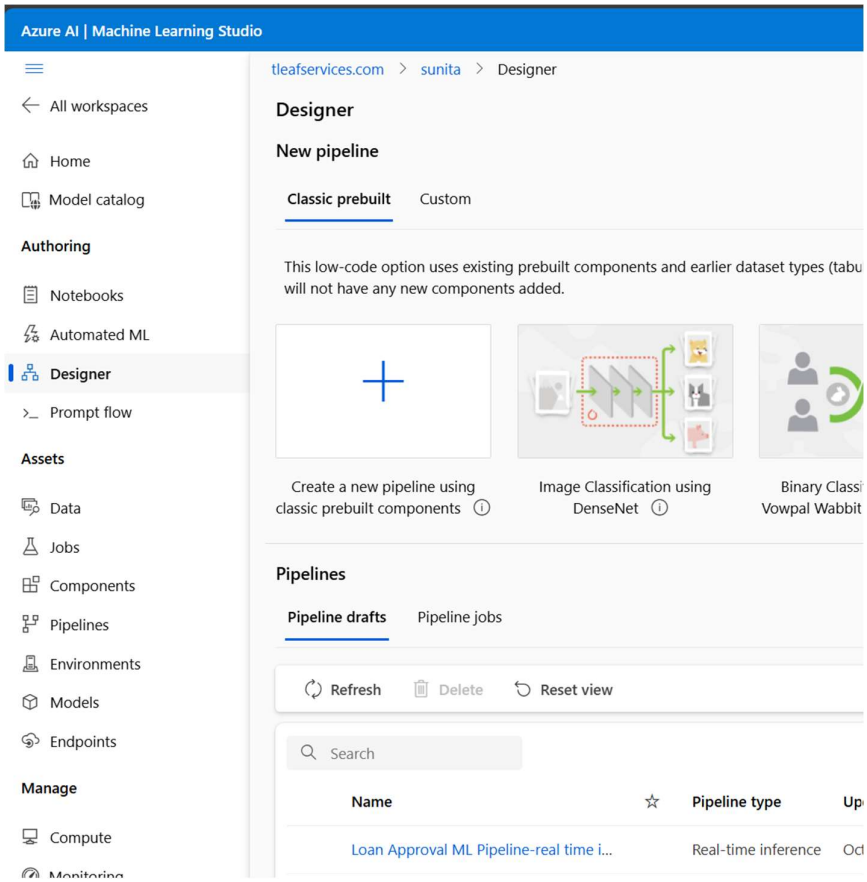
- Create ML Workspace
- Launch ML Studio from the workspace



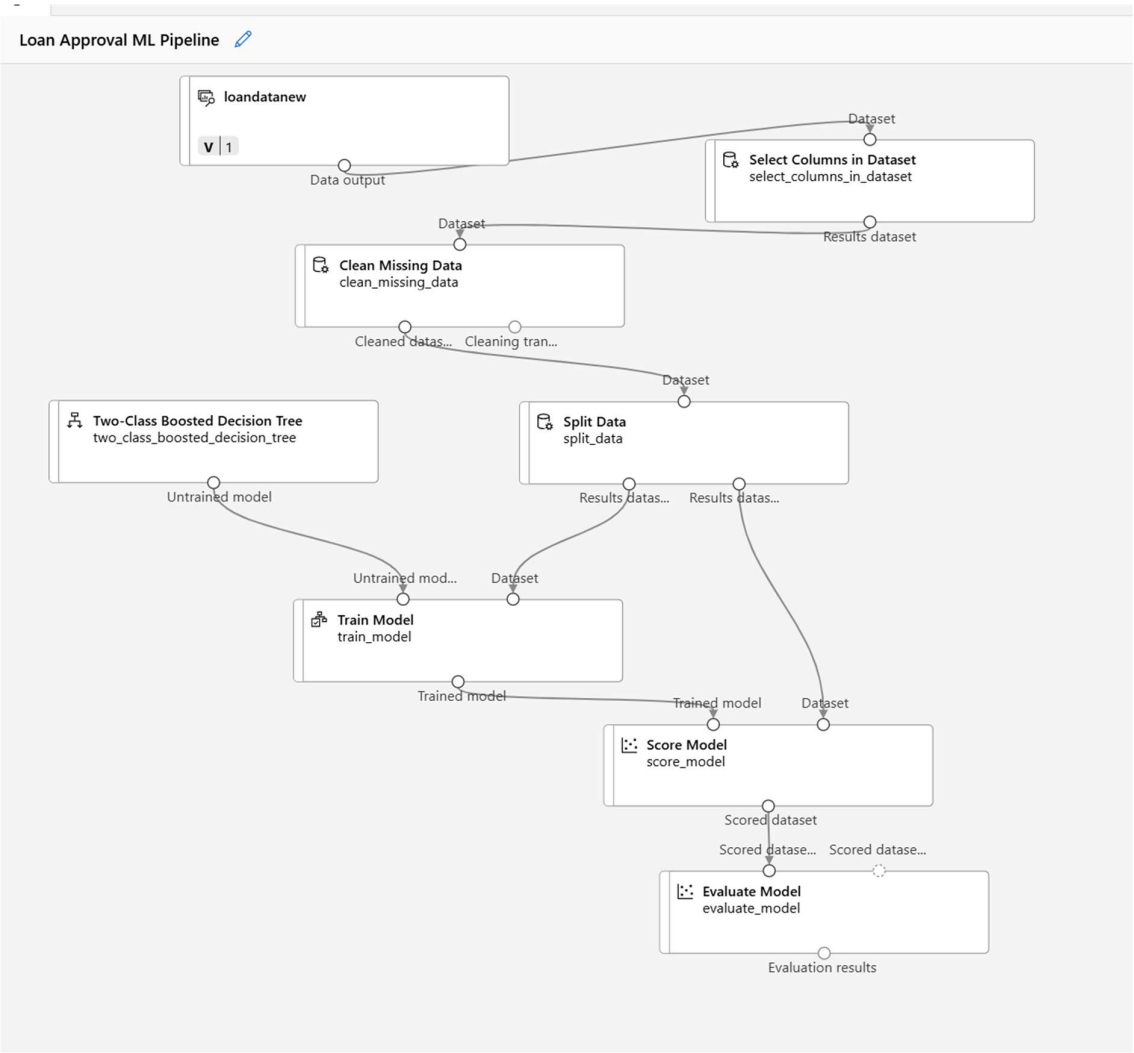
- In ML Studio:
  - Import the dataset



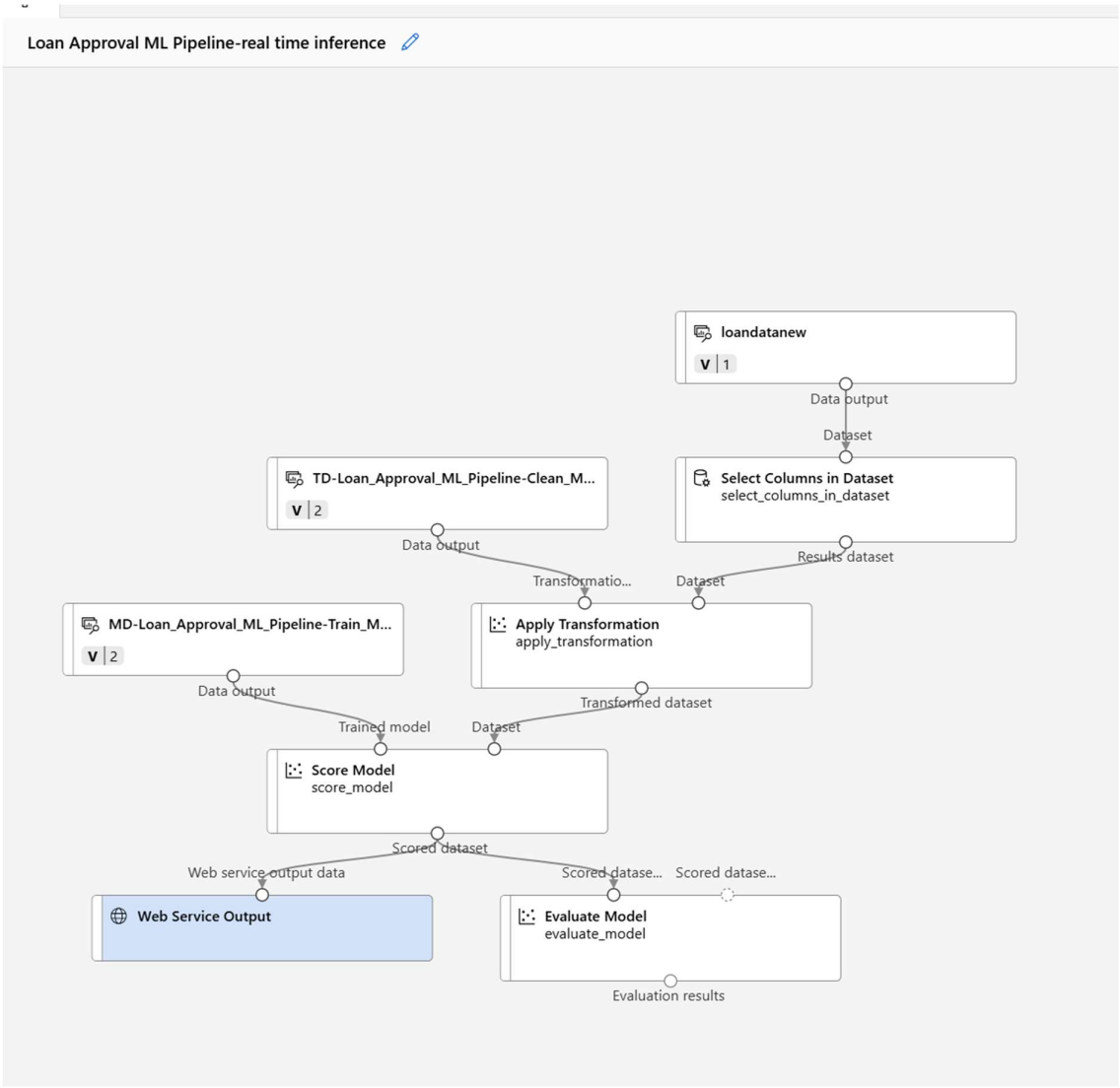
- Create a ML Pipeline



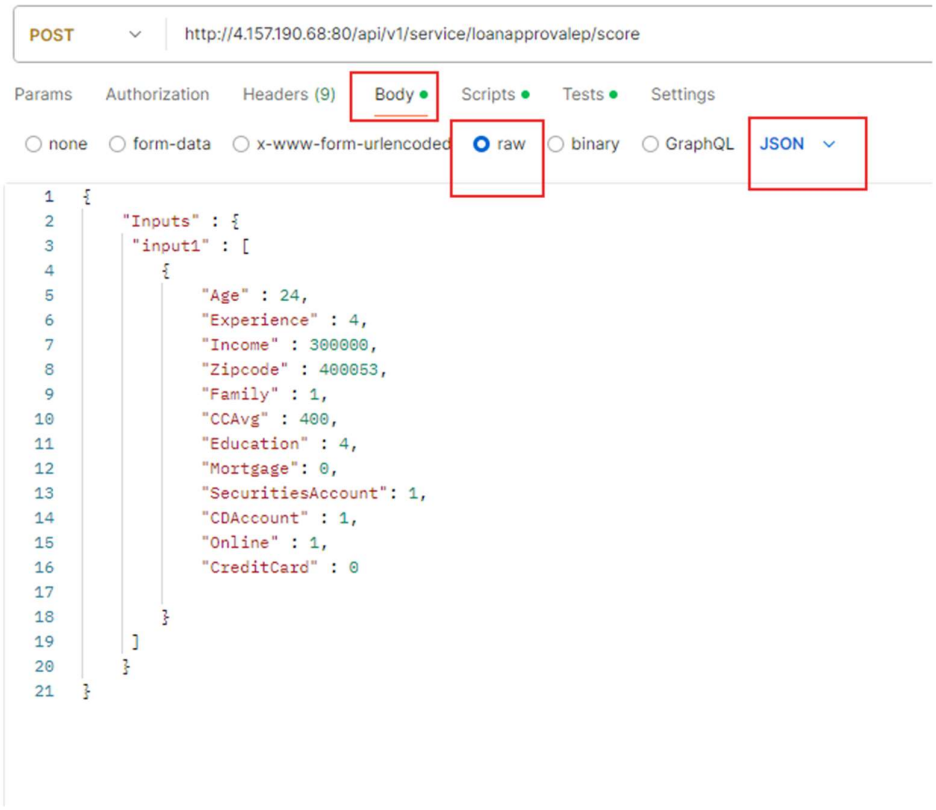
-
- Design the training pipeline



- Execute it
  - Output will be a trained model
- Create an inference pipeline that will expose the Model as an API web service



- Test the web service with postman or any other API tool



Data Science and Machine Learning References

Learning Python

https://github.com/jerry-git/learn-python3/tree/master/notebooks/beginner/notebooks

https://www.kaggle.com/code/chats351/introduction-to-numpy-pandas-and-matplotlib

For deeper understanding of data science and machine learning, recommend following reading:

1) Analytics Vidhya course (free)
   https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

Google Courses (free):
https://developers.google.com/machine-learning/crash-course/linear-regression

Top 4 Linear Regression variations:

https://towardsdatascience.com/top-machine-learning-algorithms-for-regression-c67258a2c0ac

Top 6 ML Algorithms for classification:

https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501

ML A-Z

https://drive.google.com/drive/folders/1OFNnrHRZPZ3unWdErjLHod8Ibv2FfG1d

TensorFlow Bootcamp

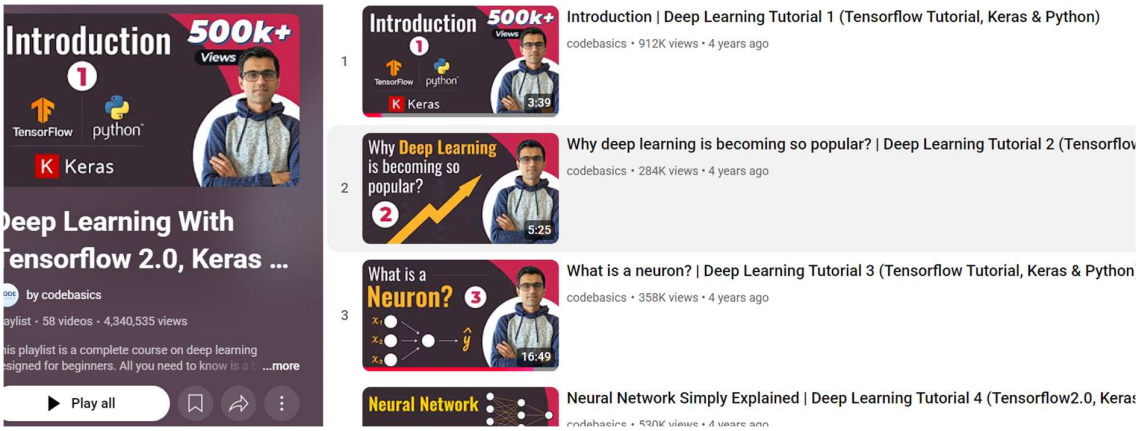
https://drive.google.com/drive/folders/1rXrgUzzIdsyJ4xp05Suq7ioR5q1tOtFY

Deep Learning

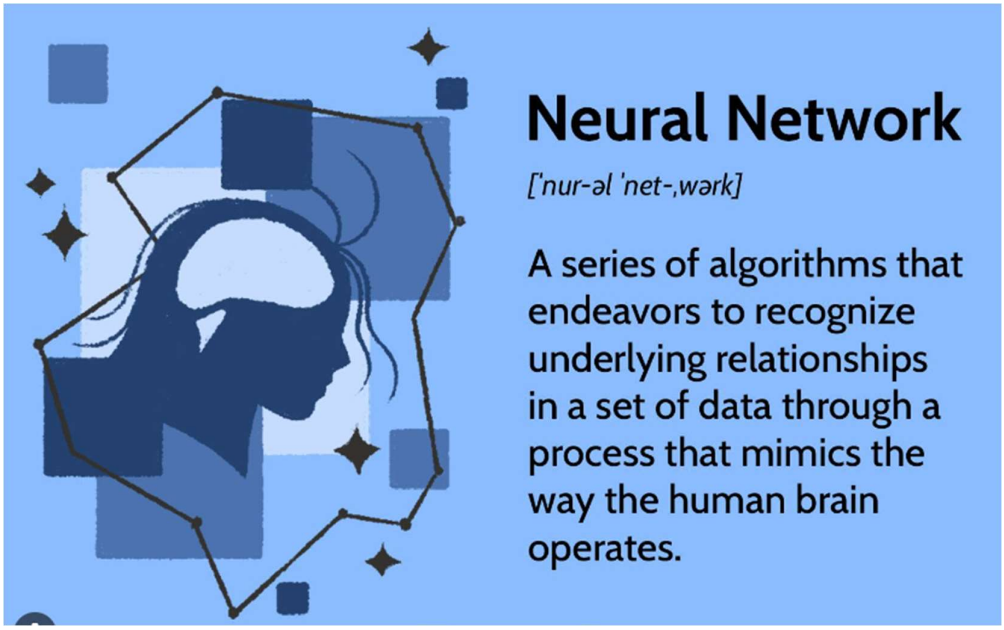
Deep learning is a subset of machine learning that uses **neural networks** to model and solve complex problems. It is particularly effective for tasks involving large volumes of data and unstructured data types such as images, audio, and text.

This playlist is good to learn it in depth:

https://www.youtube.com/playlist?list=PLeo1K3hjS3uu7CxAacxVndI4bE_o3BDtO

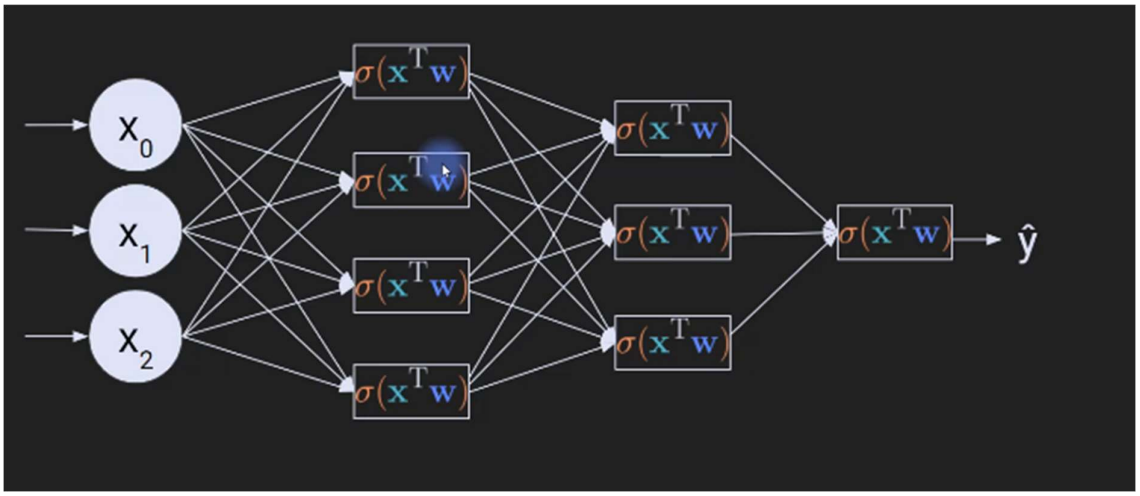


Neural Networks

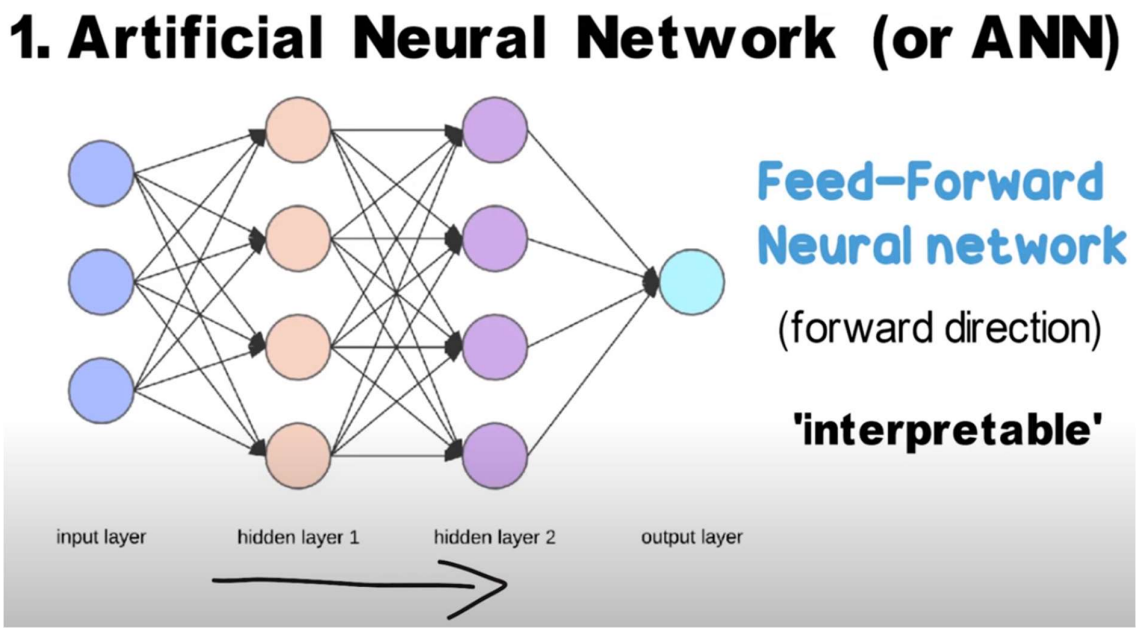


Key Components Of A Neural Network

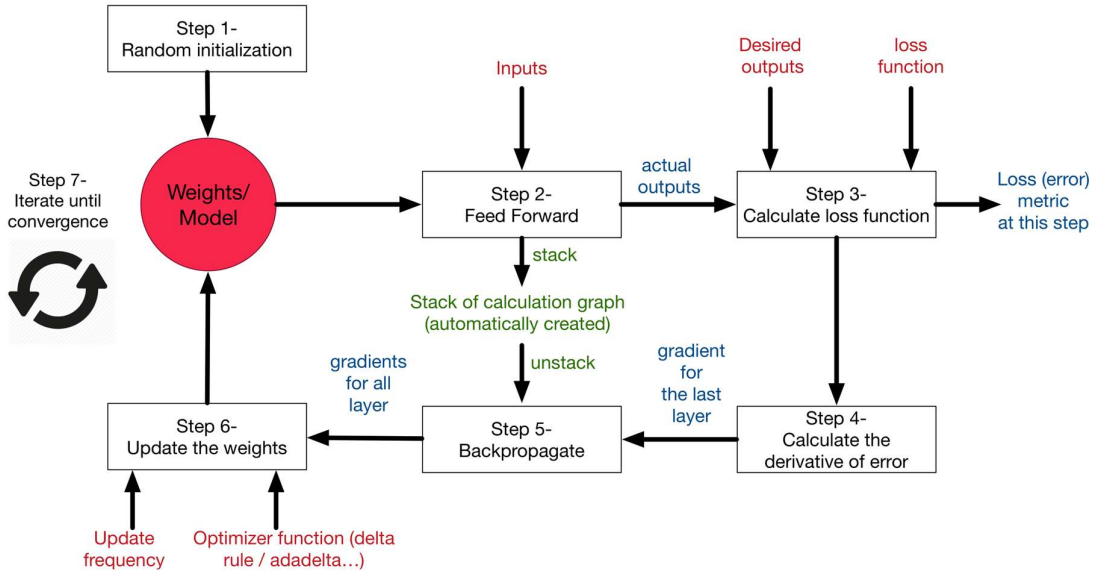
Nodes, Layers, Weights

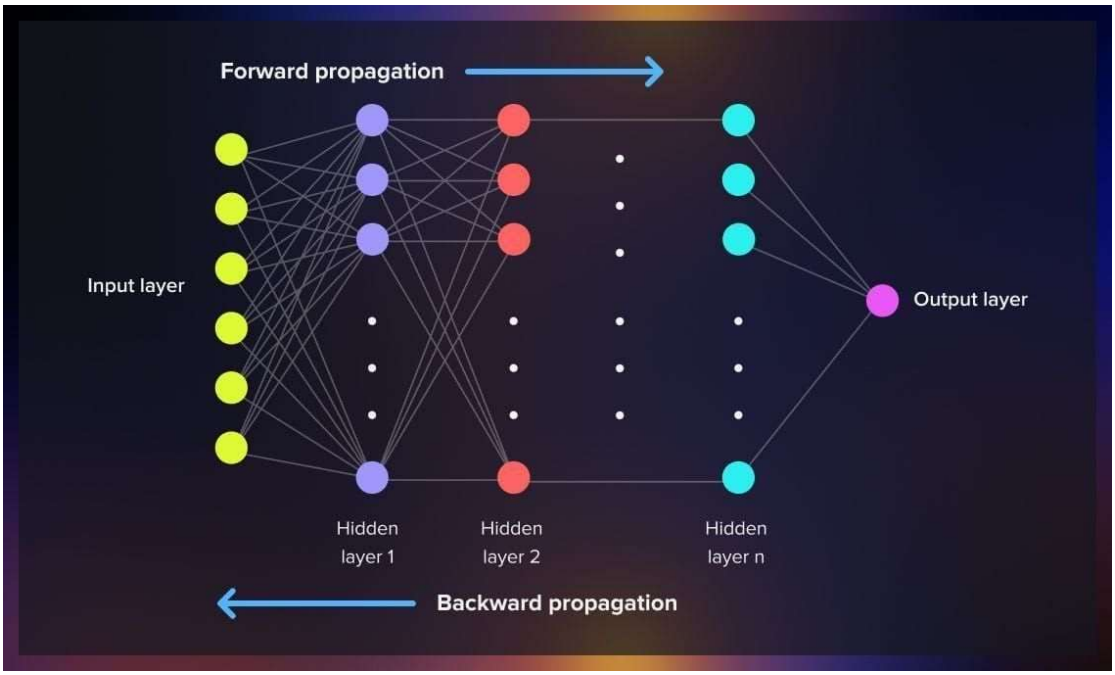


Artificial Neural Network (ANN)



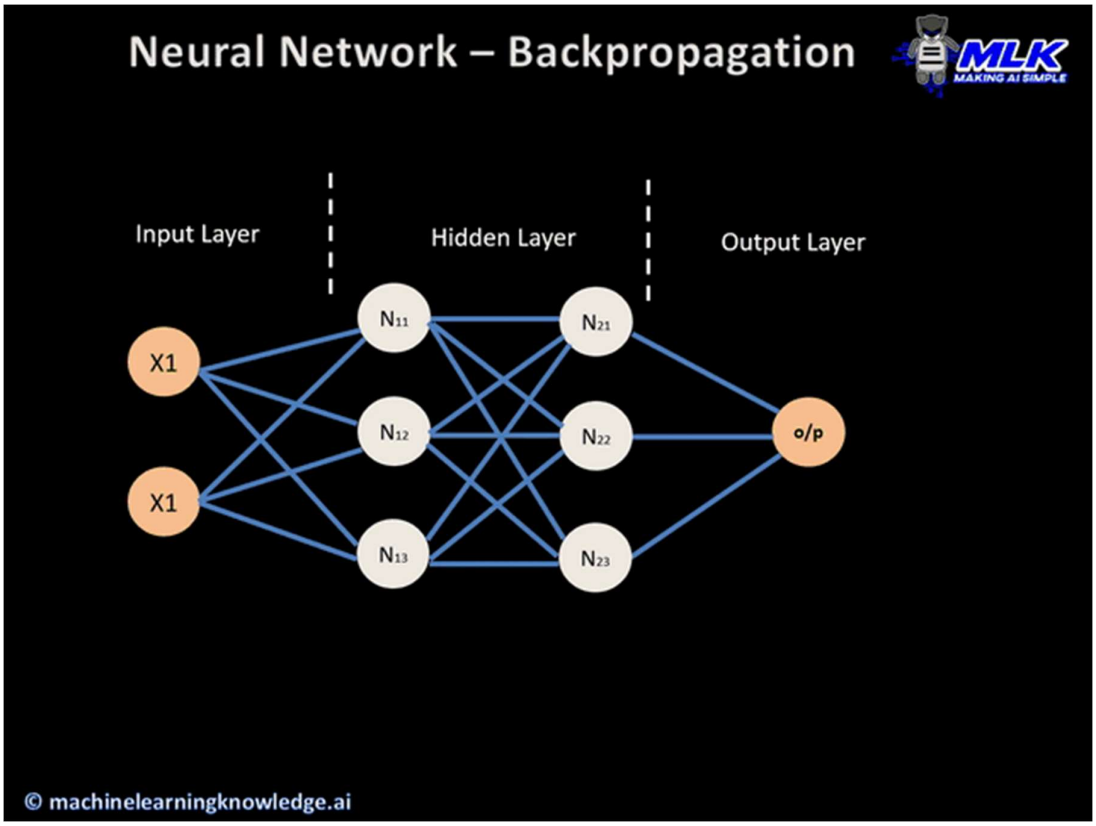
Steps In Neural Network Processing





NN Visualization: https://alexlenail.me/NN-SVG/index.html

Backpropagation



https://www.youtube.com/watch?v=Ilg3gGewQ5U&t=734s

Mastering Backpropagation:

https://www.datacamp.com/tutorial/mastering-backpropagation

Gradient Descent

Gradient Descent in 3 minutes

https://www.youtube.com/watch?v=qg4PchTECck

Gradient Descent Visualizer

https://uclaacm.github.io/gradient-descent-visualiser/#playground

Key use cases

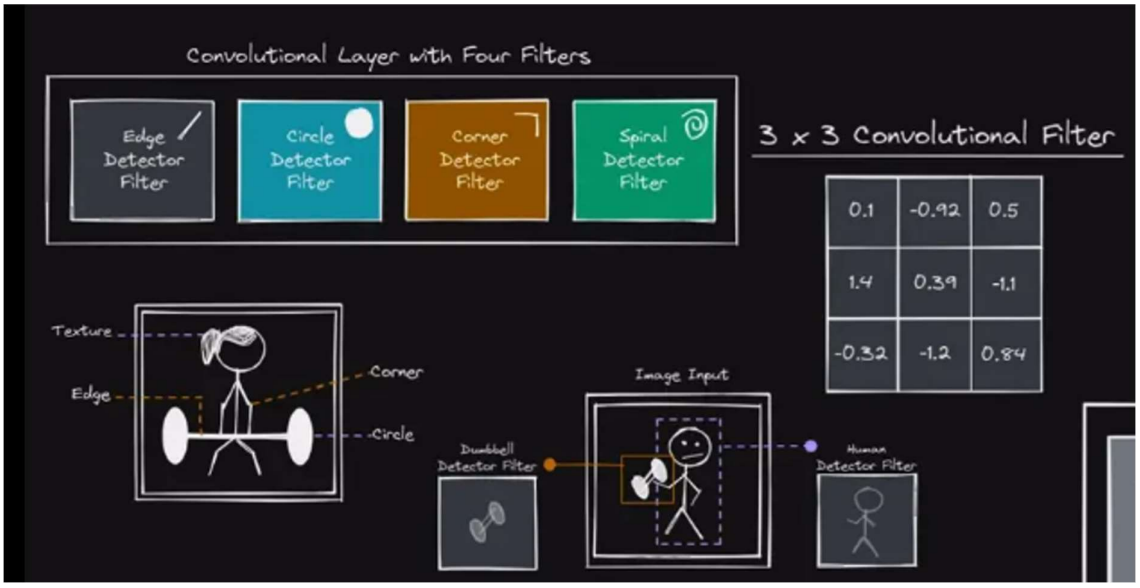
Artificial Neural Networks	Used for Regression & Classification
Convolutional Neural Networks	Used for Computer Vision
Recurrent Neural Networks	Used for Time Series Analysis
Self-Organizing Maps	Used for Feature Detection
Deep Boltzmann Machines	Used for Recommendation Systems
AutoEncoders	Used for Recommendation Systems

- Text classification using NLP (RNN, LSTM)
- Computer Vision (object detection) (Convolutional Neural Networks - CNN)

Computer vision: Convolutional Neural Network (CNN)

https://youtu.be/zfiSAzpy9NM?si=1jJF_sQJqF9HYdBn

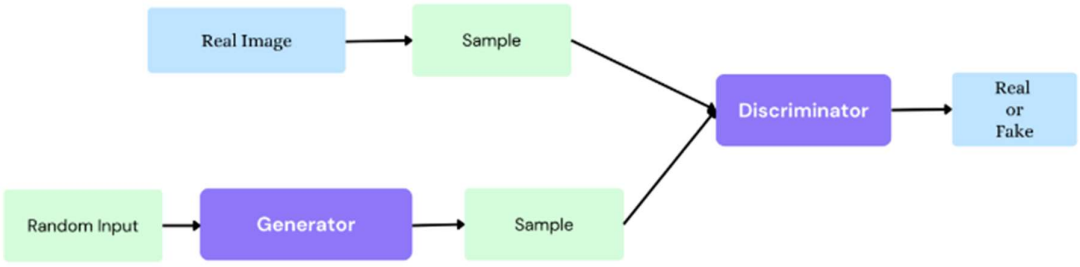
https://adamharley.com/nn_vis/



Generative Capabilities:

2014	- VAE - Encoder-decoder - GAN
2015	- Attention - U-NET - Pointer Network - NeuralStyle
2016	- ResNet - Pix2Pix
2017	- WGAN - WGAN-GP - CycleGAN - Transformers
2018	- MuseGAN - ProgressiveGAN - WorldModels - Self Attention GAN - BigGAN
2019	- BERT - StyleGAN - GPT-2 - MuseNET
2020	- GPT-3
2021	- DALL-E - CLIP .....
2025	- GPT-7

Generative Adversarial Network (GAN)

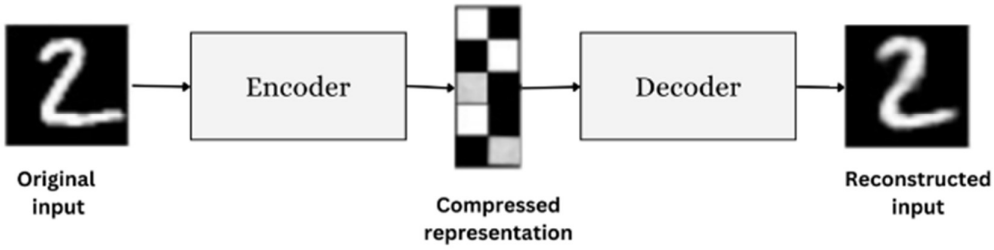


GAN architecture diagram

https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29

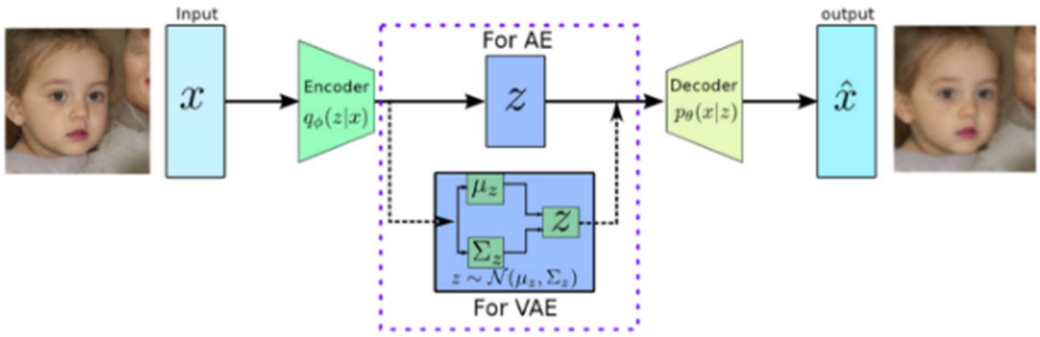
Variable Auto Encoder (VAE)

https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73



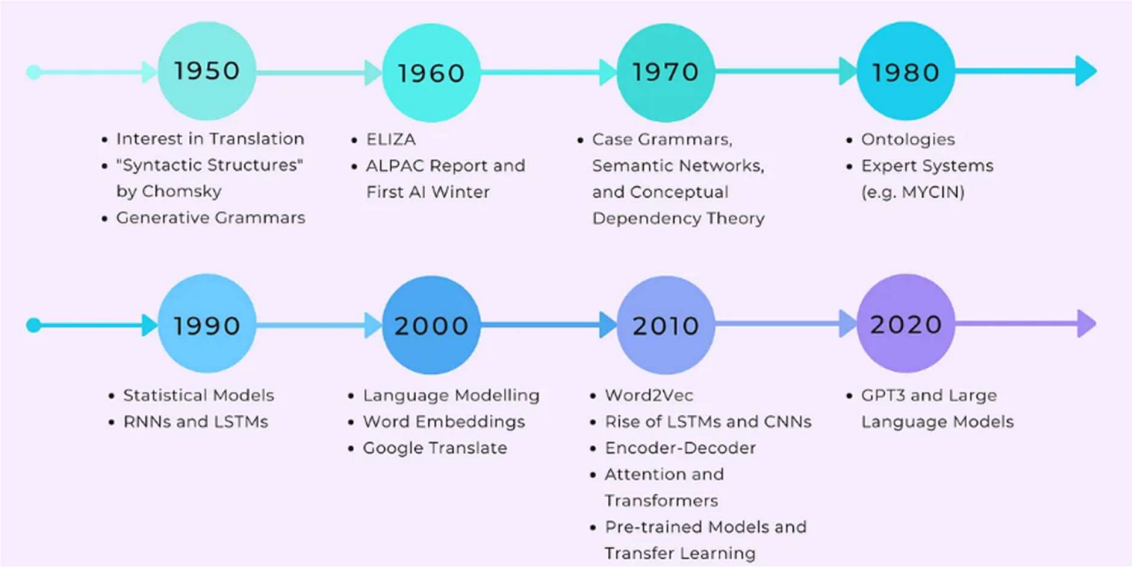
-
-
- VAEs aim to learn a probabilistic mapping between the data space and a latent space. This capability allows them to generate new samples closely resembling the patterns observed in the training data.
-
- GANs are designed to generate realistic data by training a generator network to produce data that is identical to real data.
- GANs are more popular for image generation
- Autoencoders are used for feature learning, compression, reconstruction data and minimize error between input and output.
- Both VAEs and autoencoders use a reconstruction loss function to tune the neural networks using gradient descent.
-
- The distinction between regular autoencoders (AEs) and variational autoencoders (VAEs) lies in how they handle the latent representation. In conventional autoencoders, the encoder transforms an input into a predetermined and fixed latent vector. On the other hand, in variational autoencoders, the encoder generates not a fixed latent vector but the parameters defining a probability distribution, typically a Gaussian distribution.
- The concept described above is visually depicted in the following figure.

-



-
-
- https://ubiai.tools/comparing-gan-autoencoders-and-vaes/
- https://medium.com/@ogre51/difference-between-autoencoder-ae-and-variational-autoencoder-vae-73004cc7e0fb
-

Natural Language Processing (NLP)



Tokenization

Divide the texts into words or smaller sub-texts, which will enable good generalization of relationship between the texts and the labels.

This determines the “vocabulary” of the dataset (set of unique tokens present in the data).
gpt-3.5 has a vocabulary size of ~50000 tokens

Tokens	Characters
21	105

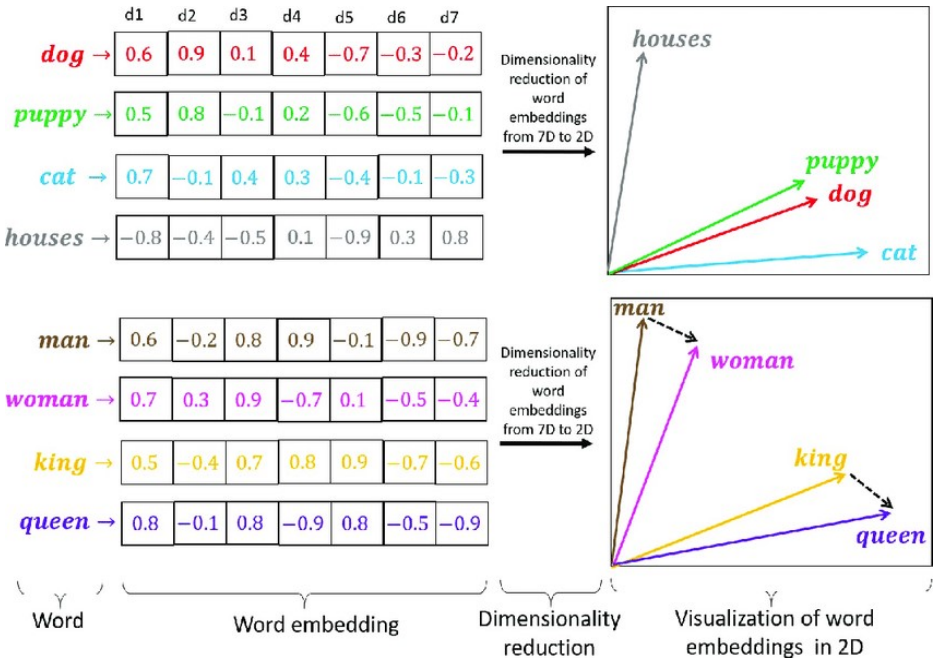
Einstein is arguably the greatest scientist of all times. He was born in germany and lived in switzerland

Vectors

Sequence of numbers - a mathematical representation for an object such as a word.

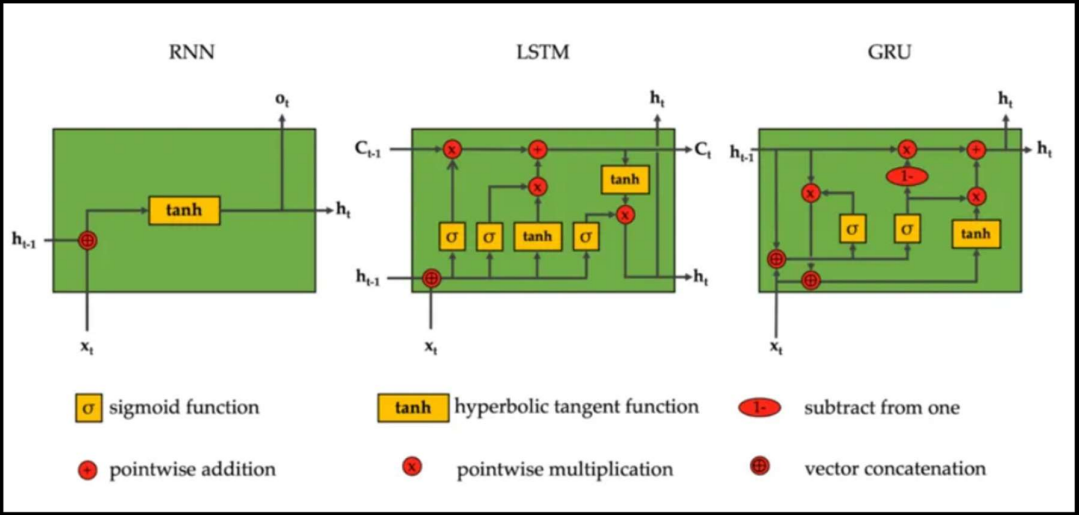
Embeddings

The sequence of numbers capture the meaning of words, the relationship between words, and the context of different words as they are used naturally. Each number in the sequence represents a dimension, such as gender



RNN, LSTM and GRUs

RNN, LSTM and GRU



- Concept of limited memory (state) enabling sequential data processing
- Examples: Timeseries analysis, text processing, Audio processing
- Enabled NLP (Natural Language Processing) use-cases

ANN	CNN	RNN
Tabular or Text Data	Image Data	Sequence data
No Parameter Sharing	Yes	Yes
Operate on Fixed Length input	Operate on Fixed Length input	Don't
No Recurrent Connections	No Recurrent Connections	They are Possible
No Spatial Relationships	They are Possible	No Spatial Relationships
ANN is considered to be less powerful than CNN, RNN	CNN is considered to be more powerful than others	RNN includes less feature compatibility when compared to CNN
Having fault tolerance, Ability to work with incomplete knowledge	High accuracy in image recognition problems, & weight sharing	Remembers each and every information, & offers time series prediction

Ref: https://arxiv.org/pdf/2401.02843.pdf

Challenges With NLP

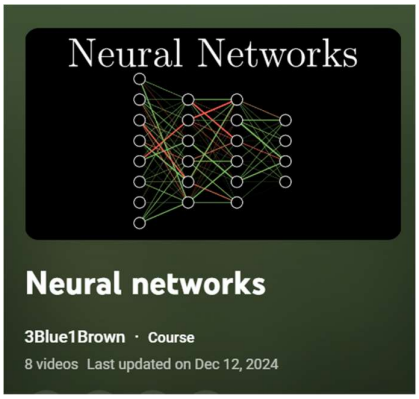
The word "**bank**" in the following two sentences has **different meanings**, but since they are assigned to the **same token id**, their word **embedding** vectors are the **same**.

1. "I went to the **bank** to withdraw some money."
2. "I went to the other side of the river **bank**."

- Earlier neural networks such as RNN, LSTM, GRU had limited memory / capacity to process context
- Suffered from problems such as vanishing gradient, exploding gradient

Deep Learning References

3Blue1Brown – Neural Networks playlist:



https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

https://www.youtube.com/playlist?list=PLqwozWPBo-FvuHWx3_aYwG2WVdbb-wC6q

https://www.youtube.com/playlist?list=PLeo1K3hjS3uu7CxAacxVndI4bE_o3BDtO

https://ml-playground.com

https://playground.tensorflow.org/

https://www.freecodecamp.org/news/deep-learning-neural-networks-explained-in-plain-english/

Deep Learning book by Ian Goodfellow, creator of G

https://www.deeplearningbook.org/

Variable Auto Encoders

https://github.com/Jackson-Kang/Pytorch-VAE-tutorial

How Stable Diffusion Works (AI Image Generation)

https://www.youtube.com/watch?v=sFztPP9qPRc

Interactive Node link visualization:

https://adamharley.com/nn_vis/

Enhanced credit card fraud detection based on attention mechanism and LSTM deep model:

https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00541-8

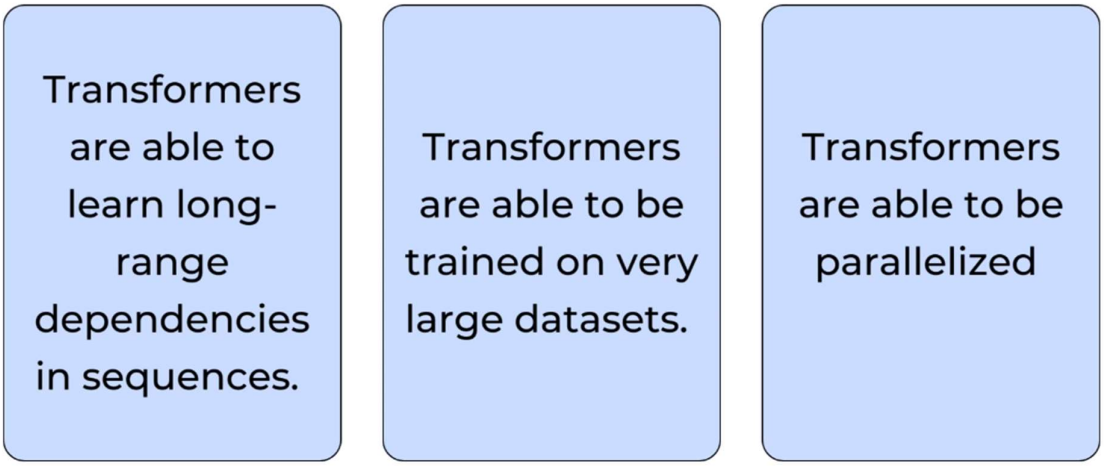
Vehicle marketplace recommendation engine with DL:

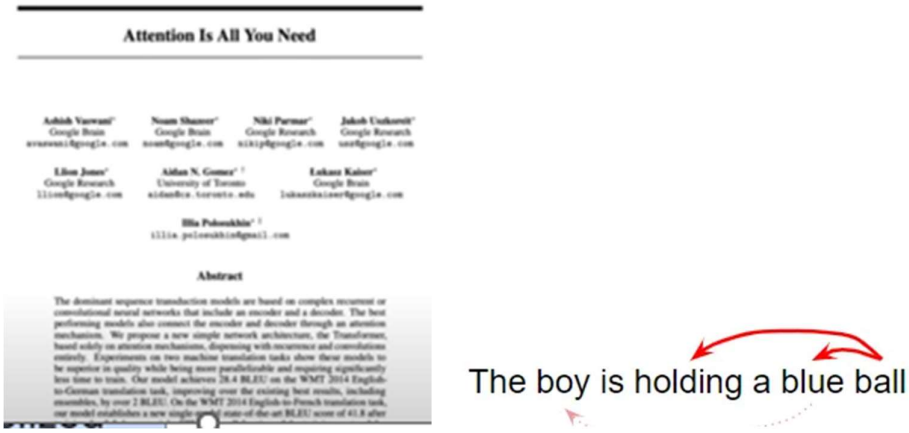
https://slideplayer.com/slide/14614243/

The Transformer Model

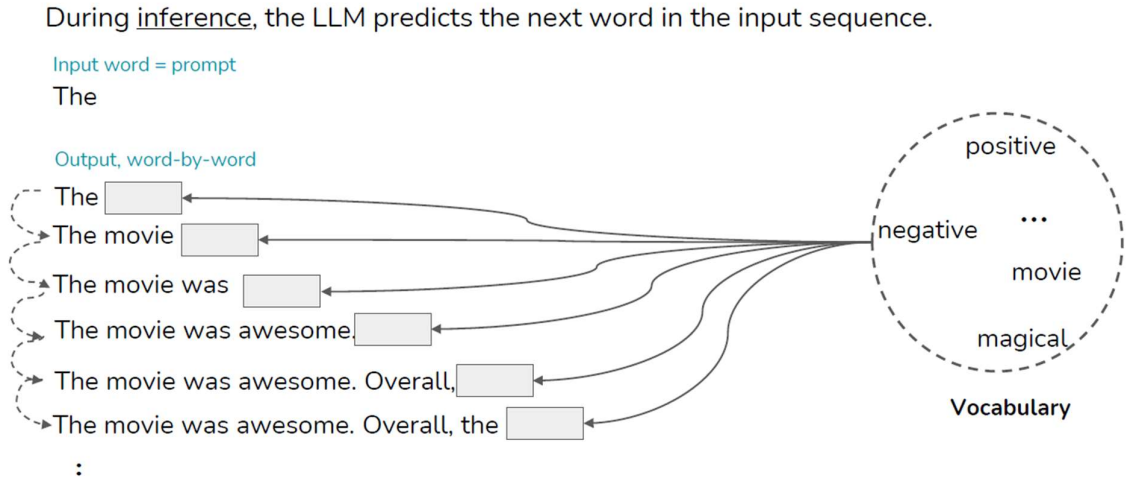
Transformers are a type of neural network that have a **unique ability to recognize long-range connections within sequences**. They are particularly useful for **tasks like generating text**, as the model needs to comprehend the preceding words in order to produce the next one.

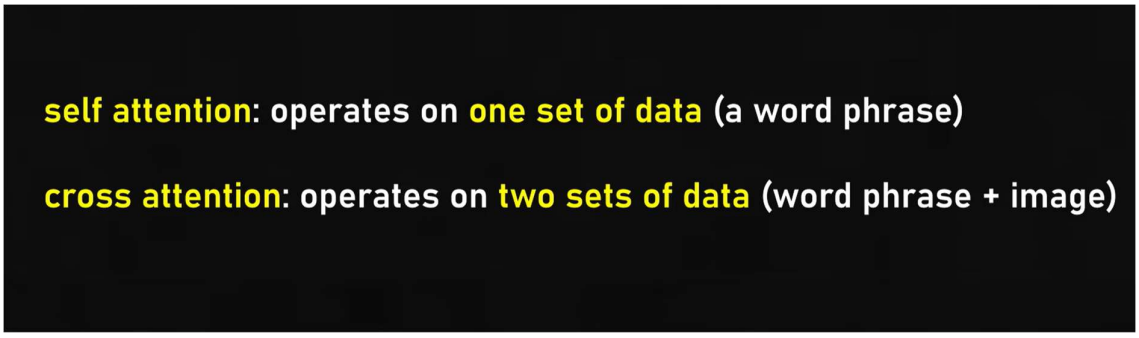
The introduction of transformers in 2018 was a groundbreaking moment for the field of natural language processing.





Transformers: Next token prediction

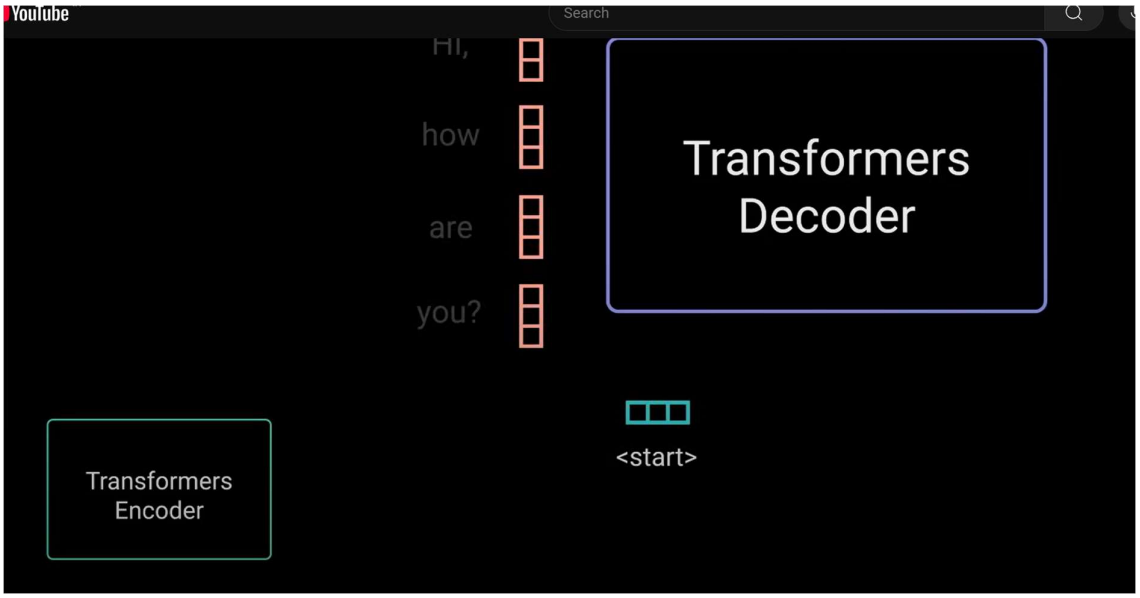




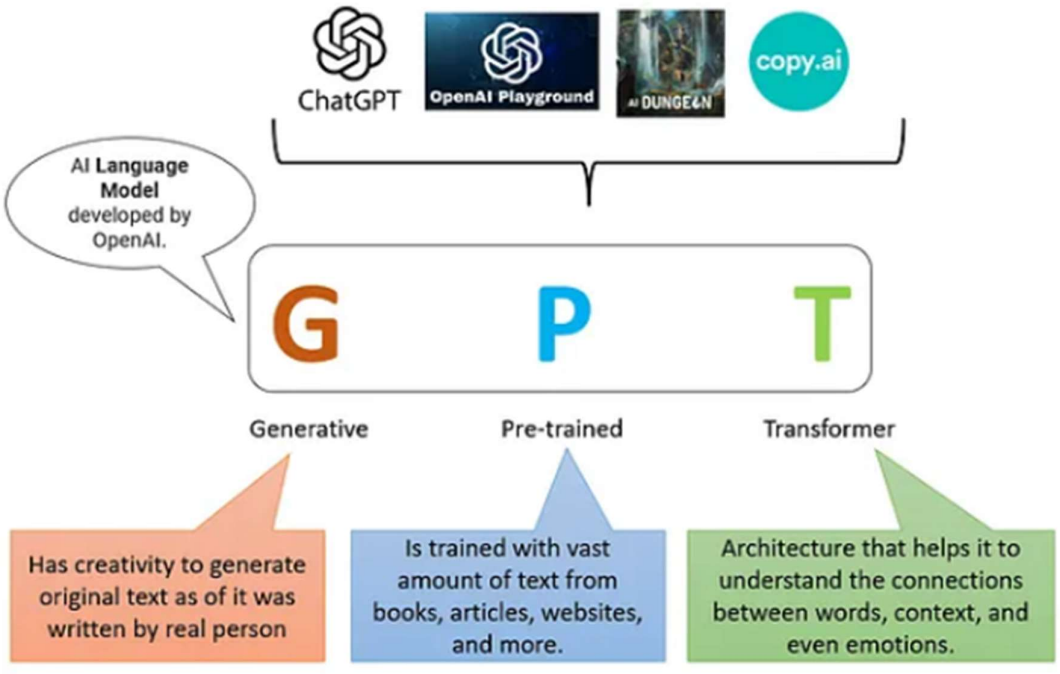
Transformer Architecture

AI Hacker: Illustrated Guide to Transformers Neural Network: A step by step explanation

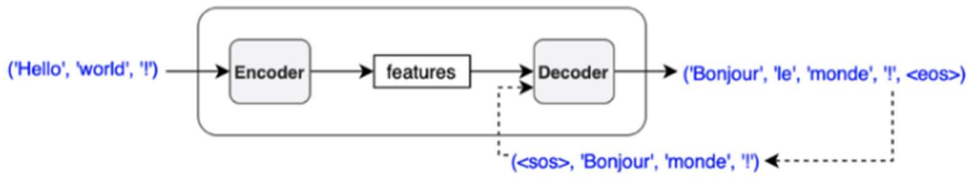
https://www.youtube.com/watch?v=4Bdc55j80l8

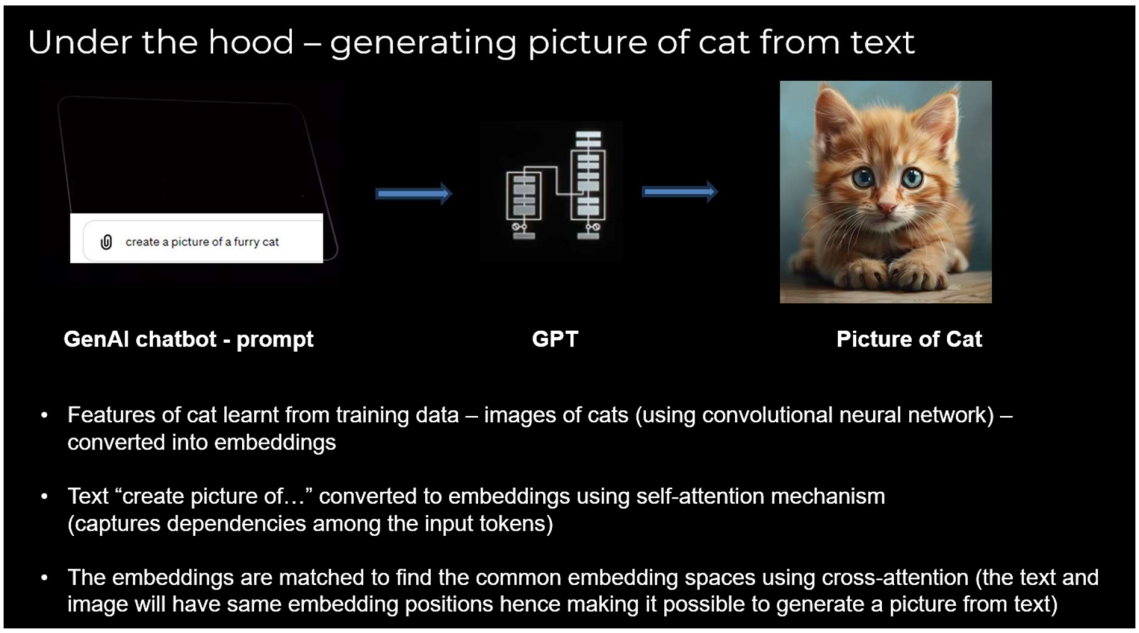


GPTs and LLMs



Transformers: Encoders and Decoders

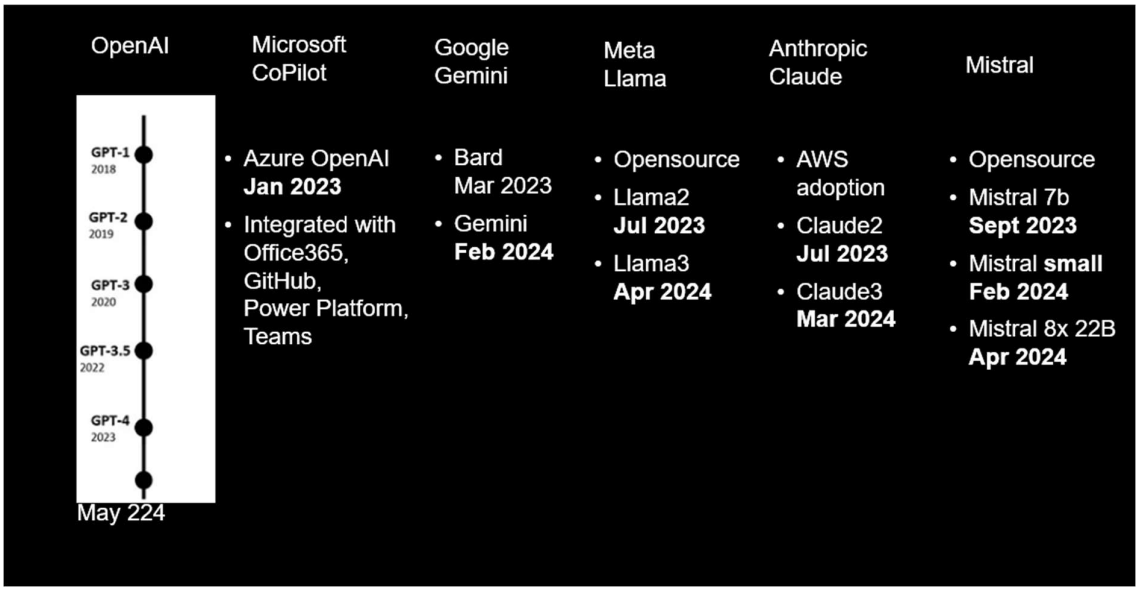




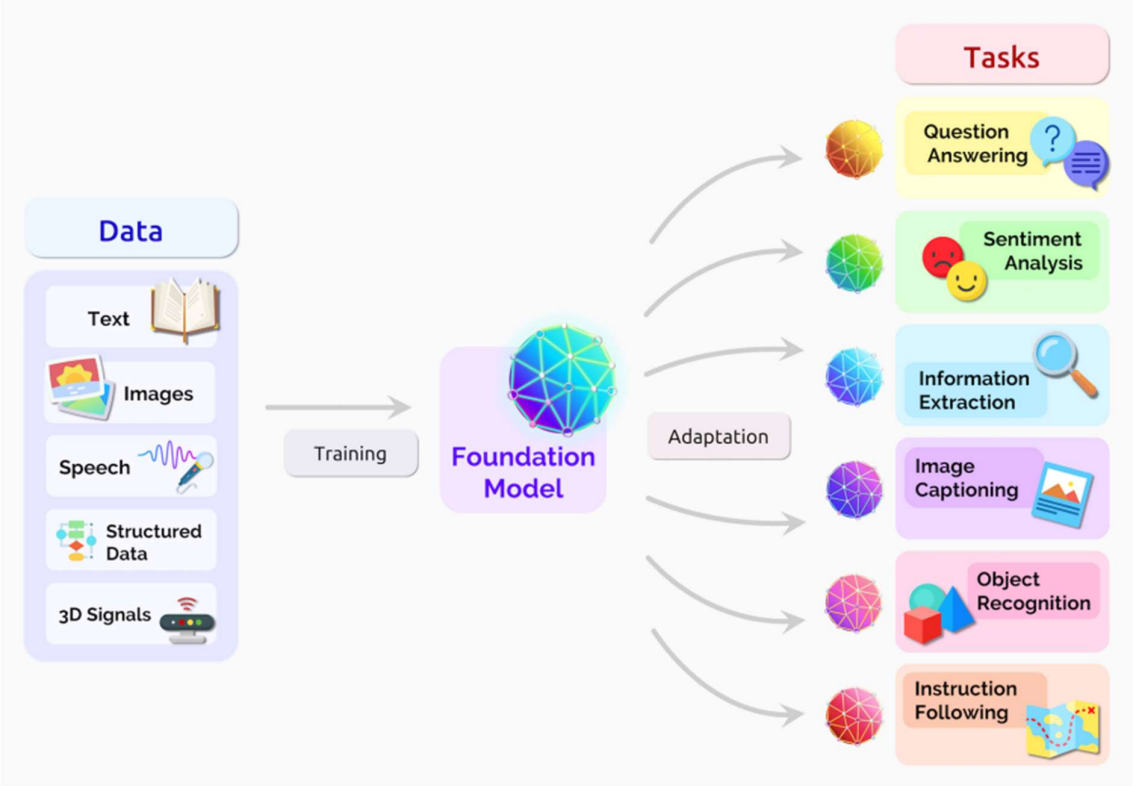
Large Language Models (LLMs)

- Pre-trained Deep Learning models
- Trained on massive textual data (tokens)
- Data source
    - Online contents and literature
    - News and current affairs
    - Social media
    - Other sources like blogs, comments, feedback, reviews etc.
- Millions of parameters (weights, neurons)
- LLMs can be fine-tuned

- for specific problems
- for specific domains
    - Using smaller datasets



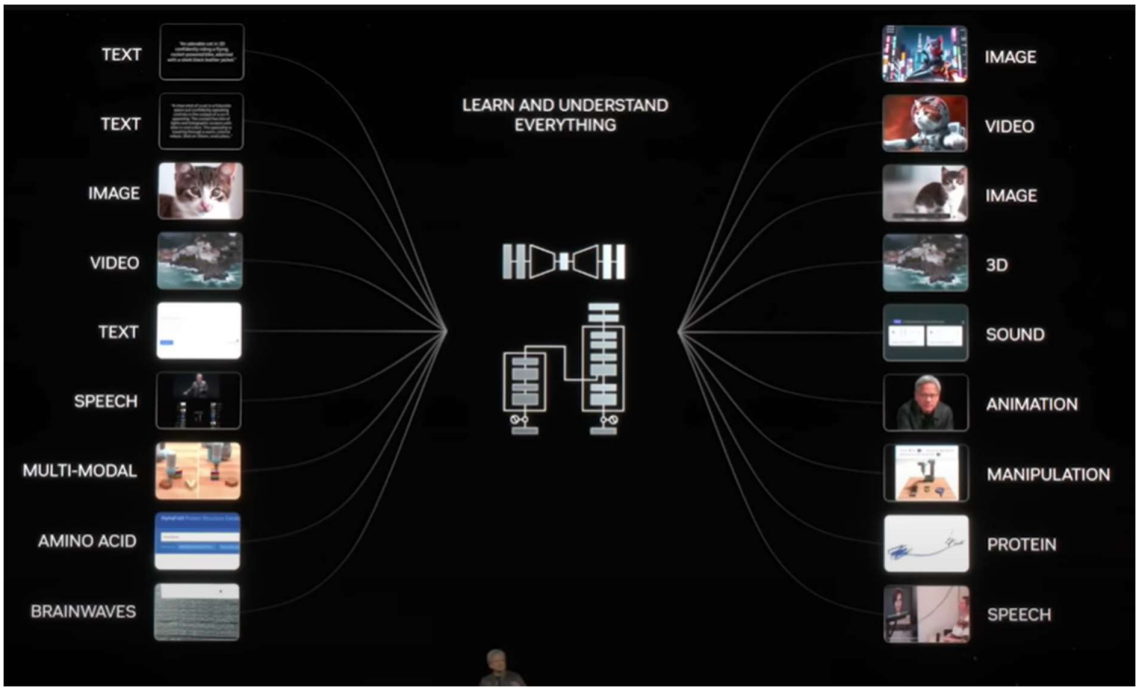
Foundation Models



Multi Modal

multi-modal refers to AI systems that can process, generate, or interact with multiple types of data modalities, such as text, images, audio, video, and more.

Multi-modal AI models are designed to integrate and understand relationships across these different modalities, enabling them to perform tasks that involve combinations of data types.



Gen AI Core capabilities

Inferencing, Question & Answer based on pre-trained knowledge

Prompt: "What is the capital of France?"
Response: "Paris."

Prompt: A detailed article about climate change.
Response: "Climate change is a significant global issue caused by human activities leading to rising temperatures and environmental impacts."

Translation: Translating text from one language to another.
Prompt: "Hello, how are you?" (English)
Response: "Hola, ¿cómo estás?" (Spanish)

Text Generation: Producing creative content such as stories, poems, or essays.
Prompt: "Once upon a time in a faraway land,"
Response: "there was a small village surrounded by towering mountains and lush forests."

Sentiment Analysis: Determining the sentiment or emotional tone of a text.

Example: Input: "I love this product!"
Output: "Positive sentiment."

Summarization

It can summarize large documents, videos, images, flowcharts, audio – any modality, multiple languages

Extractive Summarization:

The LLM selects and concatenates key sentences or phrases from the original text

Abstractive Summarization:

The LLM generates new sentences that paraphrase the original content.

Practical applications
News summarization, scientific papers summarization, legal documents, product reviews

Reasoning

Reasoning in AI involves higher-level cognitive processes (compared to inferencing) where the AI system makes decisions based on logic, understanding relationships, and drawing conclusions from given information.

Deductive reasoning:

All humans are mortal. Socrates is a human. Therefore, Socrates is mortal.

Inductive reasoning

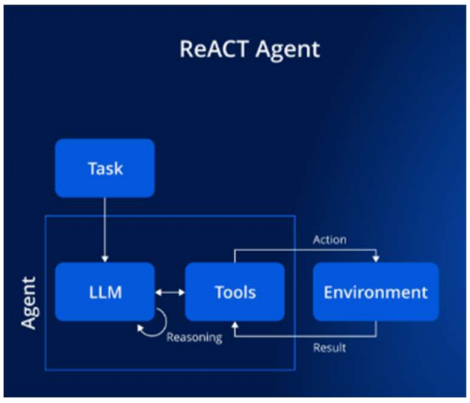
Every time I water my plant, it grows taller. Watering plants regularly helps them grow taller.

Abductive reasoning

Ability to choose best possible inference. Lawn is wet – rain? Or sprinkler?

Practical applications
Analyse market data to analyse patterns, customer behaviour analysis, drug discovery

ReAct – Reasoning and Action



1. **Input**: The agent receives a task description in natural language, which the core LLM processes.
2. **Reasoning**: The LLM breaks down the task into smaller steps, analyzes the situation, considers available information, and plans the necessary actions.
3. **Action**: Based on this reasoning, the LLM selects the appropriate tool (e.g., search engine, database, API) and performs actions to gather information or interact with the environment. This could involve querying Wikipedia for facts or retrieving data from a company database.
4. **Observation**: The agent observes the results of its actions and updates its knowledge. This new information is used to refine its reasoning in the next cycle.
5. **Response**: Finally, the agent generates a response based on its reasoning and the information gathered.

https://www.linkedin.com/pulse/react-agents-revolutionizing-ai-reasoning-action-allen-adams-fxqrc/

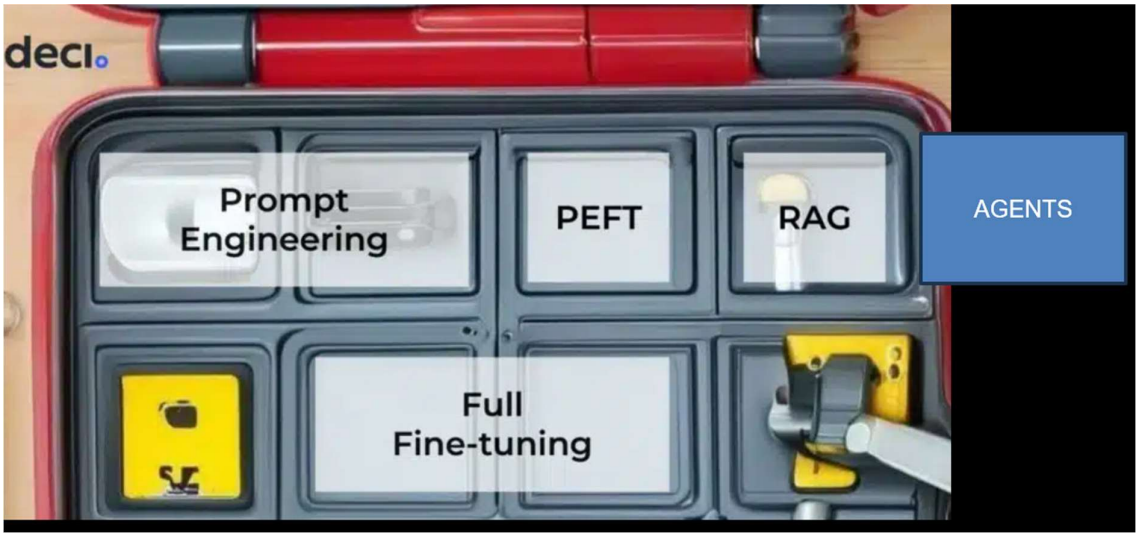
Customizing LLM/GPT behaviours

The need to customize

Why customize?

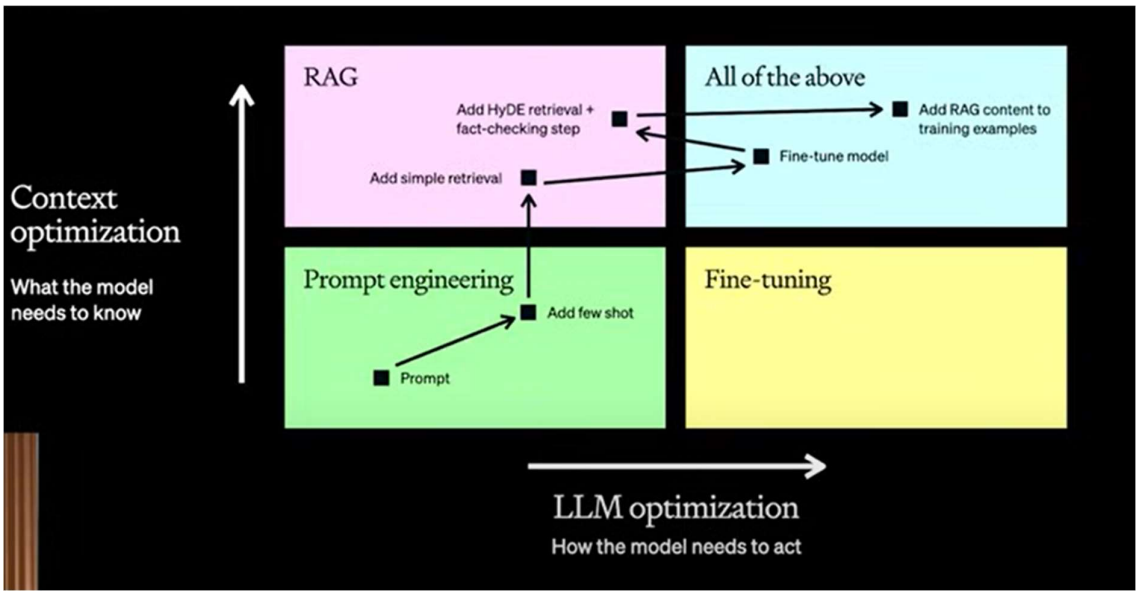
- LLMs lack up to date / real time data
  Their knowledge is restricted till the training date cut-off of the model
- Some use-cases demand models work with proprietary / IP protected / subscription based datasets / documents
- Reduce hallucinations, improve accuracy and response, change the response tone, behaviour

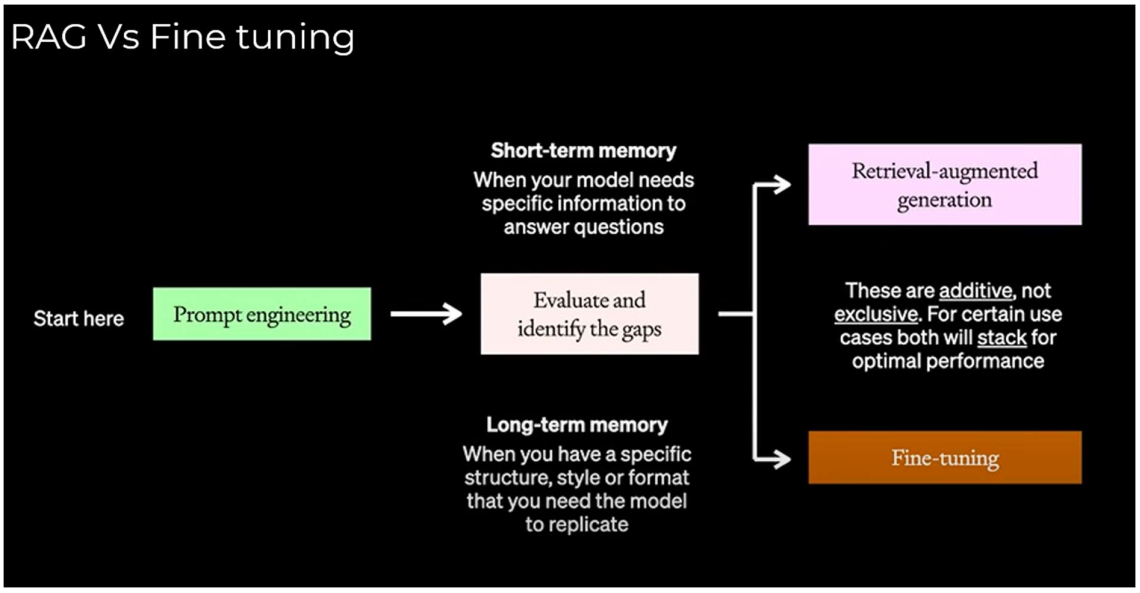
Customization options



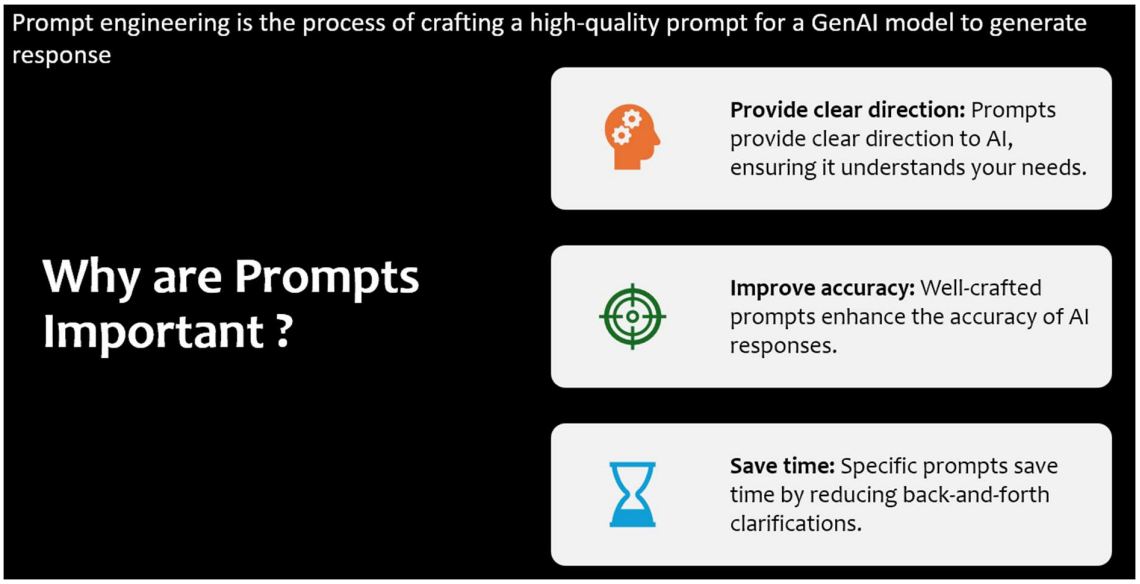
- **Prompt Engineering**
  Refines model input to guide its output.
- **Retrieval Augmented Generation (RAG)**
  Augments LLMs knowledge with custom data.
  Merges prompt engineering with index & retrieval mechanism to create a customized context for the LLM
- **Full Fine-tuning**
  Adjusts all parameters of the LLM using task-specific data.
- **Parameter-efficient Fine-tuning (PEFT)**
  Modifies select parameters for more efficient adaptation.
- **Agents**
  Agents enable tools and functions that can be invoked by the LLM to fetch specific data. LLM acts as a ReAct (Reasoning and Action) engine that can invoke any tool on demand to complete a certain assigned task (prompt)

A combination can be leveraged:

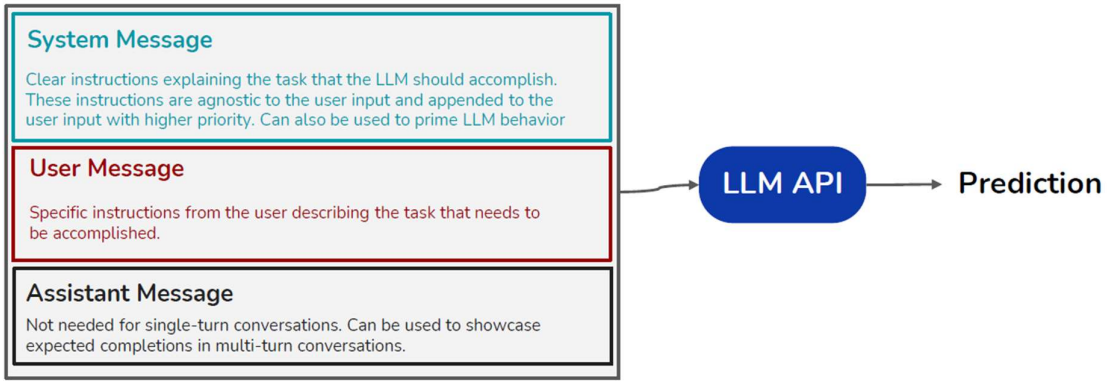




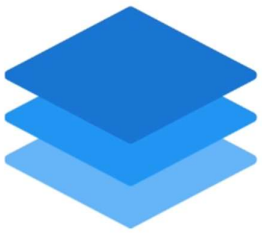
Prompt Engineering



Azure Open AI APIs are compatible with the Open AI APIs and have the following three components.



Zero, One, and Few-shot Prompting



Gives task directly

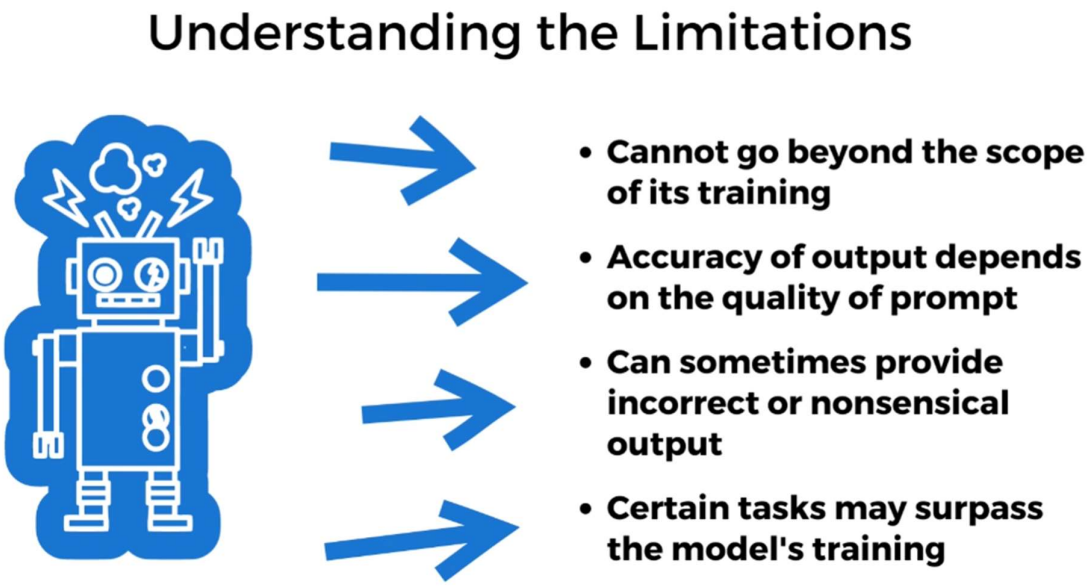
e.g. translate "this cat is sleeping" into French

Provides one example

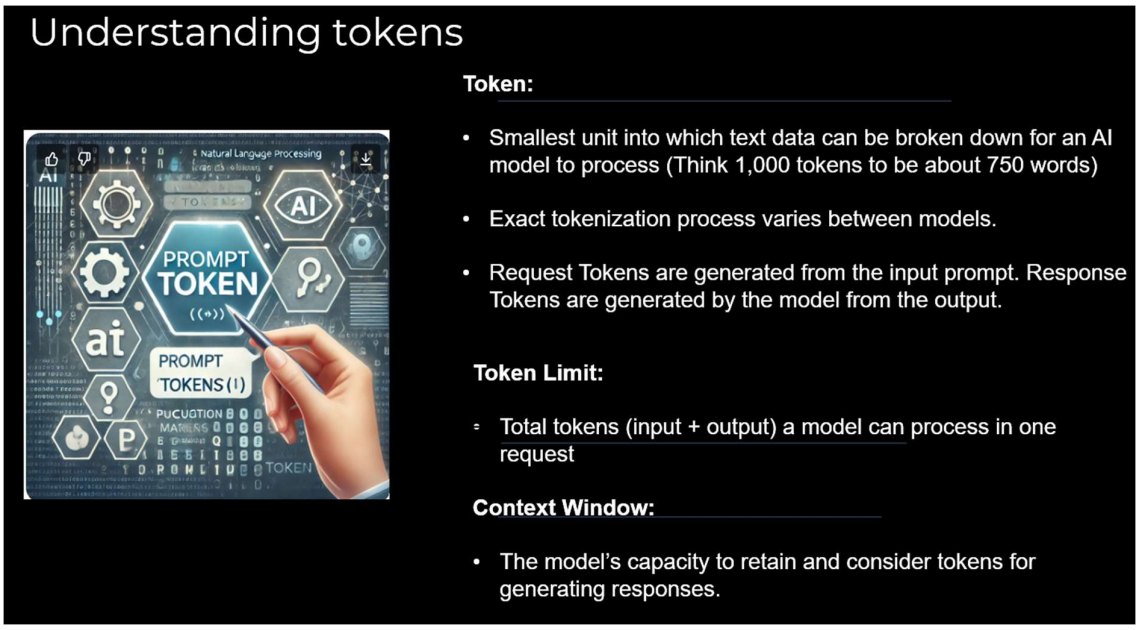
e.g. continue this story, here is an example

Provides multiple examples

e.g. write a review based on these samples



Understanding Tokens

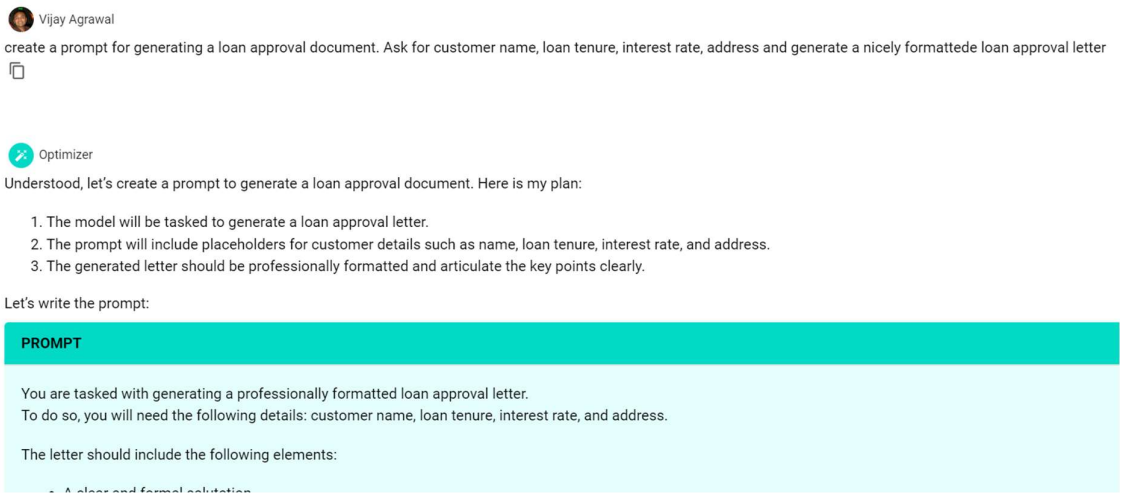


Tokenizer tool:

https://platform.openai.com/tokenizer

Prompt Engineering Tools

1. PromptPerfect: Create a perfect prompt by describing the use-case:



Tool URL:

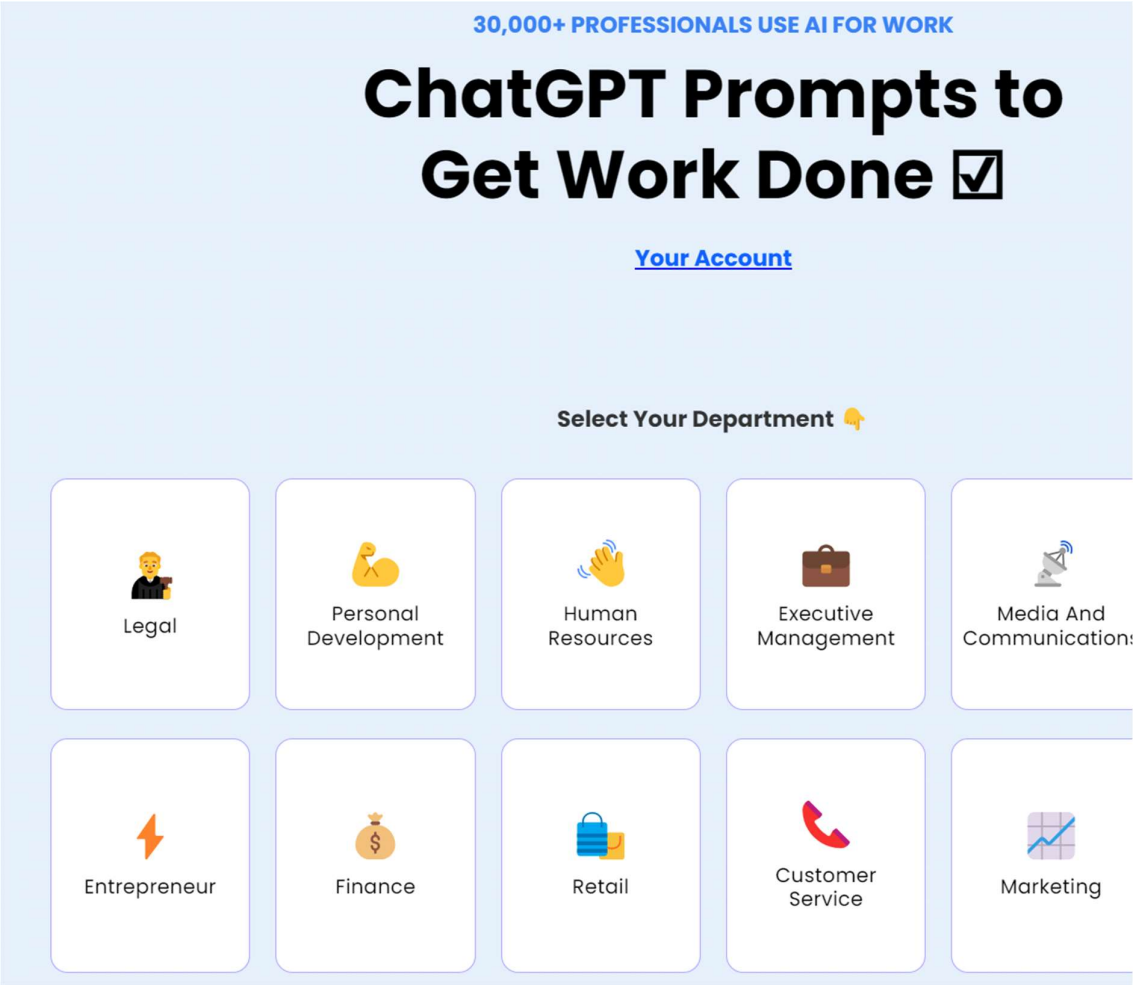
https://promptperfect.jina.ai/

2. AI For Work

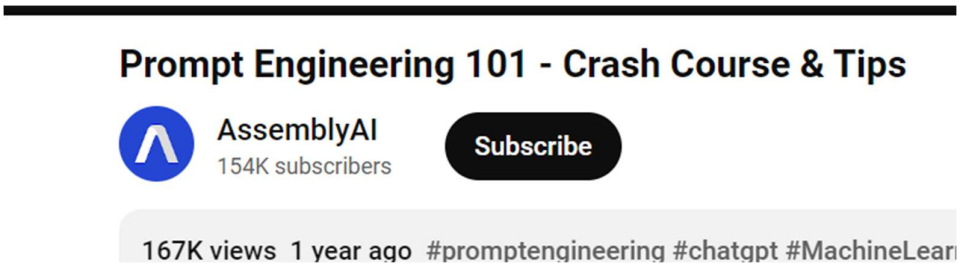
https://www.aiforwork.co/

Google AI Studio – Prompt Gallery

https://aistudio.google.com/app/gallery



Prompt Engineering References



https://www.youtube.com/watch?v=aOm75o2Z5-o

Retrieval Augmented Generation (RAG)

RAG Use Cases

The development of RAG technique is rooted in use cases that were limited by the inherent weaknesses of the LLMs. As of today some commercial applications of RAG are in -

Document Question Answering Systems

By providing access to proprietary enterprise document to an LLM, the responses are limited to what is provided within them. A retriever can search for the most relevant documents and provide the information to the LLM. Check out this blog for an example

Conversational agents

LLMs can be customised to product/service manuals, domain knowledge, guidelines, etc. using RAG. The agent can also route users to more specialised agents depending on their query. SearchUnify has an LLM+RAG powered conversational agent for their users.

Real-time Event Commentary

Imagine an event like a sports or a new event. A retriever can connect to real-time updates/data via APIs and pass this information to the LLM to create a virtual commentator. These can further be augmented with Text To Speech models.IBM leveraged the technology for commentary during the 2023 US Open

Content Generation

The widest use of LLMs has probably been in content generation. Using RAG, the generation can be personalised to readers, incorporate real-time trends and be contextually appropriate. Yarnit is an AI based content marketing platform that uses RAG for multiple tasks.

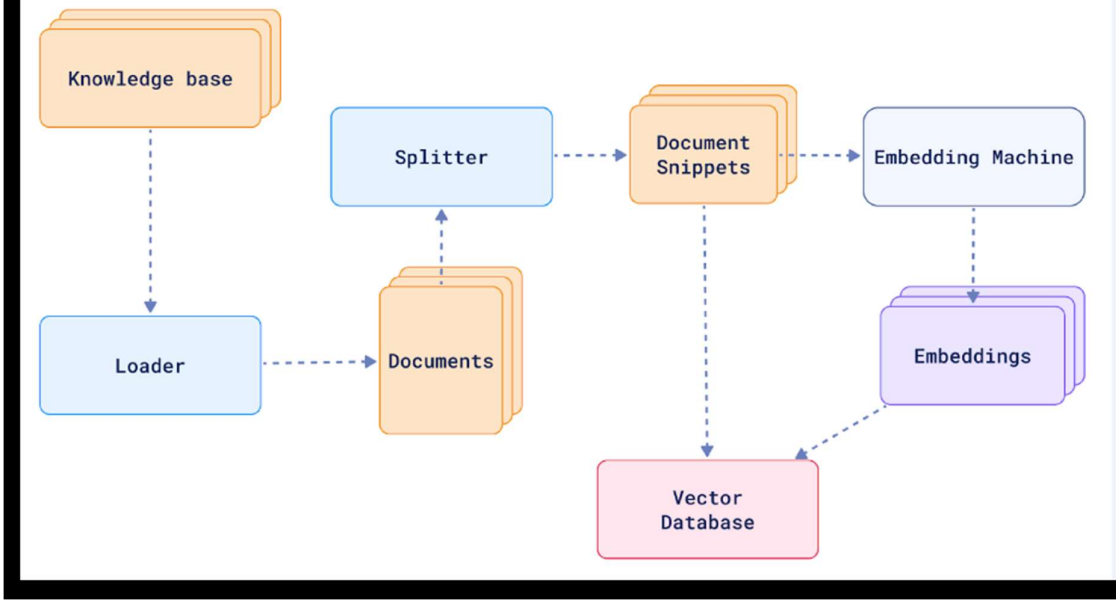
Personalised Recommendation

Recommendation engines have been a game changes in the digital economy. LLMs are capable of powering the next evolution in content recommendations. Check out Aman's blog on the utility of LLMs in recommendation systems.

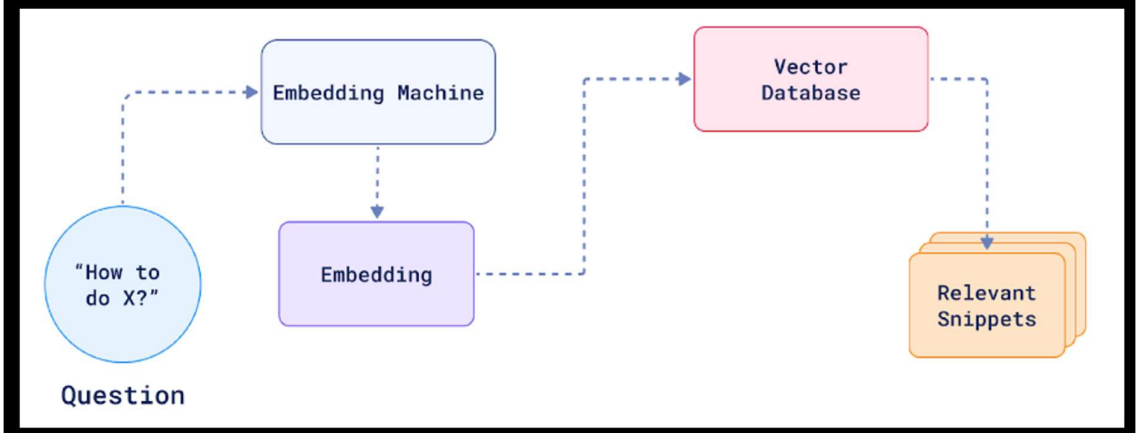
Virtual Assistants

Virtual personal assistants like Siri, Alexa and others are in plans to use LLMs to enhance the experience. Coupled with more context on user behaviour, these assistants can become highly personalised.

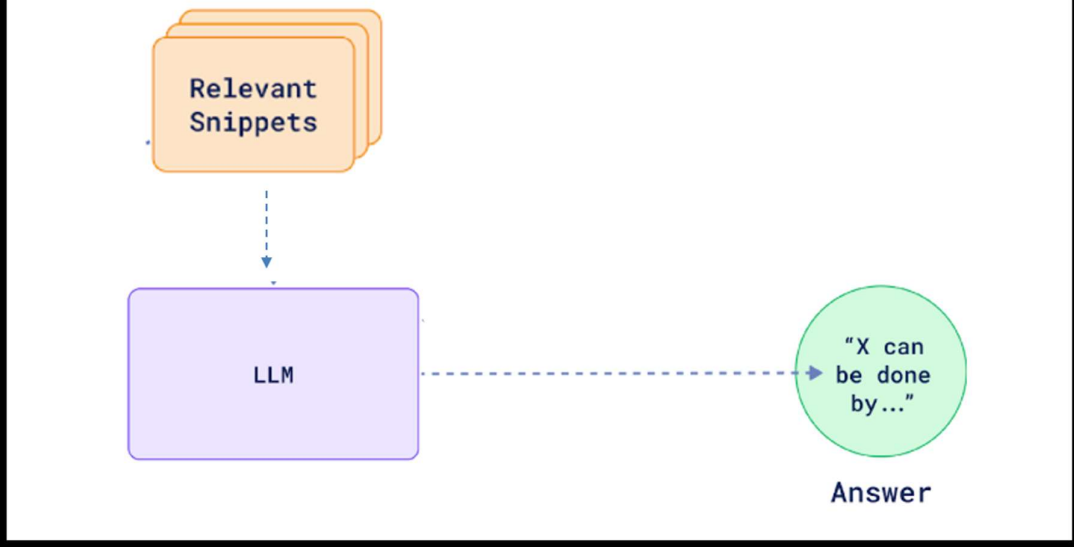
RAG step 1 – Indexing (of new knowledge)



RAG step 2: Retrieval



RAG step 3: Generation



Popular Embedding Models

word2vec Google's Word2Vec is one of the most popular pre-trained word embeddings. The official paper - https://arxiv.org/pdf/1301.3781.pdf

GLOVE The 'Global Vectors' model is so termed because it captures statistics directly at a global level. The official paper - https://nlp.stanford.edu/pubs/glove.pdf

fastText Facebook's AI research, fastText builds embeddings composed of characters instead of words. The official paper - https://arxiv.org/pdf/1607.04606.pdf

Elmo Embeddings from Language Models, are learnt from the internal state of a bidirectional LSTM. The official paper - https://arxiv.org/pdf/1802.05365.pdf

BERT Bidirectional Encoder Representations from Transformers is a transformer bases approach. The official paper - https://arxiv.org/pdf/1810.04805.pdf

Souce: LangChain: https://x.com/LangChainAI/status/1869812624998969836

RAG Optimization knobs

Chunking
Semantic chunking
Agentic chunking

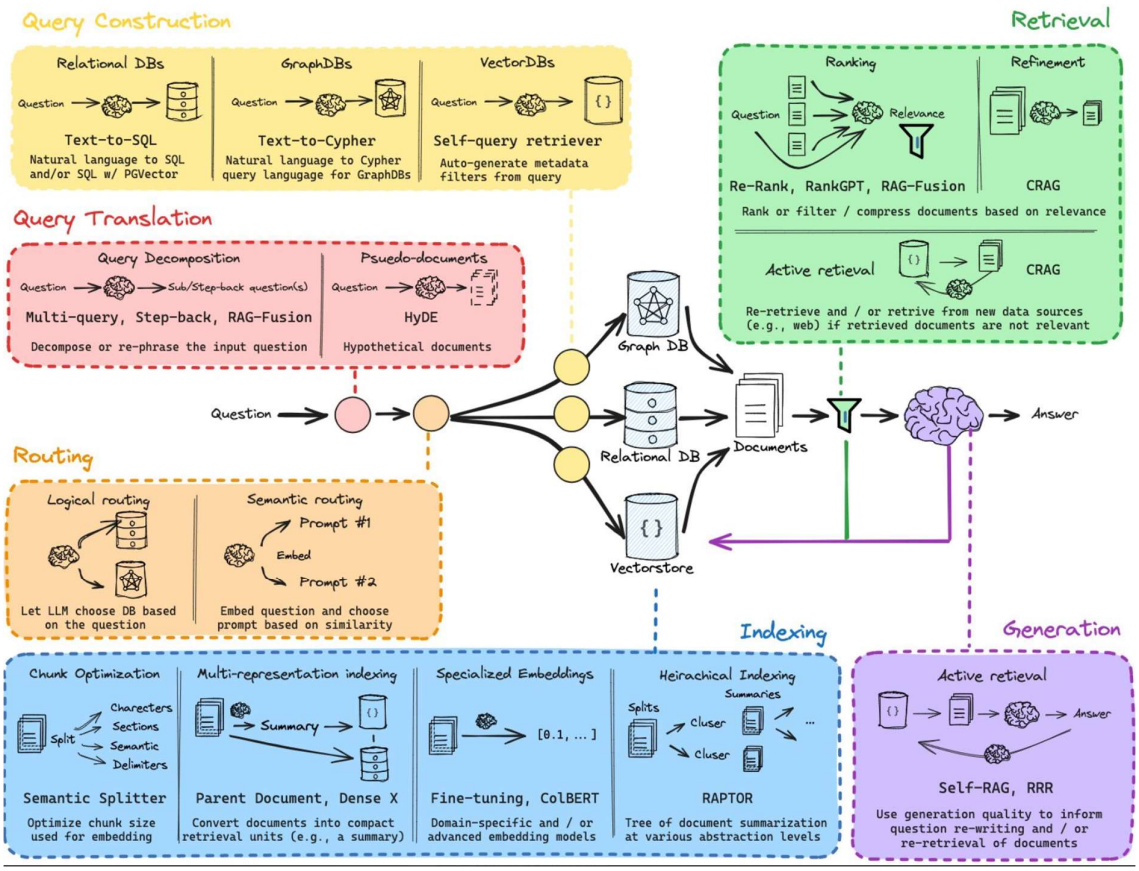
Indexing
vector compression
Metadata
Hyperparameter tuning
Choose right ANN algo

Query optimization
Prompt engineering
Query transformation, splitting
Query expansion, compression
Query routing, augmentation
Lost in the middle problem

Retrieval knobs
Embedding models
hybrid search weighting
re-ranker models
Multi-index search configurations
Metadata filtering.
Time-weighted retrieving
Summary based retrieval
Ensemble retriever

Generation knobs
Choice of LLM
LLM configuration & Tuning
(PEFT, FFT)

Advanced RAG Implementation:



Source: https://x.com/LangChainAI/status/1754915914796216654

Quantization to improve RAG performance:

https://qdrant.tech/articles/what-is-vector-quantization/

RAG References

Top 9 Vector databases you should know:

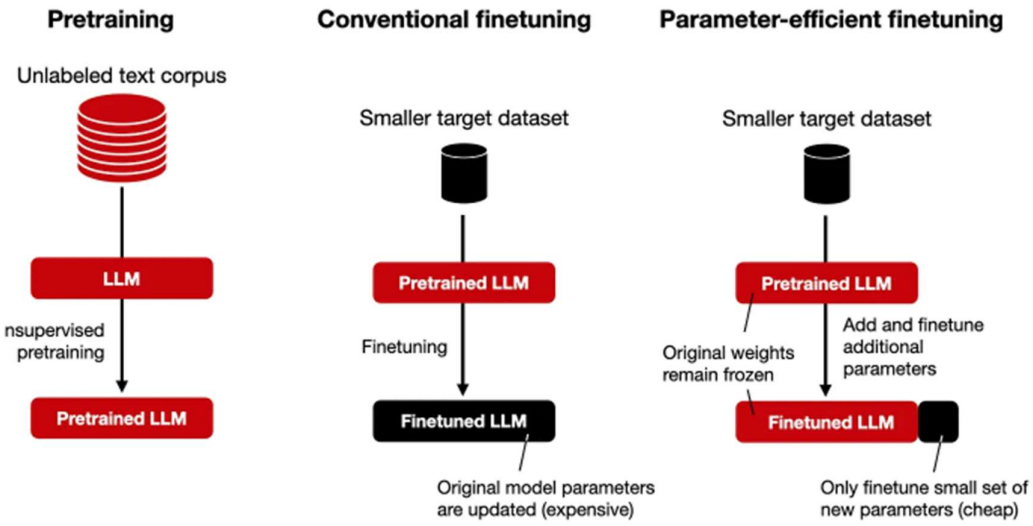
https://machinelearningknowledge.ai/top-vector-databases-you-should-know/#google_vignette

How does similarity search work?

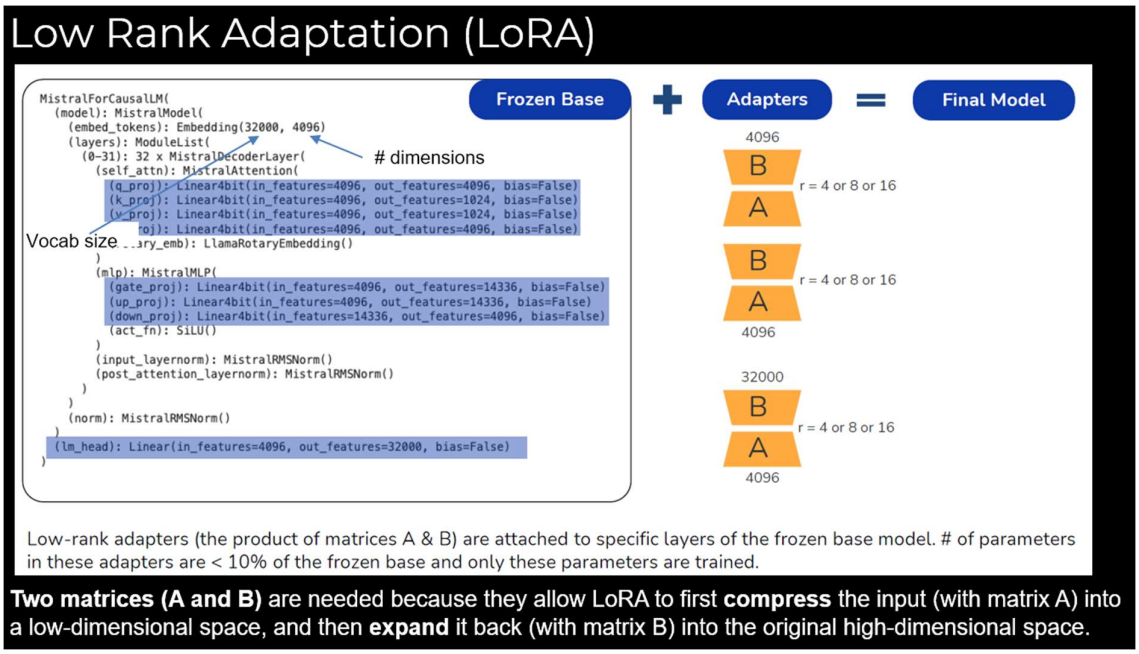
https://towardsdatascience.com/similarity-search-part-4-hierarchical-navigable-small-world-hnsw-2aad4fe87d37/

https://www.youtube.com/watch?v=QvKMwLjdK-s

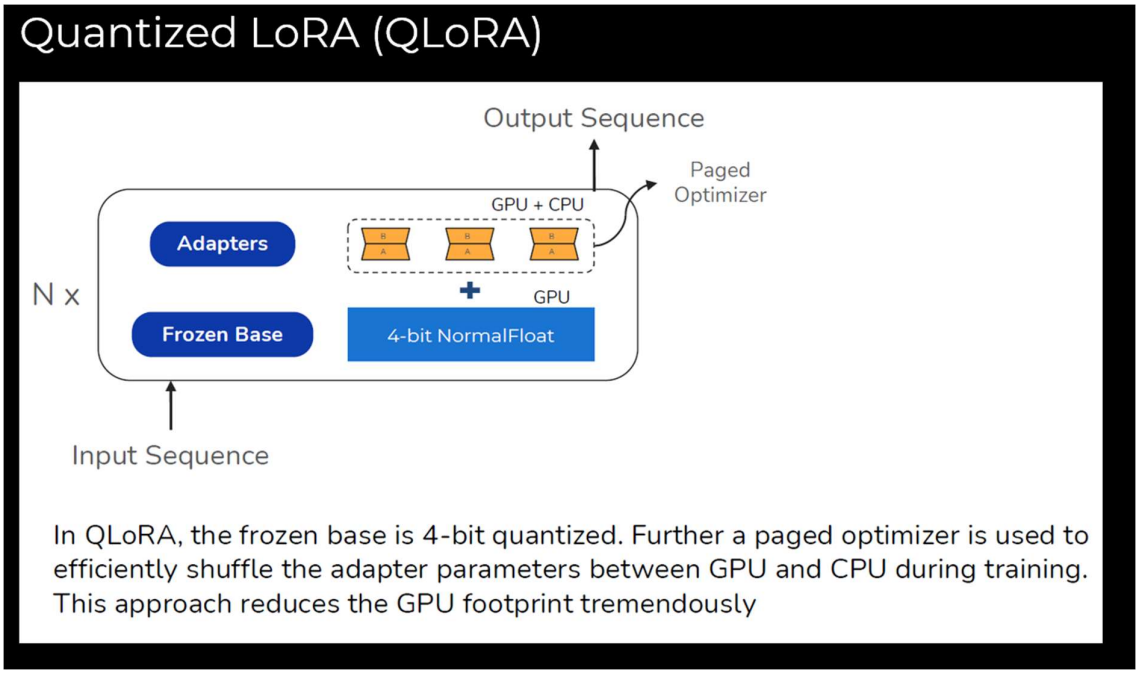
LLM Fine Tuning



Low Rank Adaptation (LoRA)



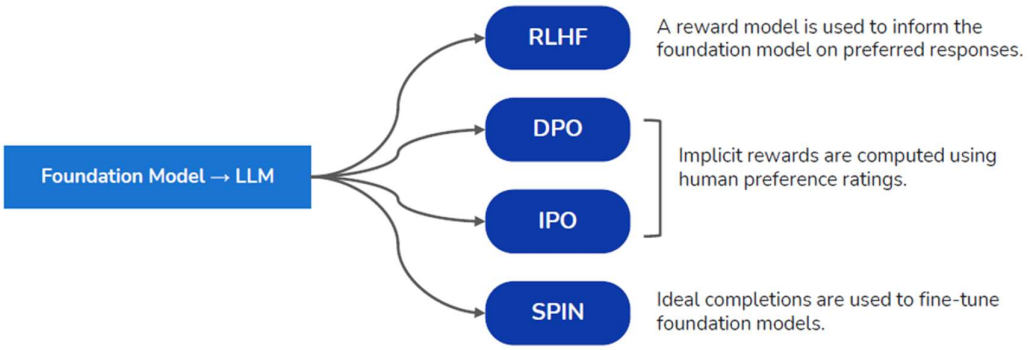
Quantized Low Rank Adaptation (QLoRA)



Alignment Fine Tuning (RLHF etc.)

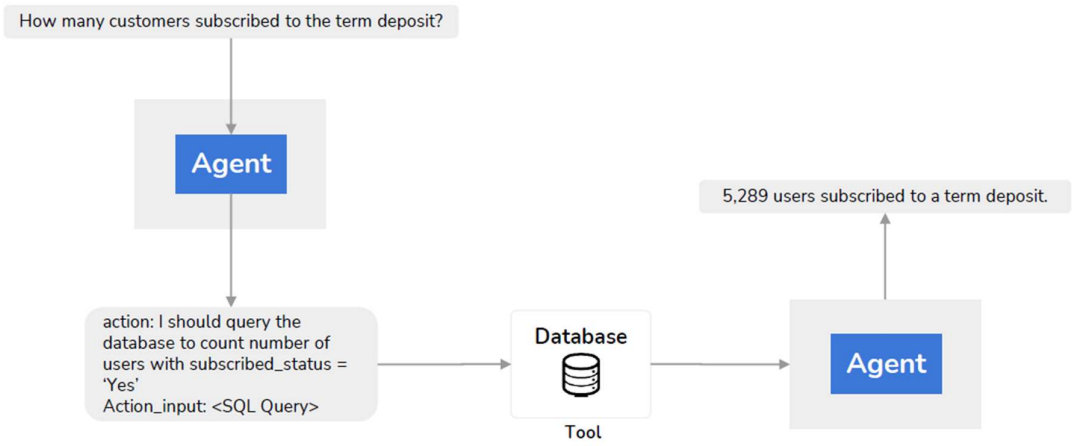
Alignment fine tuning techniques

Foundation models are aligned with human preferences by rewarding high human ratings of model responses.



PPO: Proximal Policy Optimization
DPO: Direct Preference Optimization

Agents



NVIDIA CEO on Agents Being the Future of AI

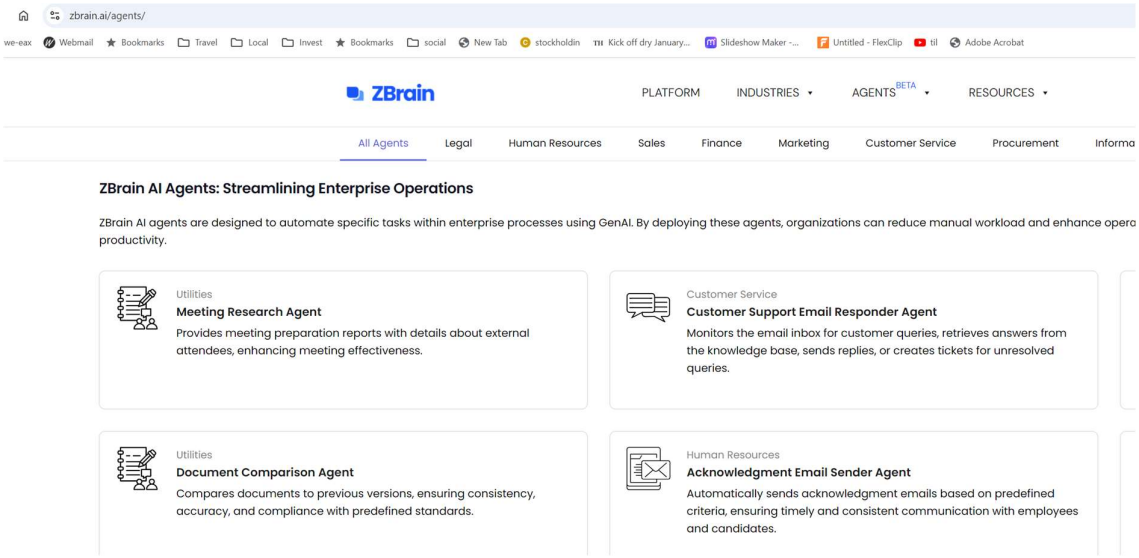
Matthew Berman
369K subscribers

Join

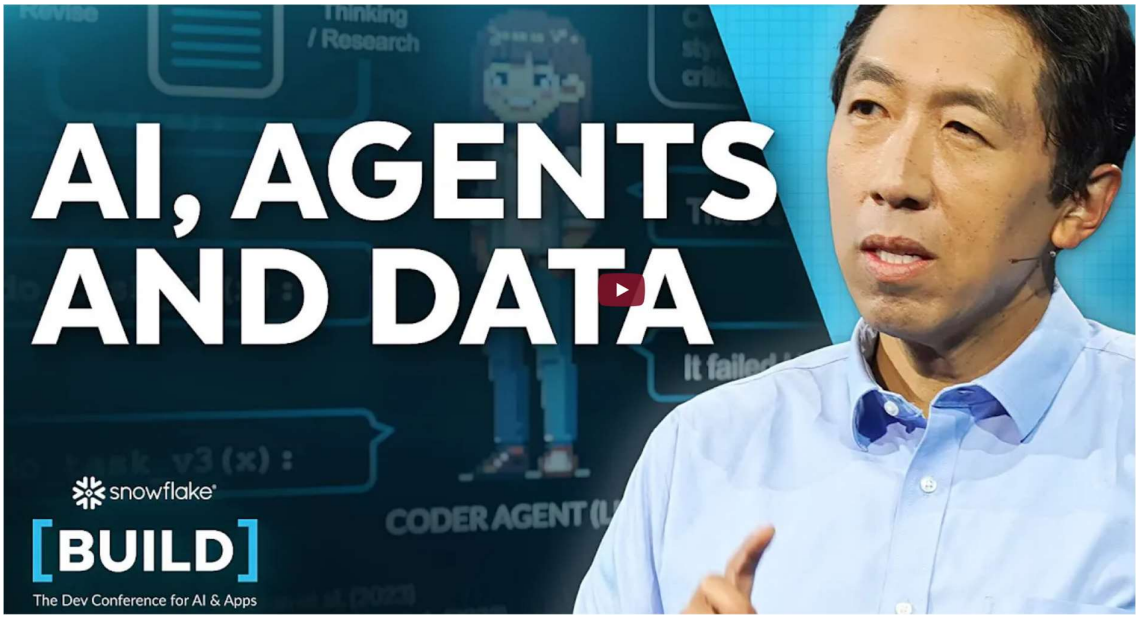
Subscribed

93K views 3 months ago ✪ Members first
Join My Newsletter for Regular AI Updates 👇
https://forwardfuture.ai
...more

https://www.youtube.com/watch?v=SMnBnGQqlR4

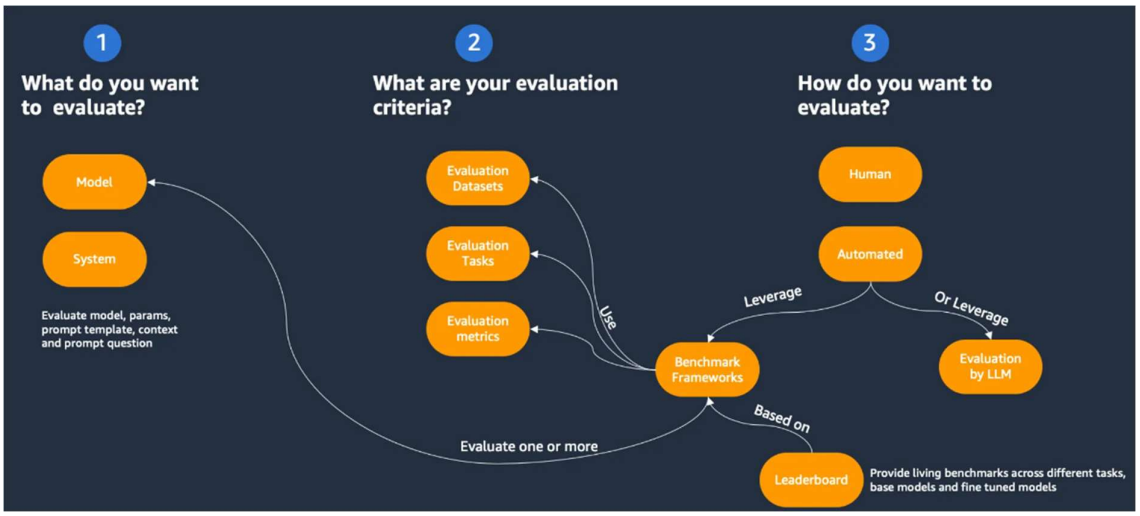


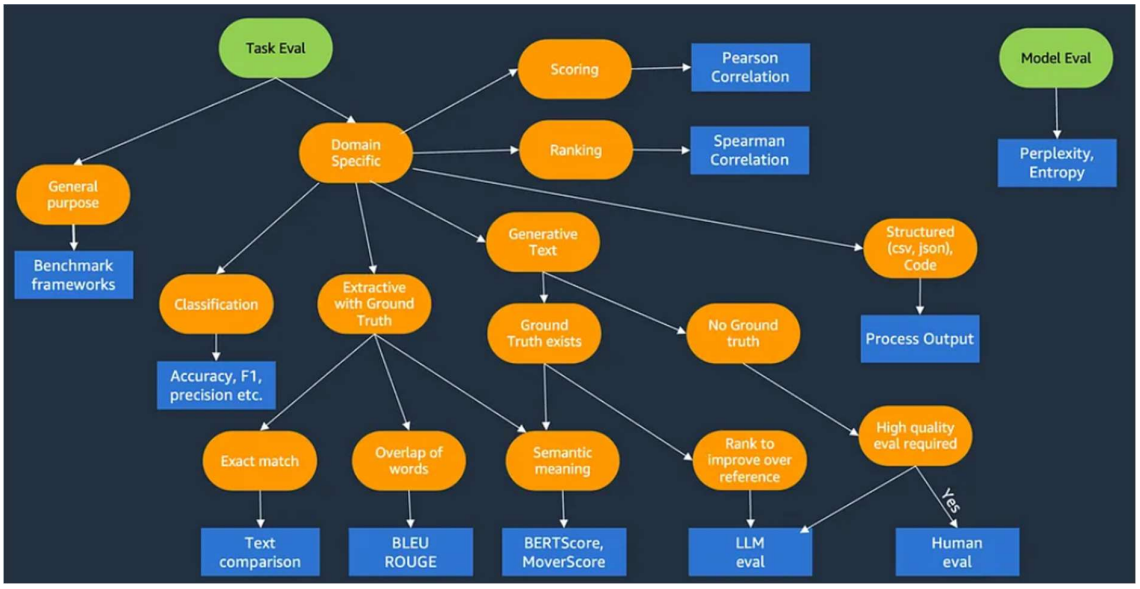
Andrew Ng Explores The Rise Of AI Agents And Agentic Reasoning | BUILD 2024 Keynote, Snowflakes Inc

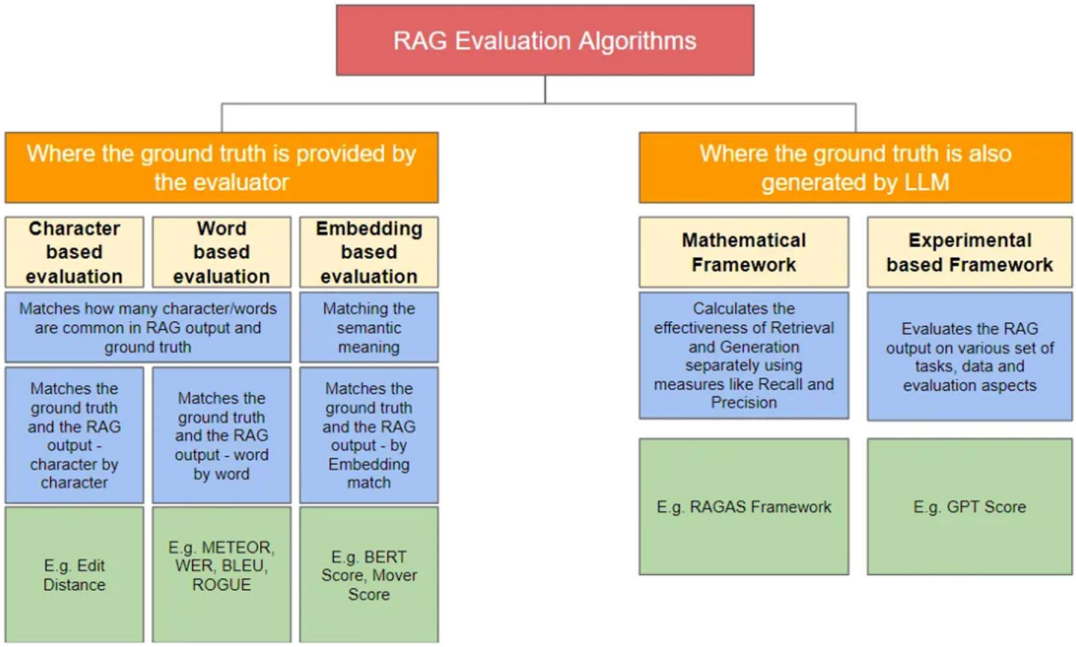


https://www.youtube.com/watch?v=KrRD7r7y7NY

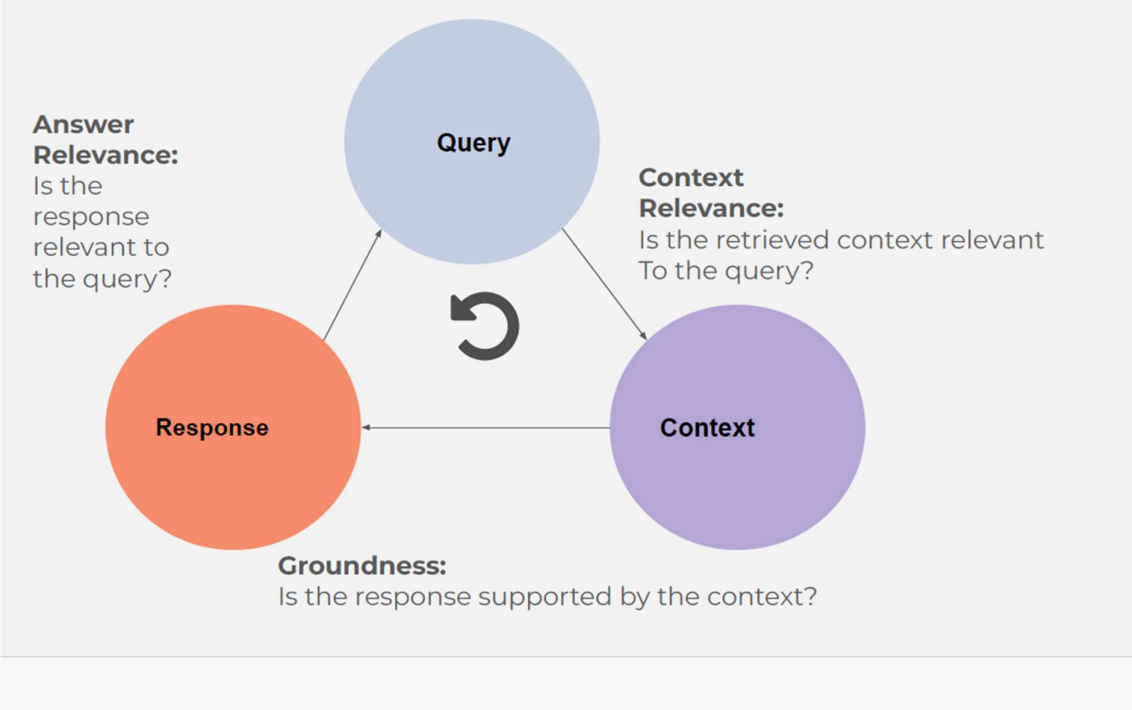
Evaluating LLMs

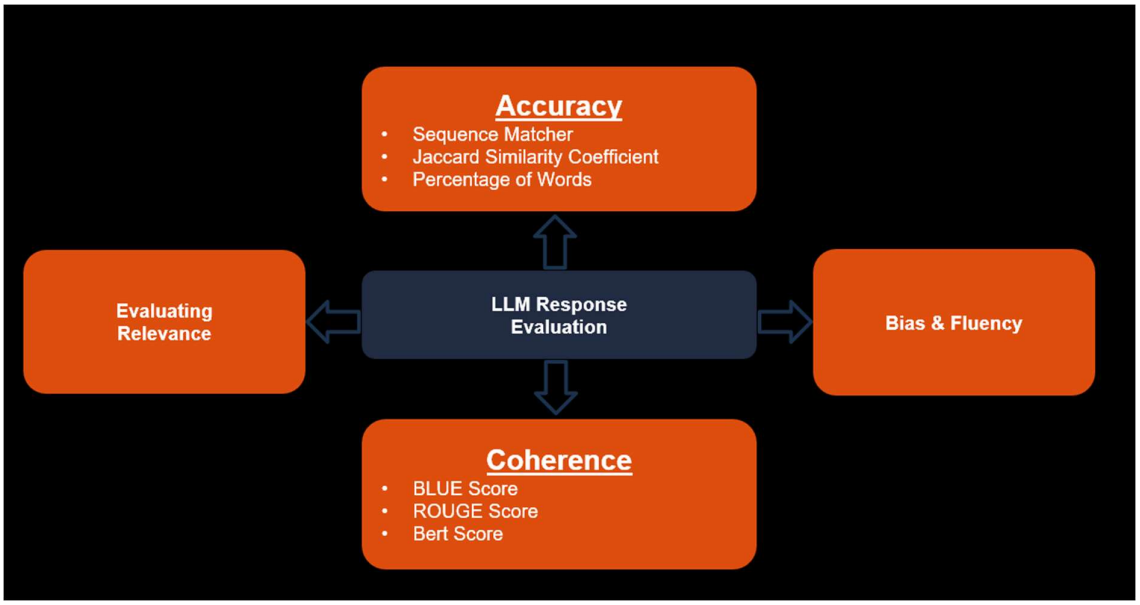


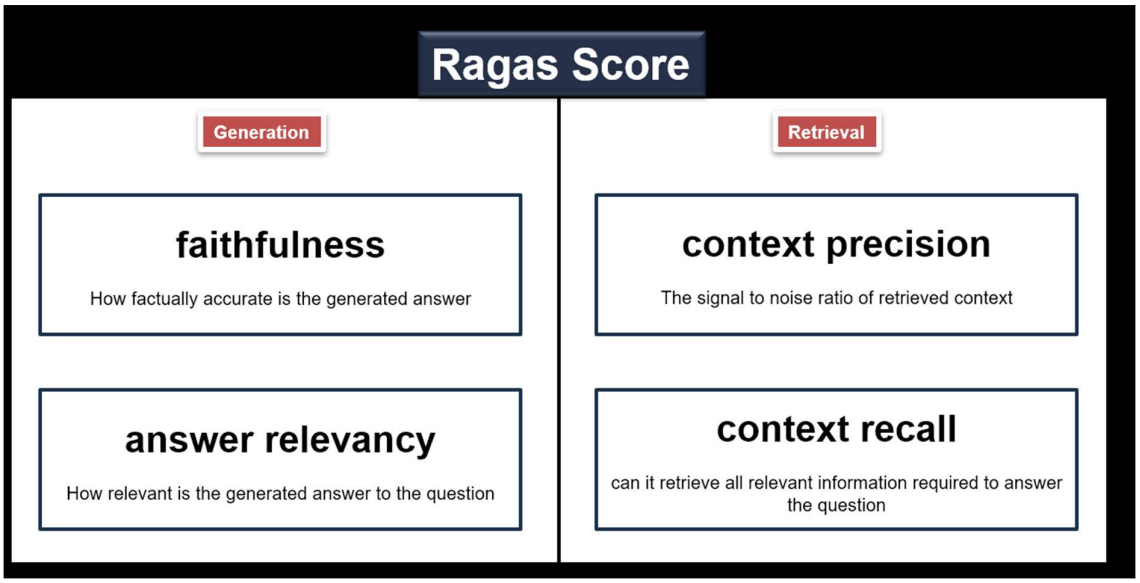




The RAG Triad

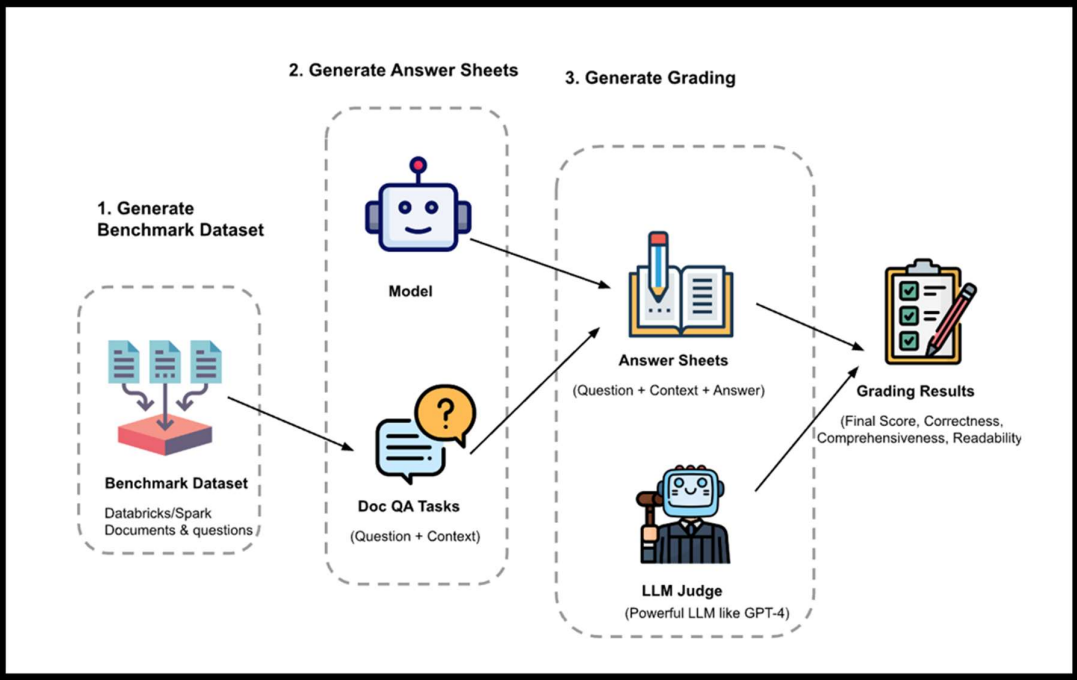






LLM as a Judge

LLM as a judge



Role of Human Reviewers

Role of Human reviewers in assessing response quality

- Human reviewers play a crucial role in evaluating how well the responses from LLM align with the intended goals and expectations
- It makes sure that LLM is not only informative but also produces human-like text
- Researchers assess LLM for empathetic responses and effective addressing of user concerns

Platforms And Tools for Gen AI App Development

Langchain

Langflow

Llamaindex

LmStudio

Replit

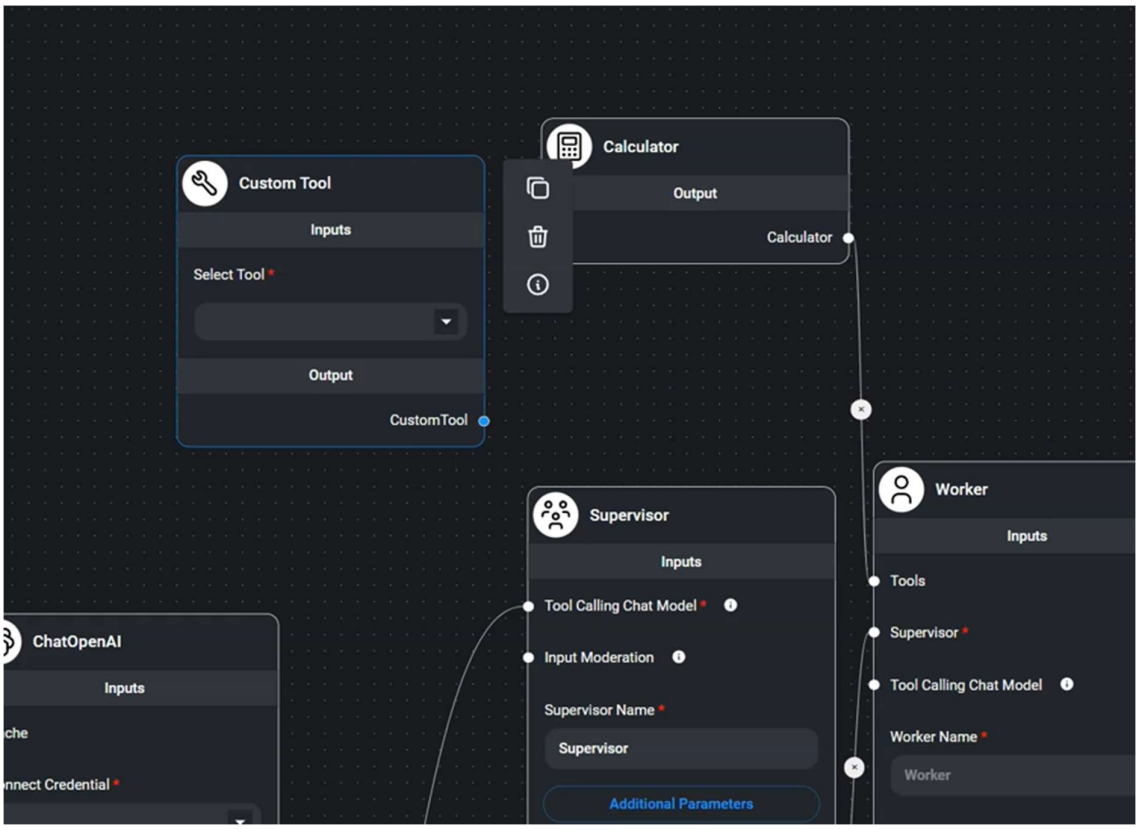
Anything LLM

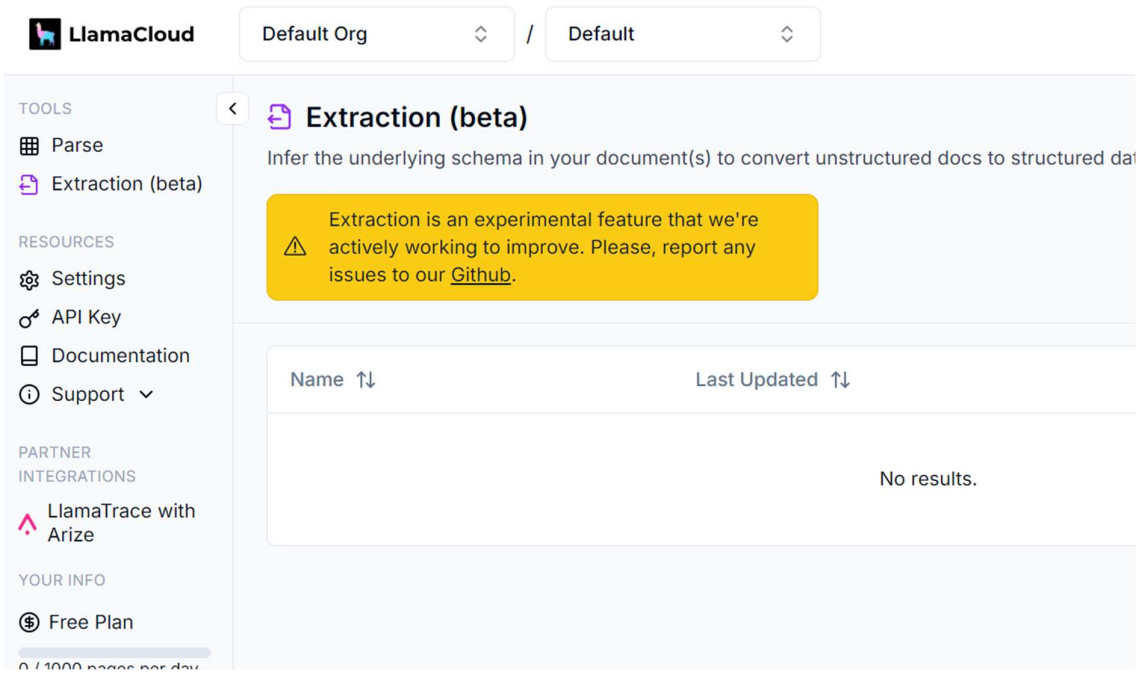
Make.com

Firecrawl

Autotrain

Flowise





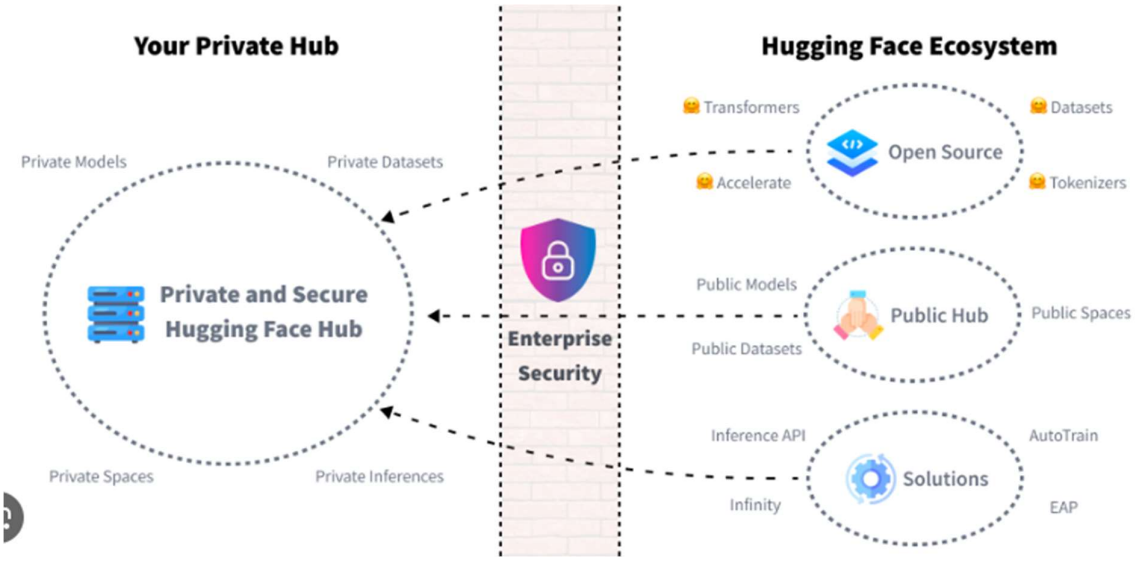
Huggingface

https://huggingface.co

HF cookbooks:

https://huggingface.co/learn/cookbook/mlflow_ray_serve

HF Hub:



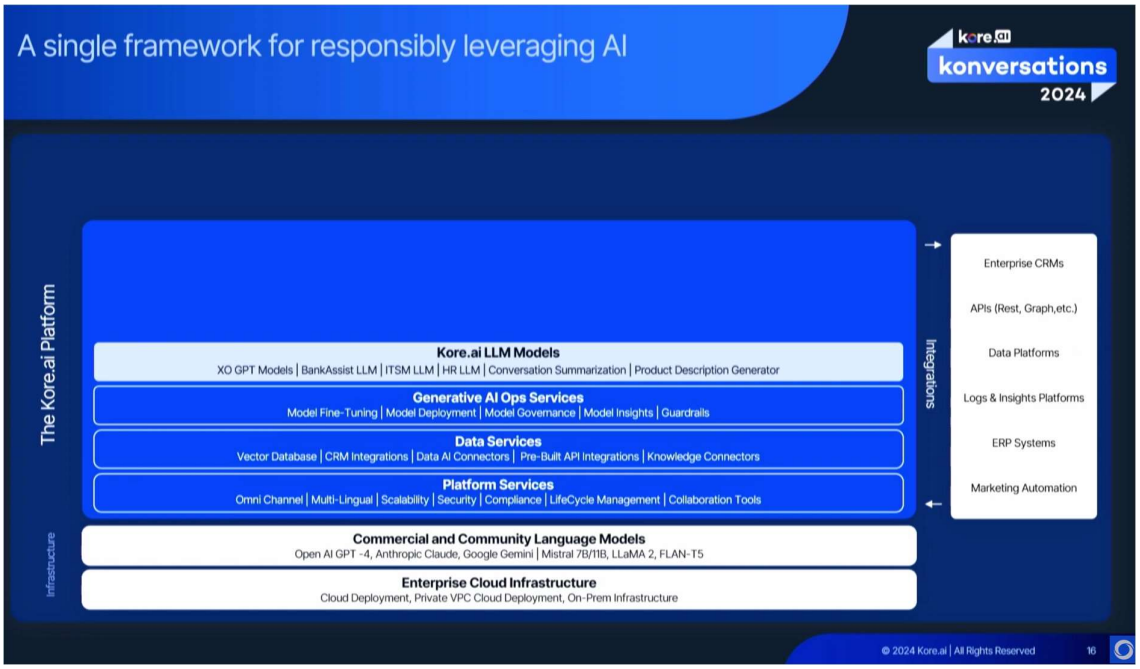
LLM Finetuning:

https://github.com/ashishpatel26/LLM-Finetuning

Kore.ai

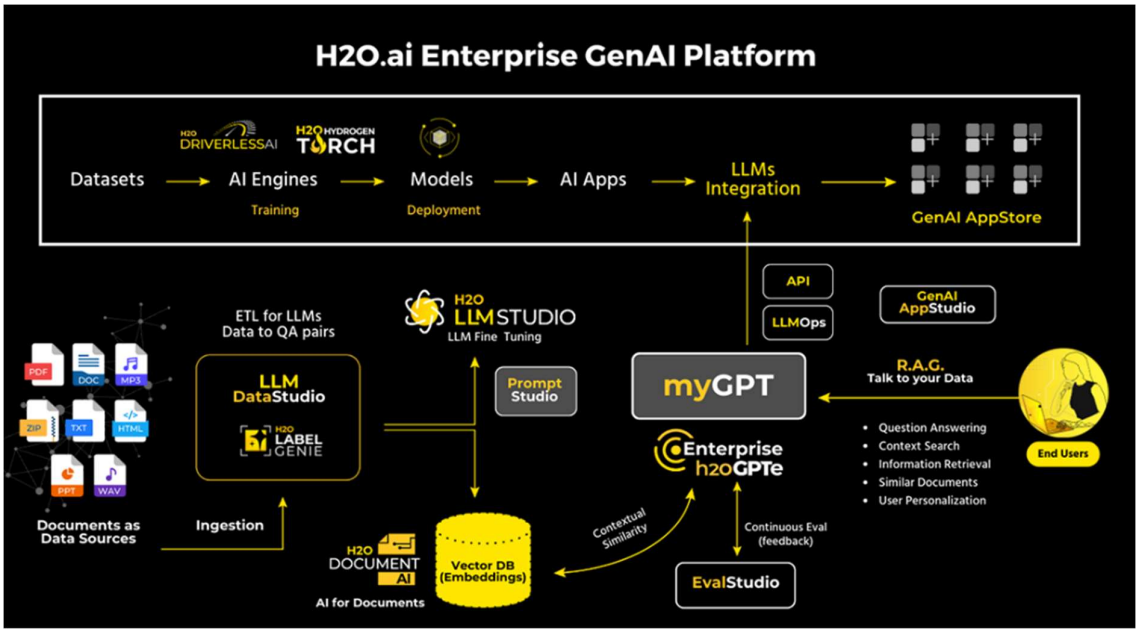
https://www.youtube.com/watch?v=p06tMqyohBI

https://kore.ai

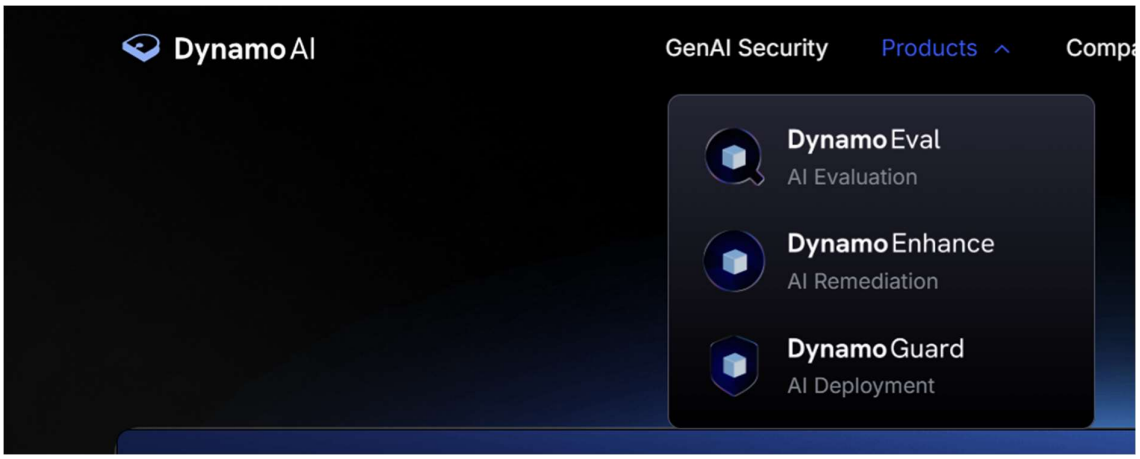


H2O.AI

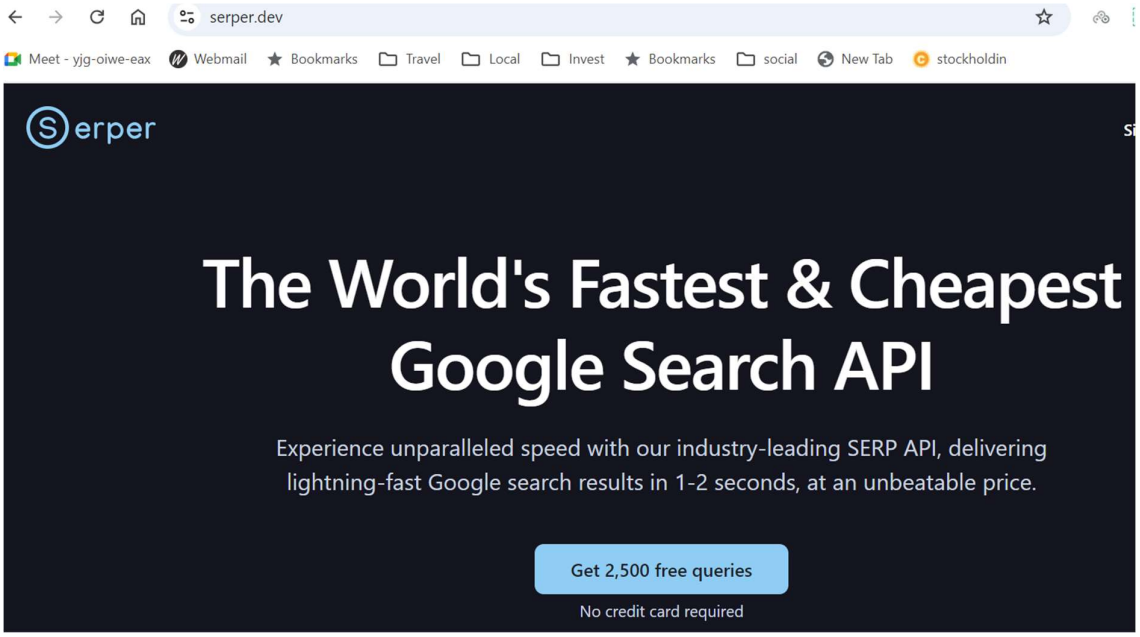
https://h2o.ai/



https://dynamo.ai/



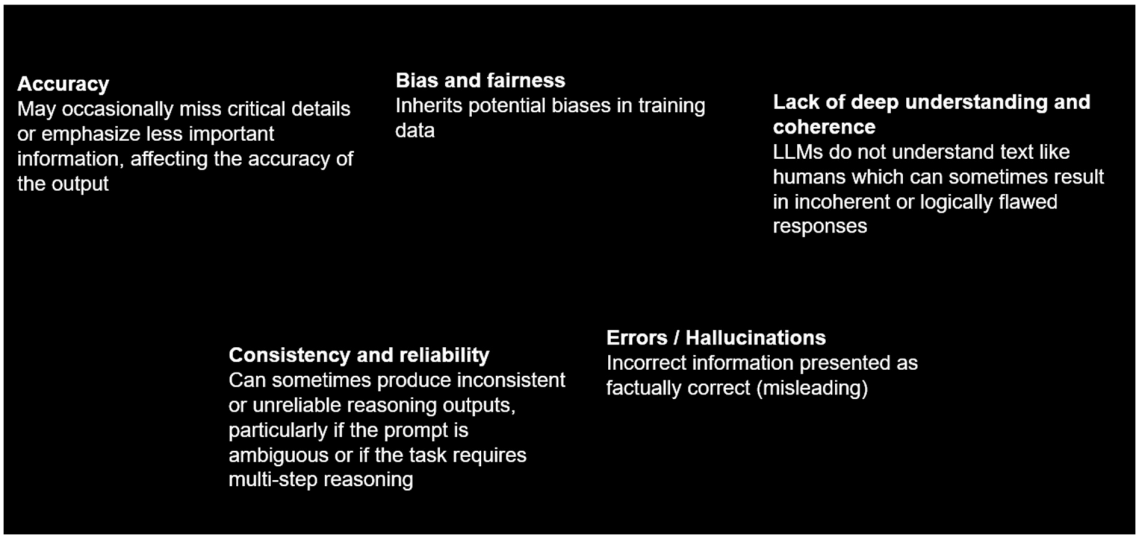
Serper.dev for google search API



Implementation Challenges

1) Lack of good training data
   Mitigation: Synthetic data – use LLMs to generate data

2) Quality of output:



Problems

1. Controllability of AI output
2. Accuracy and reliability
3. IP protection
4. Latency
5. Governance challenges
6. Costs
7. Privacy concerns

Mitigation: Prompt Engineering, RAG, Fine Tuning

Hallucination

Hallucinations occur when AI model generates incorrect or misleading information but presents it as if it were a fact.

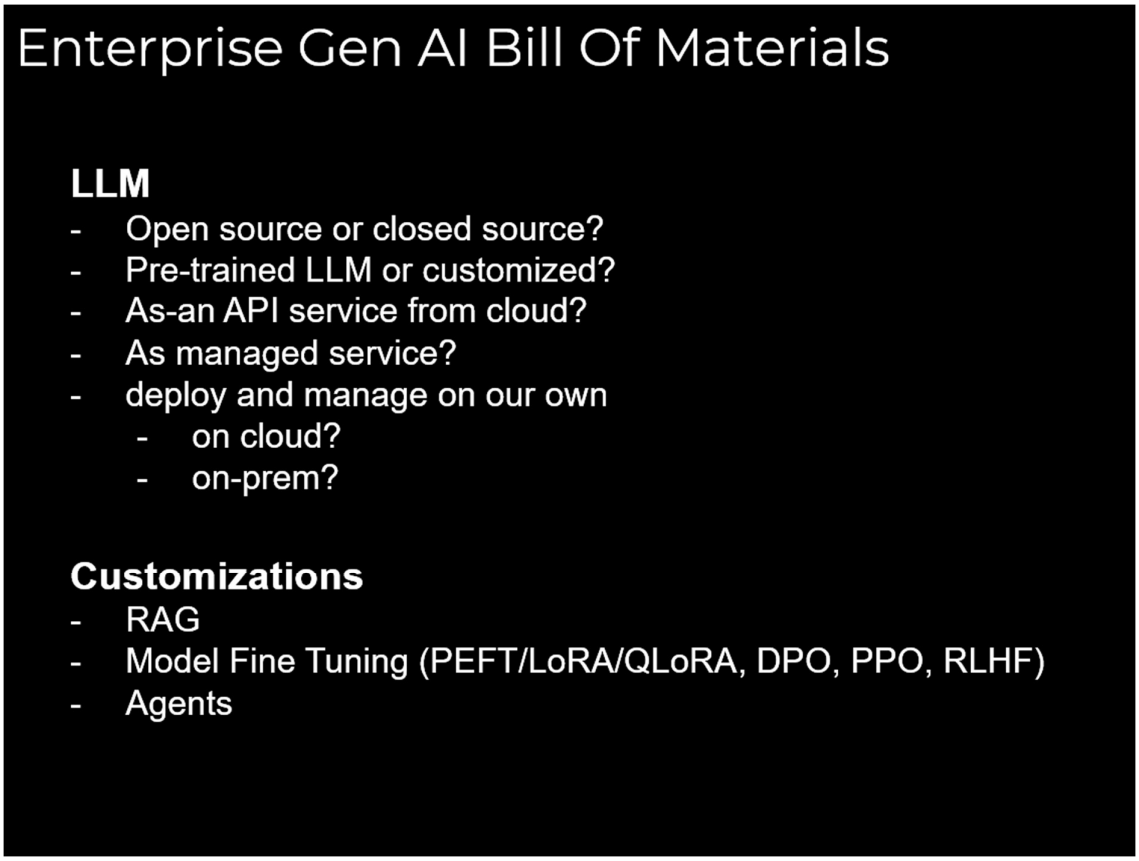
Example: LLM designed to generate summaries of news articles may produce a summary that includes details not present in the original article, or even fabricates information entirely.

Reasons

- Flawed training data – Inaccurate, Incomplete.
- Overfitting to training data
- Extrapolation beyond training.
- Insufficient context or context switching
- Ambiguous prompt

AI, Investment Decisions, and Inequality by Alex G. Kim, David S. Kim, Maximilian Muhn, Valeri V. Nikolaev , Eric C. So :: SSRN

Costing



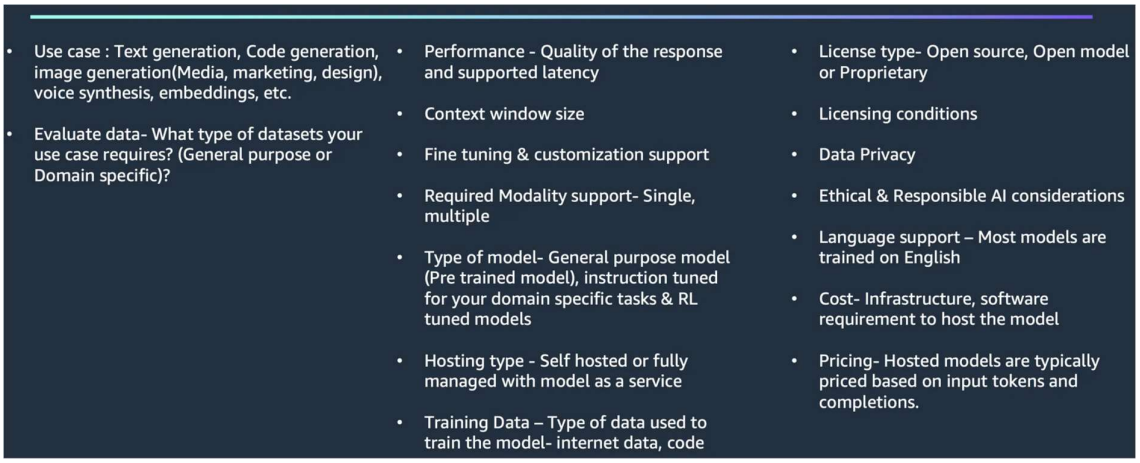
Recent costing (check latest from vendors)

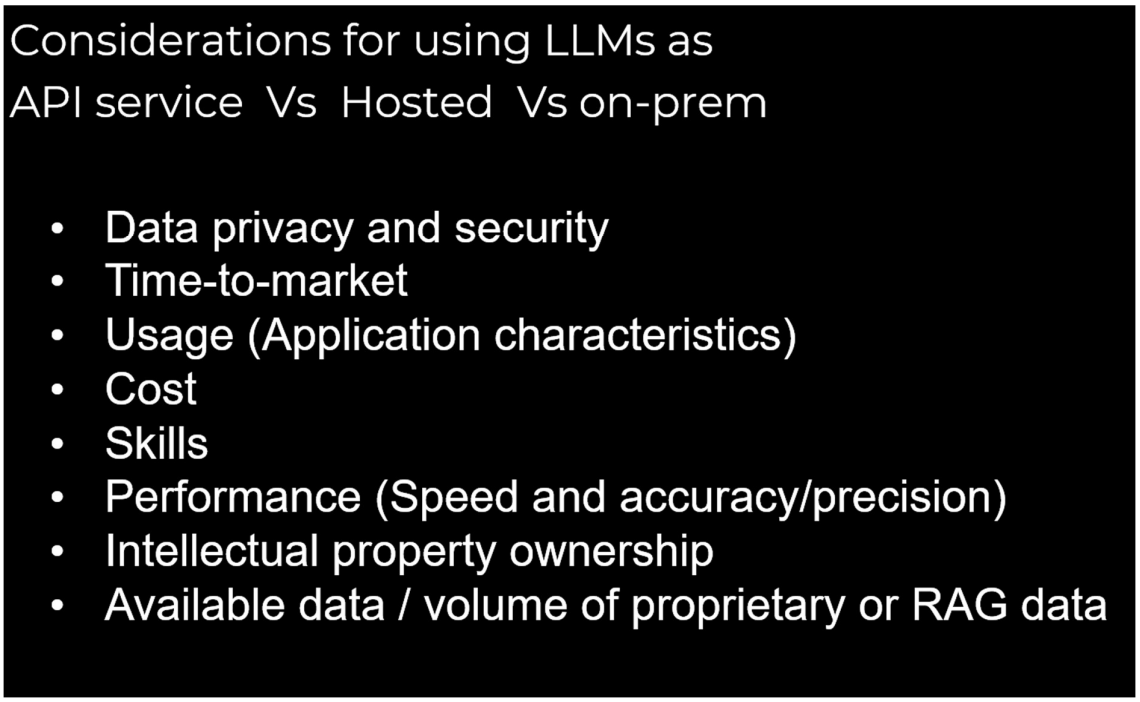
Model	Quality	Resolution	Price
DALL·E 3	Standard	1024×1024	$0.040 / image
	Standard	1024×1792, 1792×1024	$0.080 / image
DALL·E 3	HD	1024×1024	$0.080 / image
	HD	1024×1792, 1792×1024	$0.120 / image
DALL·E 2		1024×1024	$0.020 / image
		512×512	$0.018 / image
		256×256	$0.016 / image

GPT-4o mini is our most cost-efficient small model that's smarter and cl
than GPT-3.5 Turbo, and has vision capabilities. The model has 128K cor
an October 2023 knowledge cutoff.

Learn about GPT-4o mini ↗

Model	Pricing	Pricing with Batch API*
gpt-4o-mini	0,150 $ / 1M input tokens	0,075 $ / 1M input tokens
	0,600 $ / 1M output tokens	0,300 $ / 1M output tokens
gpt-4o-mini-2024-07-18	0,150 $ / 1M input tokens	0,075 $ / 1M input tokens
	0,600 $ / 1M output tokens	0,300 $ / 1M output tokens

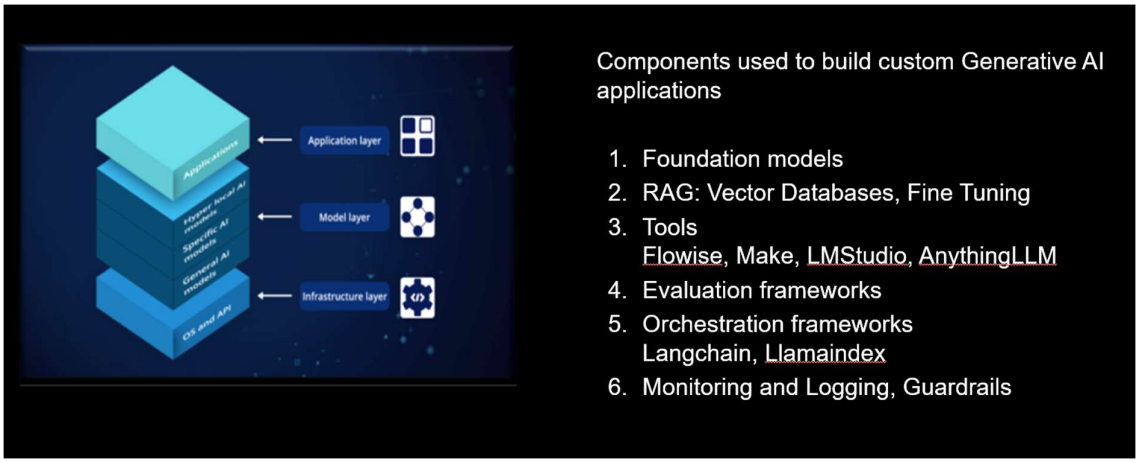


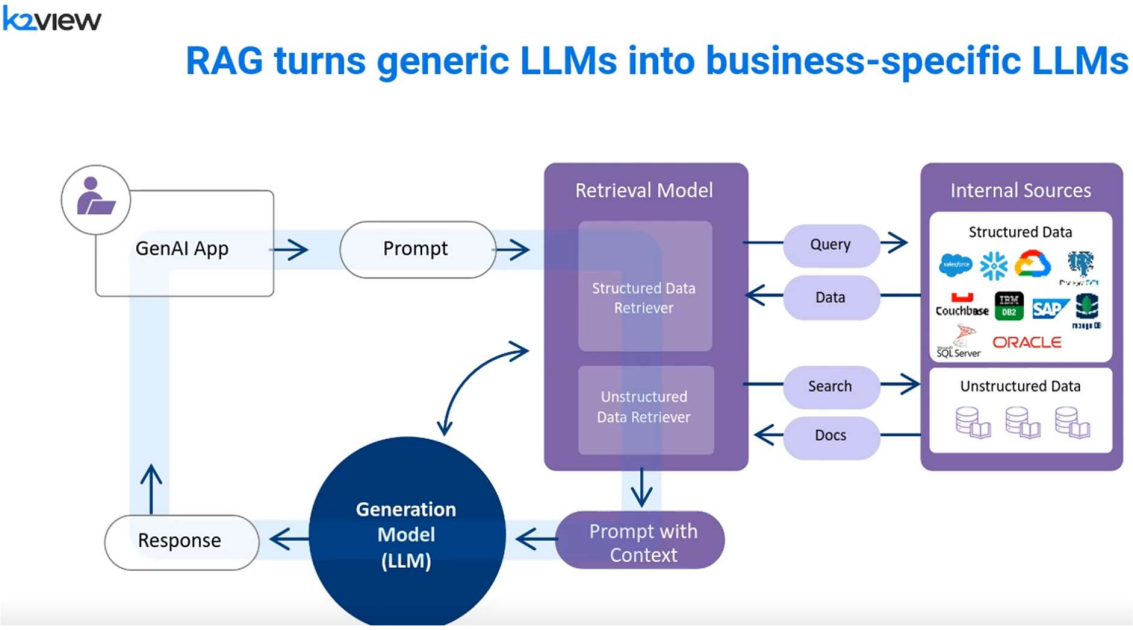


https://medium.com/@gopikwork/comprehensive-guide-for-model-selection-and-evaluation-fcd7fe299a50

https://medium.com/emalpha/the-economics-of-large-language-models-2671985b621c

Enterprise Considerations

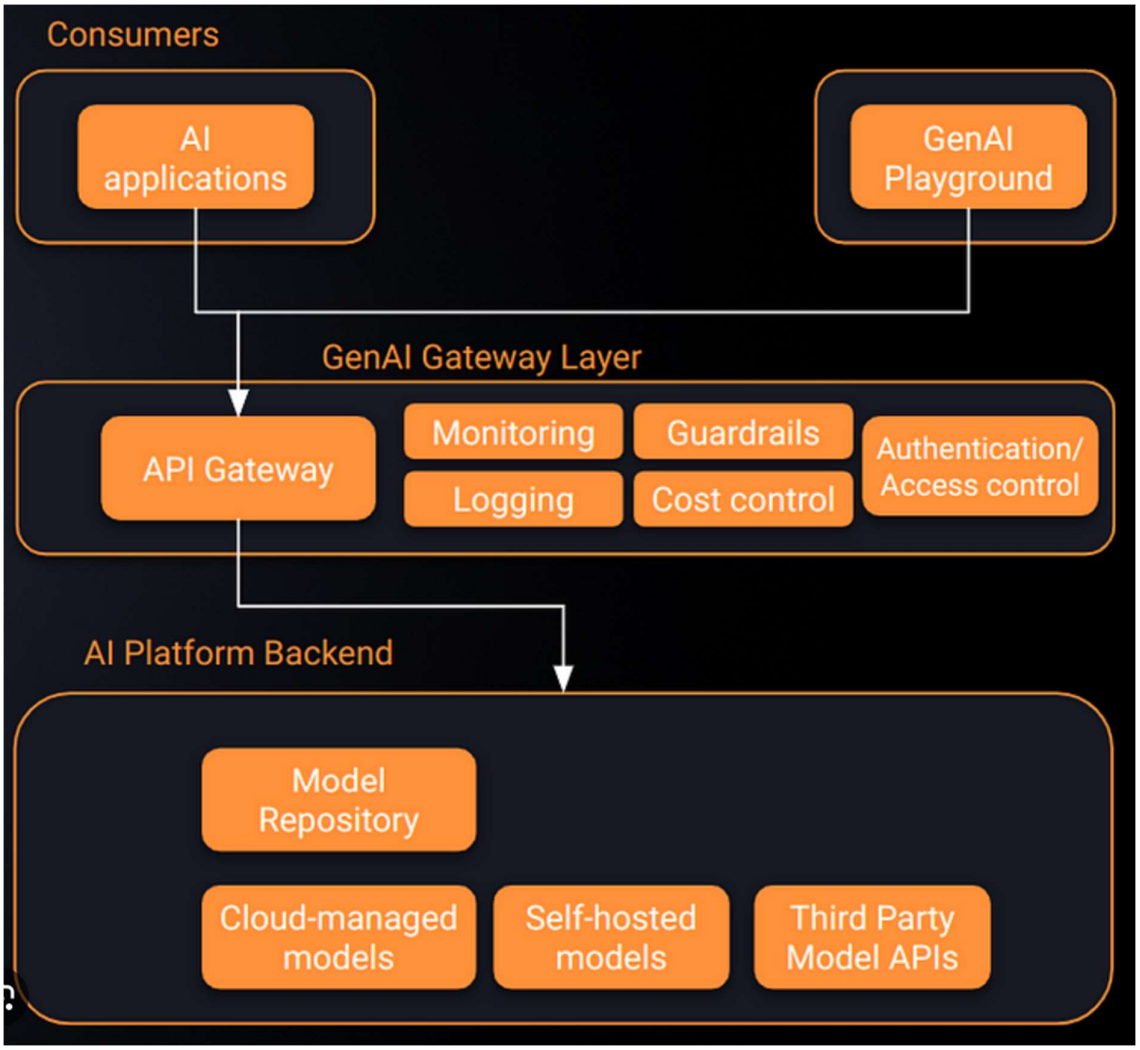




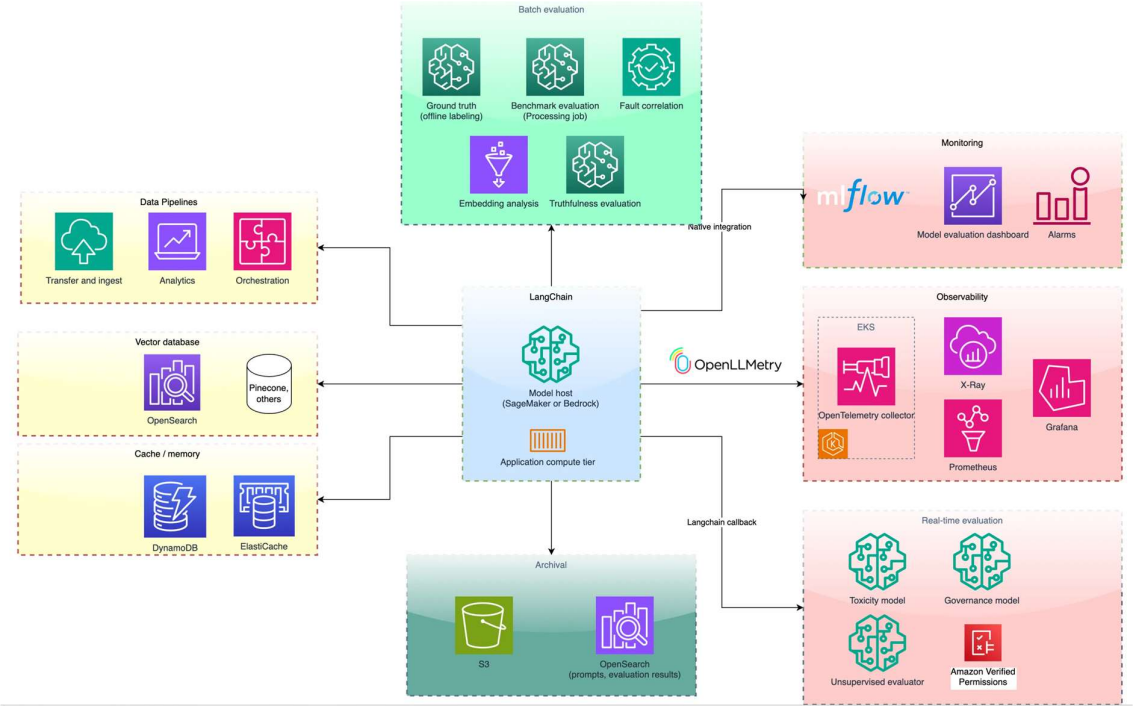
Productionalizing LLM Applications

What We've Learned From A Year of Building with LLMs – Applied LLMs

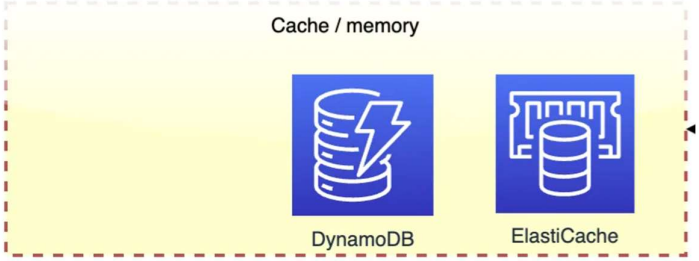
https://community.aws/content/2i7BzRVM4ppZSGaZKoEHdpNwick/observability-monitoring-and-layered-guardrails-for-genai

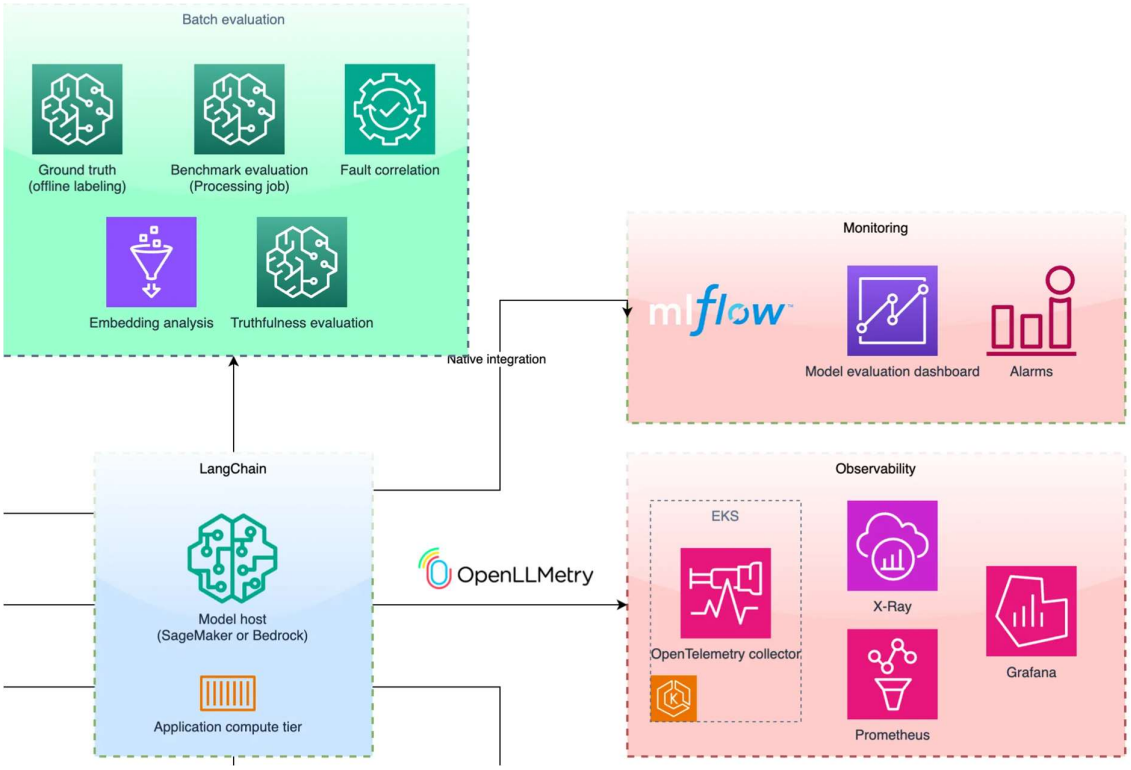


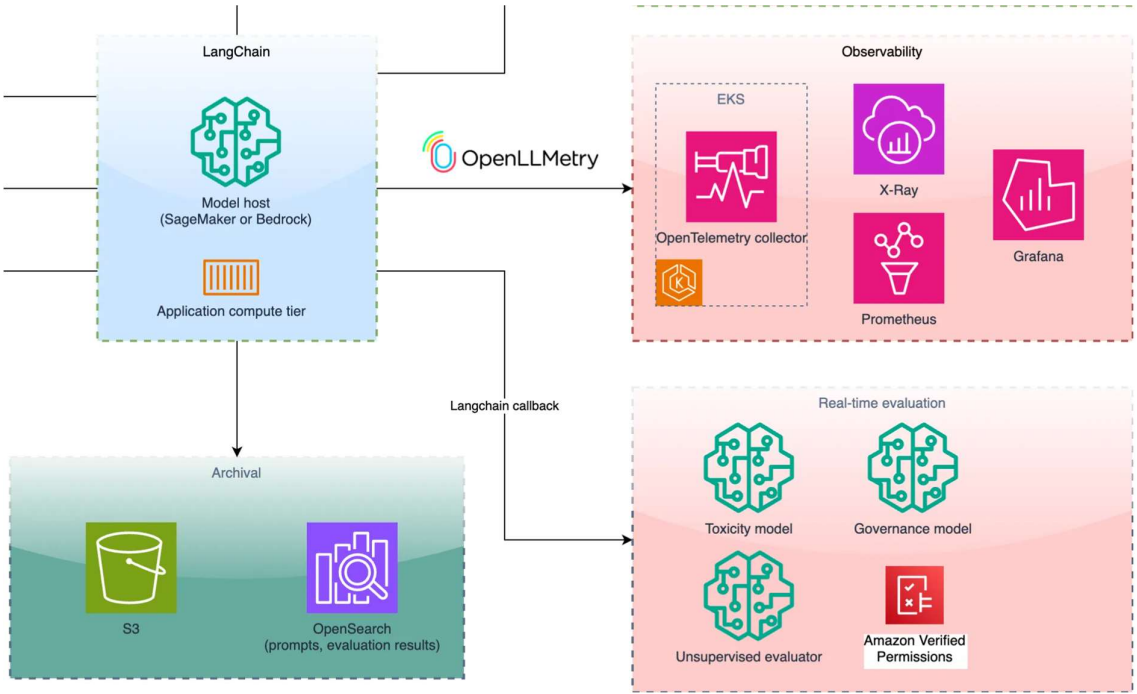
https://www.ml6.eu/blogpost/why-you-need-a-genai-gateway



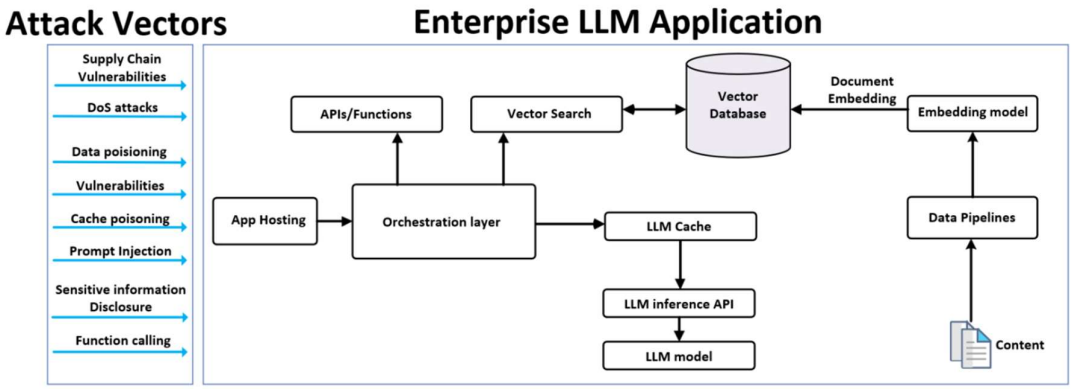
https://community.aws/content/2i7BzRVM4ppZSGaZKoEHdpNwick/observability-monitoring-and-layered-guardrails-for-genai




Batch evaluation
Ground truth
(offline labeling)
Benchmark evaluation
(Processing job)
Fault correlation
Embedding analysis
Truthfulness evaluation
Monitoring
mlflow
Model evaluation dashboard
Alarms
Native integration
LangChain
Model host
(SageMaker or Bedrock)
Application compute tier
OpenLLMetry
Observability
EKS
OpenTelemetry collector
X-Ray
Grafana
Prometheus

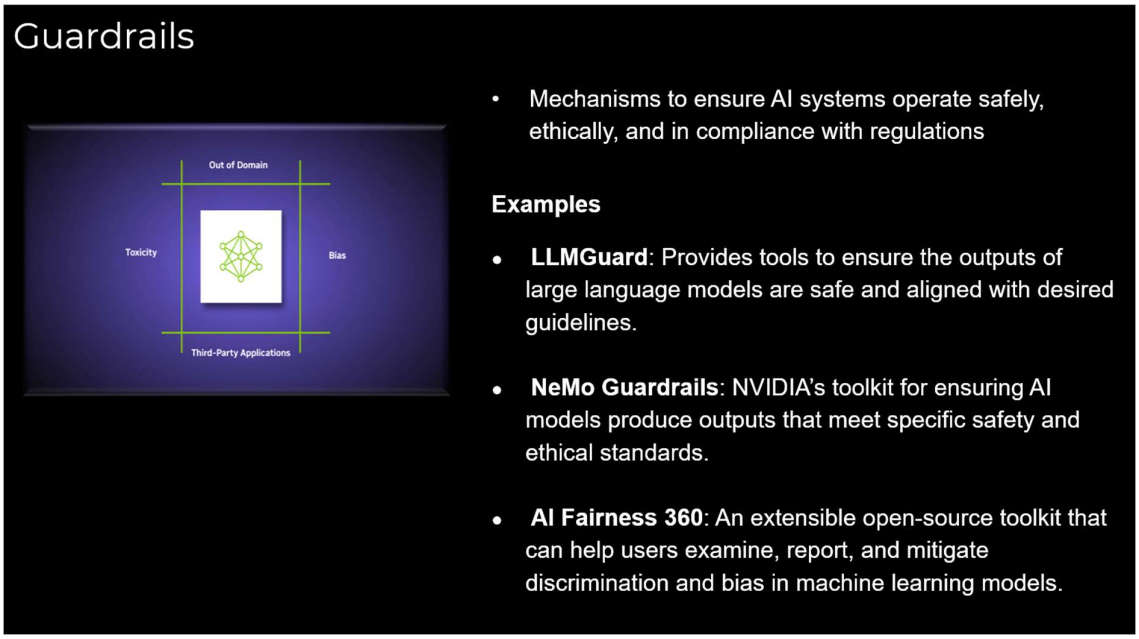

LangChain
Model host
(SageMaker or Bedrock)
Application compute tier
OpenLLMetry
Observability
EKS
OpenTelemetry collector
X-Ray
Grafana
Prometheus
Langchain callback
Archival
S3
OpenSearch
(prompts, evaluation results)
Real-time evaluation
Toxicity model
Governance model
Unsupervised evaluator
Amazon Verified
Permissions

Securing LLM Applications



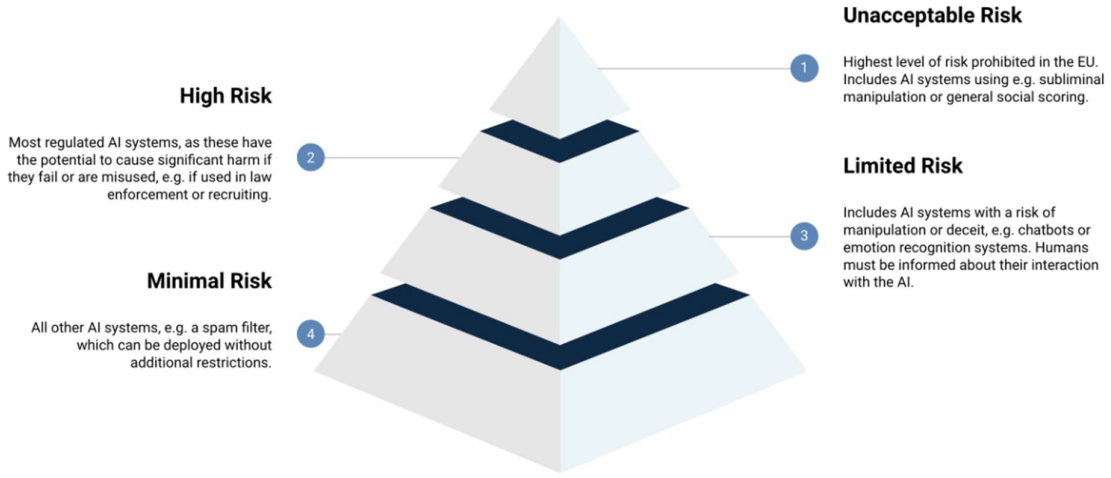
https://msandbu.org/how-do-you-secure-an-genai-application-or-service/

Guardrails

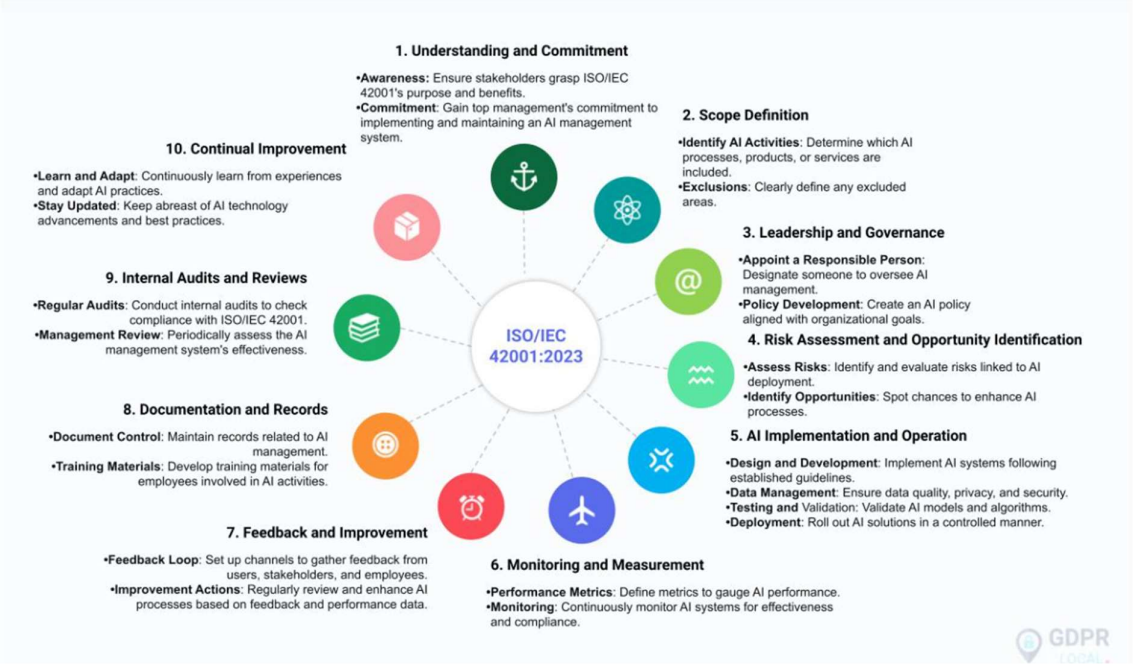


Risks, Governance And Compliance

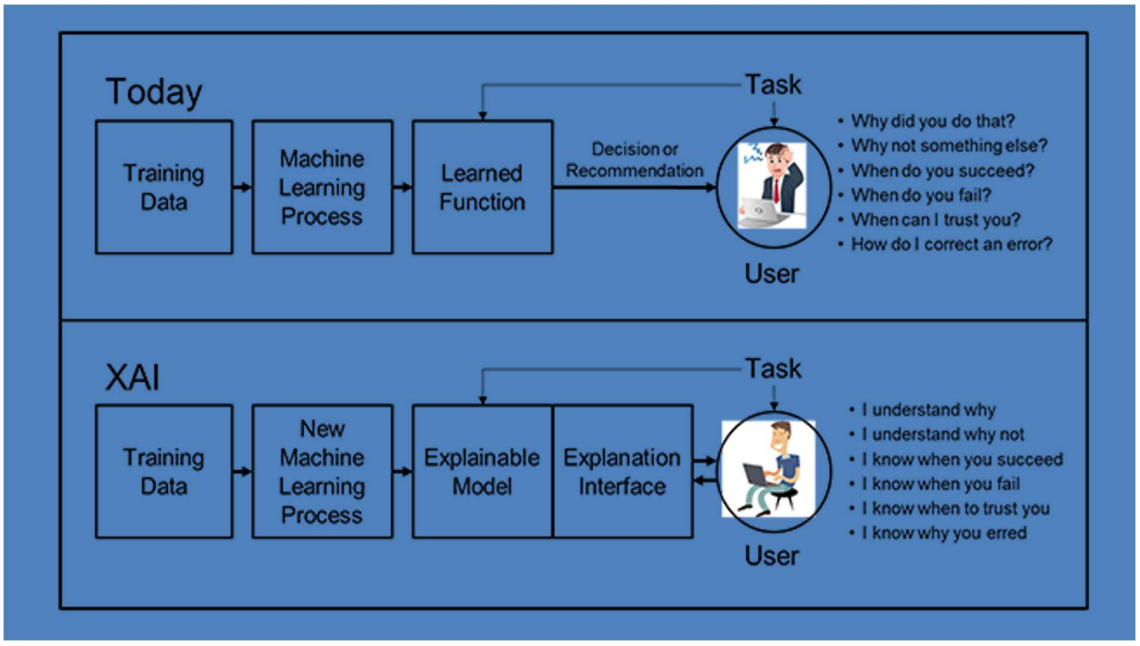
EU AI ACT

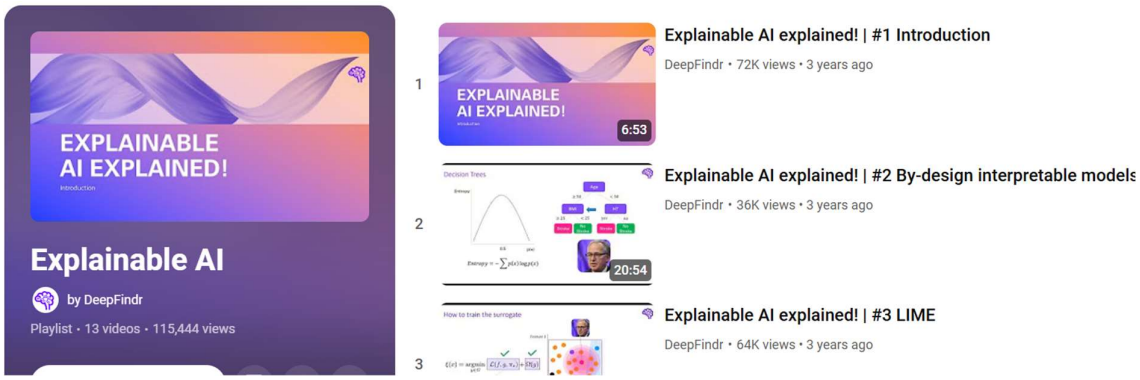


ISO 42001



Explainable AI





https://www.youtube.com/playlist?list=PLV8yxwGOxvvovp-j6ztxhF3QcKXT6vORU

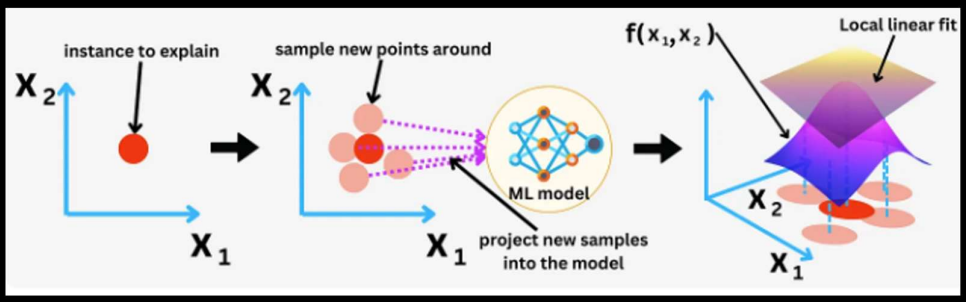
Explainable AI (XAI) frameworks – LIME and SHAP

LIME: Local Interpretable Model-agnostic Explanations

- Perturb input data
- Model predictions on perturbed data
- Train a simple interpretable model
- Local explanation

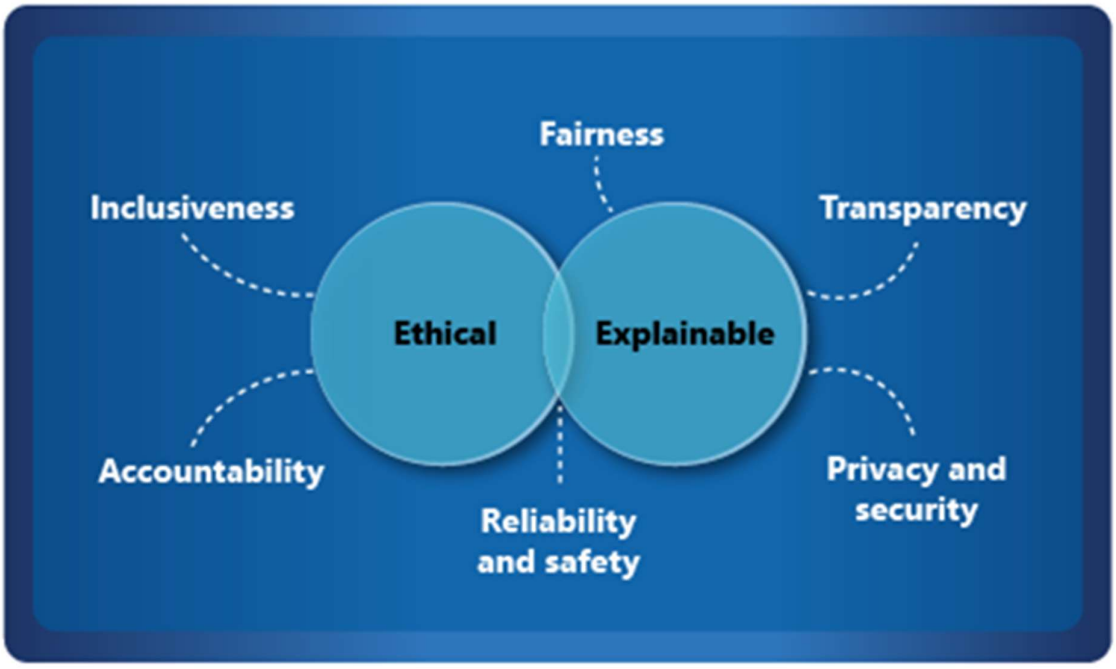
SHAP (Shapley Additive exPlanations)

- Shapley Values (Game theory concept)
- Evaluate feature contributions
- Build a SHAP explanation

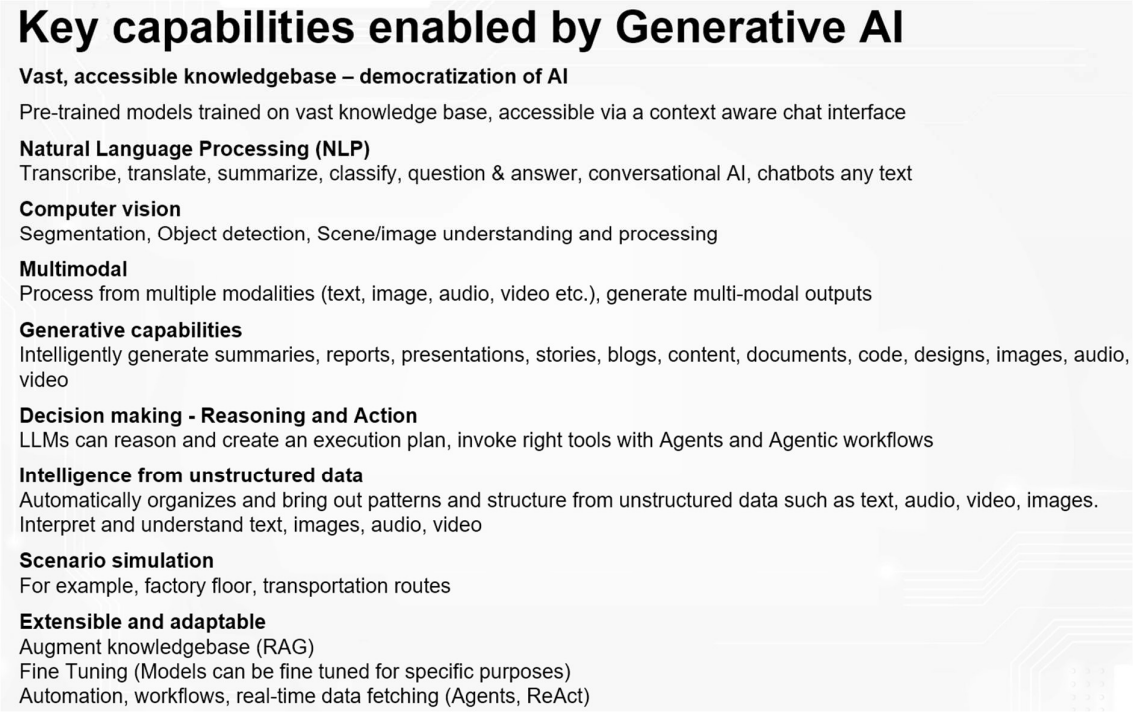


LIME

Responsible AI Framework



Gen AI Use cases Across Job Functions



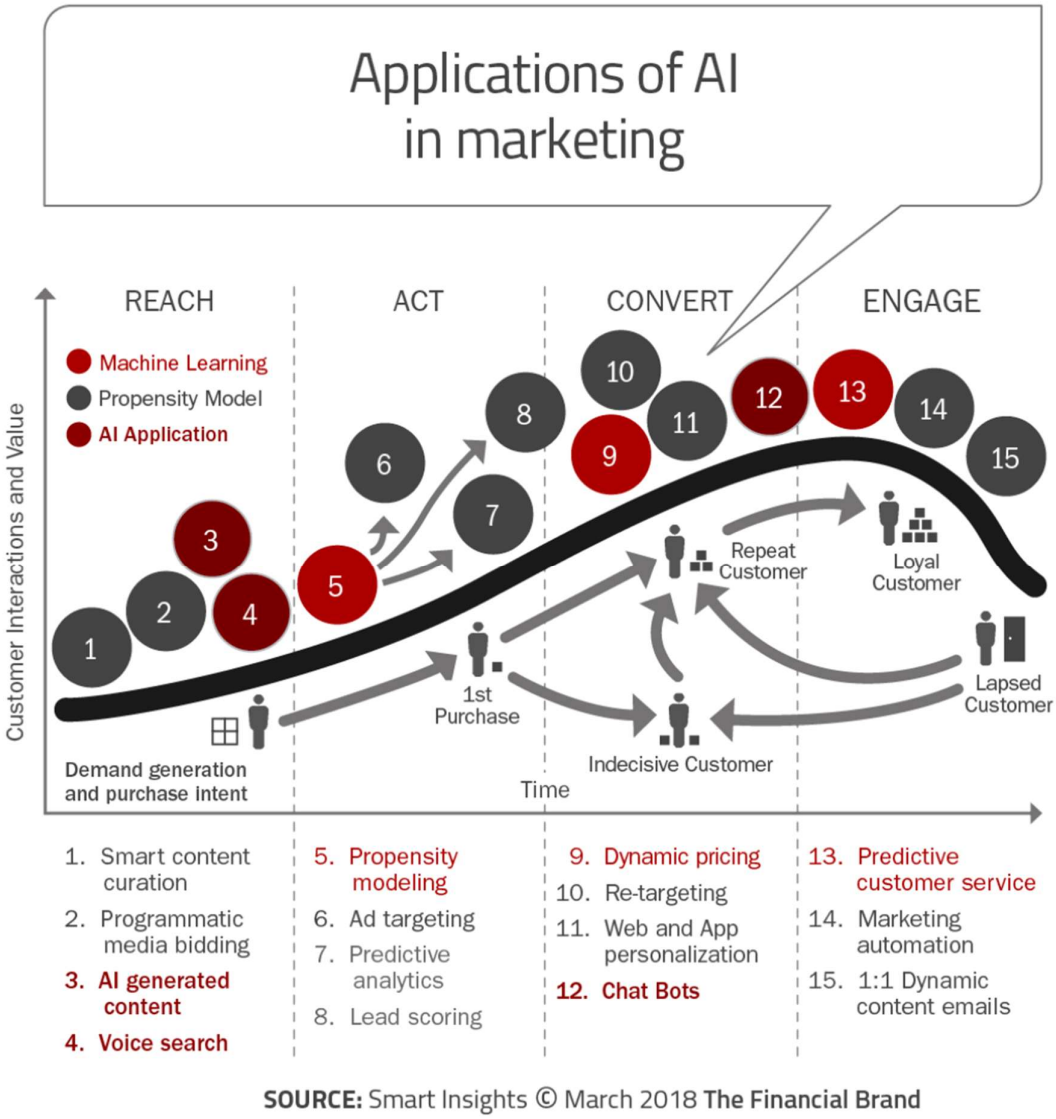
Gen AI – Impact by business functions, across industries

Generative AI productivity impact by business functions[1]

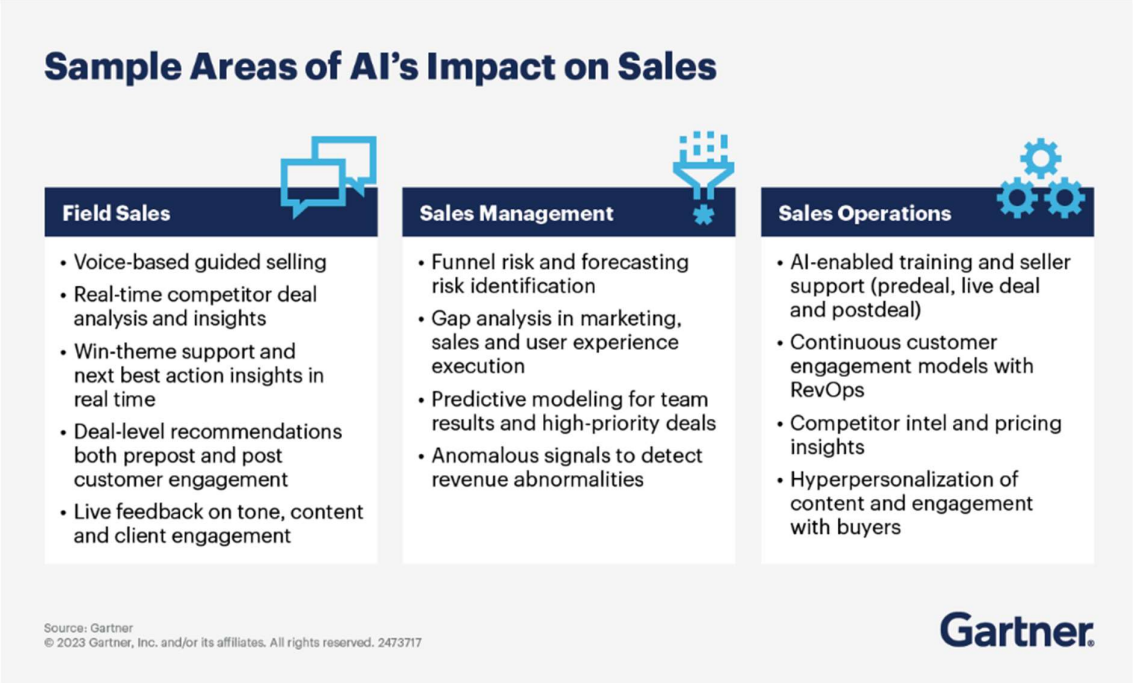
Low impact — High impact

	Total, % of industry revenue	**Total,** $ billion	Marketing and sales	Customer operations	Product R&D	Software engineering	Supply chain and operations	Risk and legal	Strategy and finance	Corporate IT[2]	Talent and organization
			760–1,200	340–470	230–420	580–1,200	290–550	180–260	120–260	40–50	60–90
High tech	4.8–9.3	240–460									
Banking	2.8–4.7	200–340									
Pharmaceuticals and medical products	2.6–4.5	60–110									
Education	2.2–4.0	120–230									
Telecommunications	2.3–3.7	60–100									
Healthcare	1.8–3.2	150–260									
Insurance	1.8–2.8	50–70									
Media and entertainment	1.8–3.1	80–130									
Advanced manufacturing[3]	1.4–2.4	170–290									
Consumer packaged goods	1.4–2.3	160–270									
Advanced electronics and semiconductors	1.3–2.3	100–170									

Marketing and Sales



SOURCE: Smart Insights © March 2018 **The Financial Brand**



HubSpot AI in Sales 2024 trends report

https://www.hubspot.com/hubfs/Content%20Offers/HubSpot%20AI%20Sales%20Trends%20Report%202024.pdf?hubs_signup-url=offers.hubspot.com%2Fai-sales&hubs_signup-cta=Submit&hubs_offer=offers.hubspot.com%2Fai-sales&_gl=1*15wcawt*_gcl_au*MjExMDM1NjAzOS4xNzM0MzY1MjUx&_ga=2.267153549.261305445.1734406649-259091869.1734406649

AI Sales Tools list:

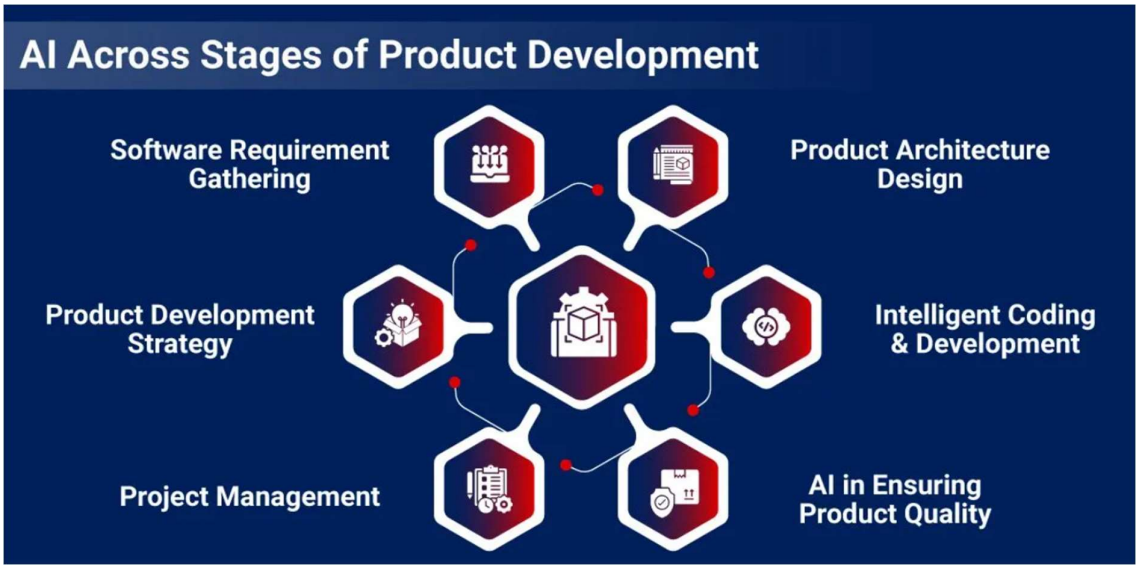
Hubspot: 10 best AI tools in 2024:

https://blog.hubspot.com/sales/ai-sales-tools

A long list of AI tools for Sales:

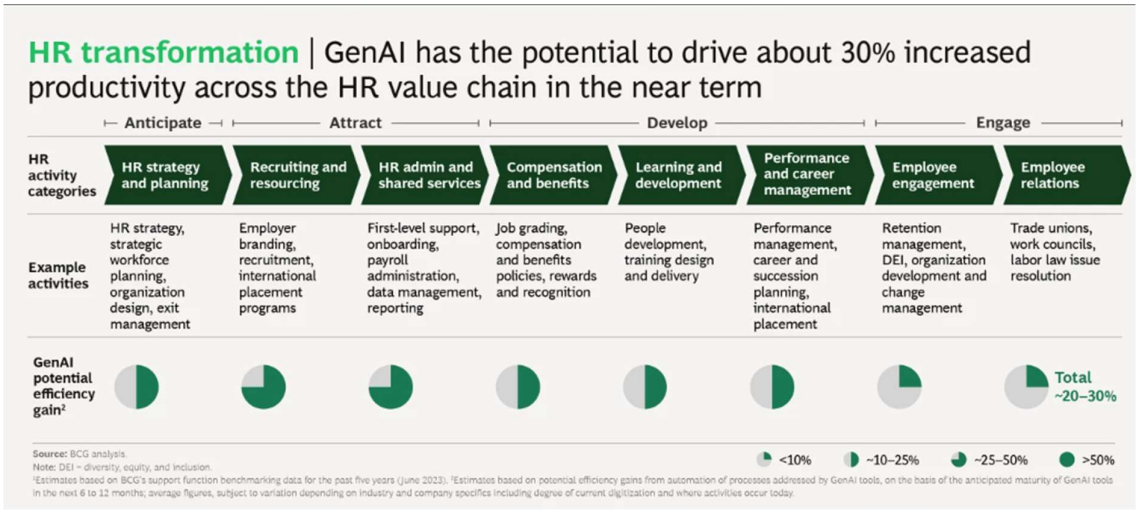
https://www.coldiq.com/ai-sales-tools

Software Development and Quality Assurance

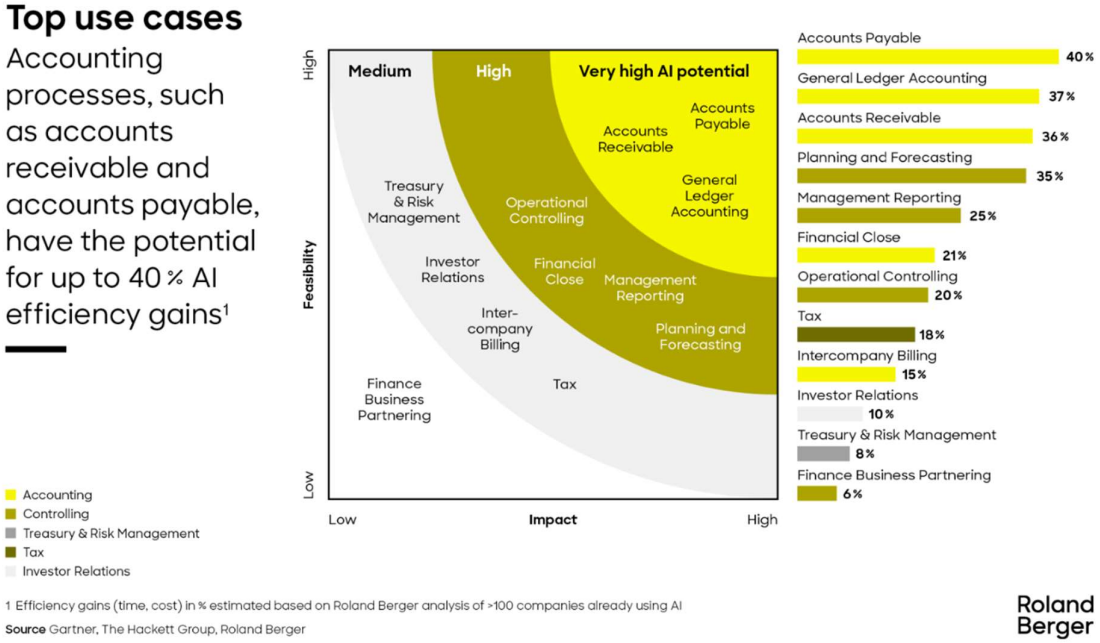


https://www.rishabhsoft.com/blog/ai-in-product-development

Human Resources



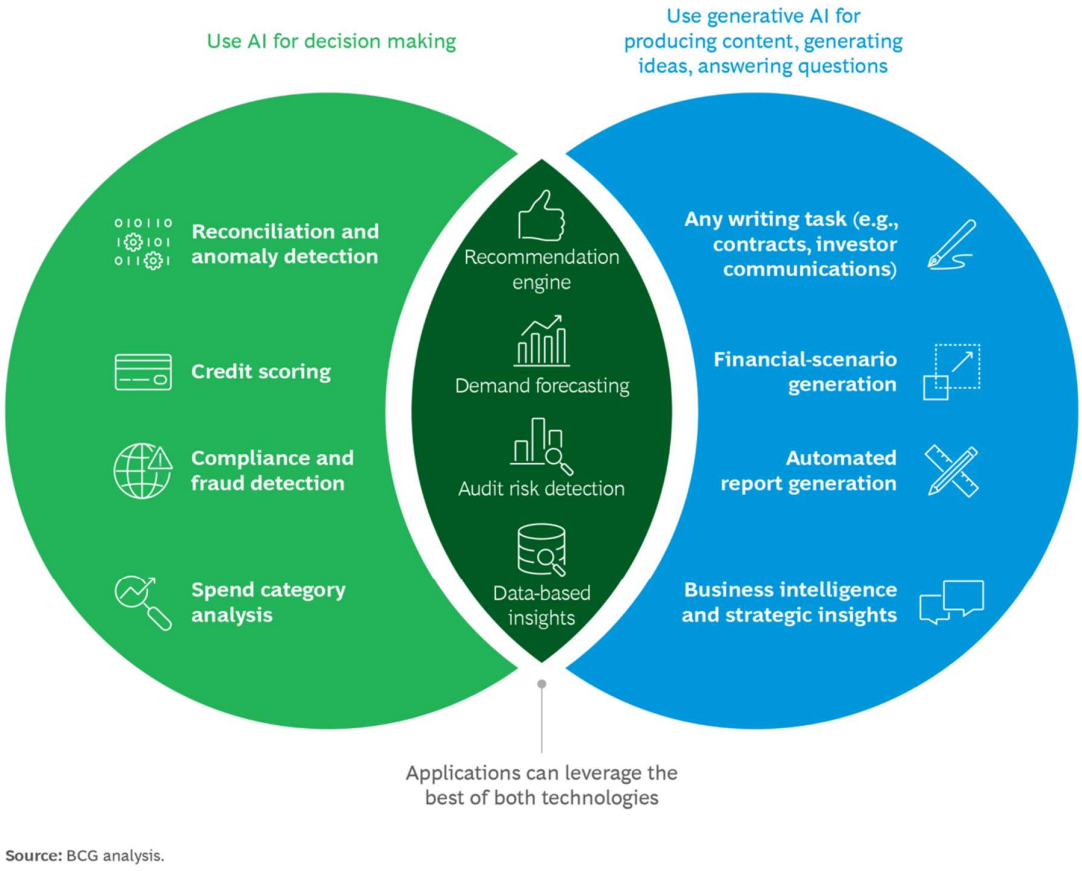
Accounting and Finance



https://www.rolandberger.com/en/Insights/Publications/Mastering-AI-in-the-finance-function.html

Gen AI Use cases Across Industries

Finance and BFSI



Source: BCG analysis.

https://www.bcg.com/publications/2023/generative-ai-in-finance-and-accounting

Moody's copilot:

https://www.youtube.com/watch?v=7O8jtu3bpS8

Morgan Stanley uses AI evals to shape the future of financial services

https://openai.com/index/morgan-stanley/

https://www.youtube.com/watch?v=BO1n3NdUWPs

Blog > Reports And Research

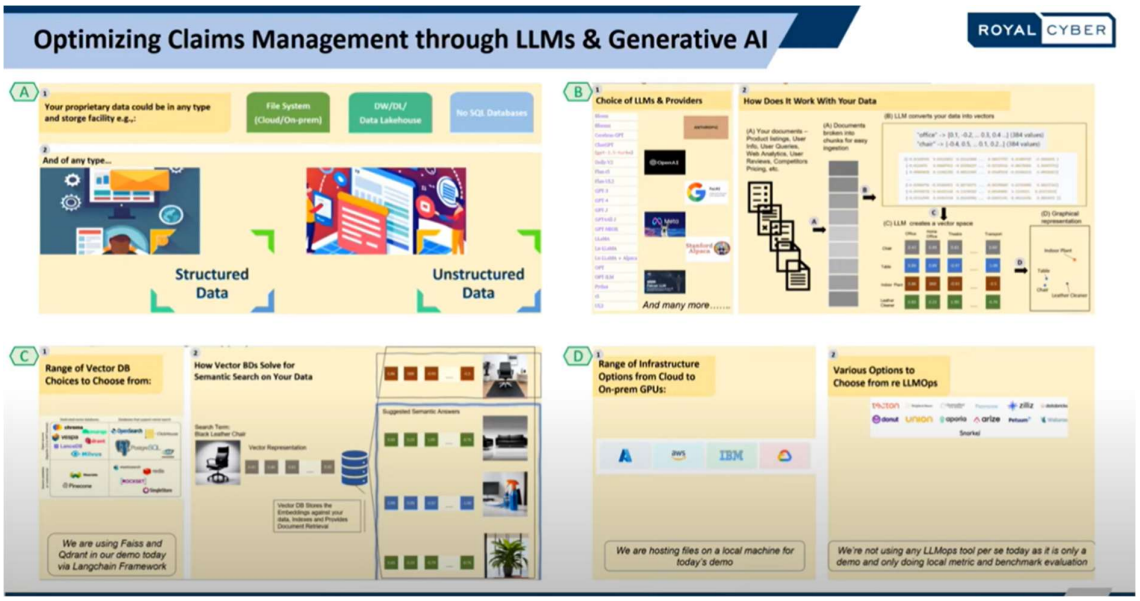
Generative AI in Finance: Real-World Examples of LLMs in Banking in 2023

August 04, 2023 5 min read Kate Kopyl

https://tovie.ai/blog/generative-ai-in-finance-real-world-examples-of-llms-in-banking-in-2023

https://www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking

INSURANCE



https://www.royalcyber.com/resources/videos/webcasts/transforming-insurance-operations-through-generative-ai/

Company financial analysis:

https://chat.financialdatasets.ai/

https://github.com/virattt/ai-financial-agent

Stocks sentiment analysis:

https://huggingface.co/spaces/Karthikeyen92/Stock-Sentiment-Analysis

LLM in fraud detection

https://www.youtube.com/watch?v=lRRtDrwd2z8

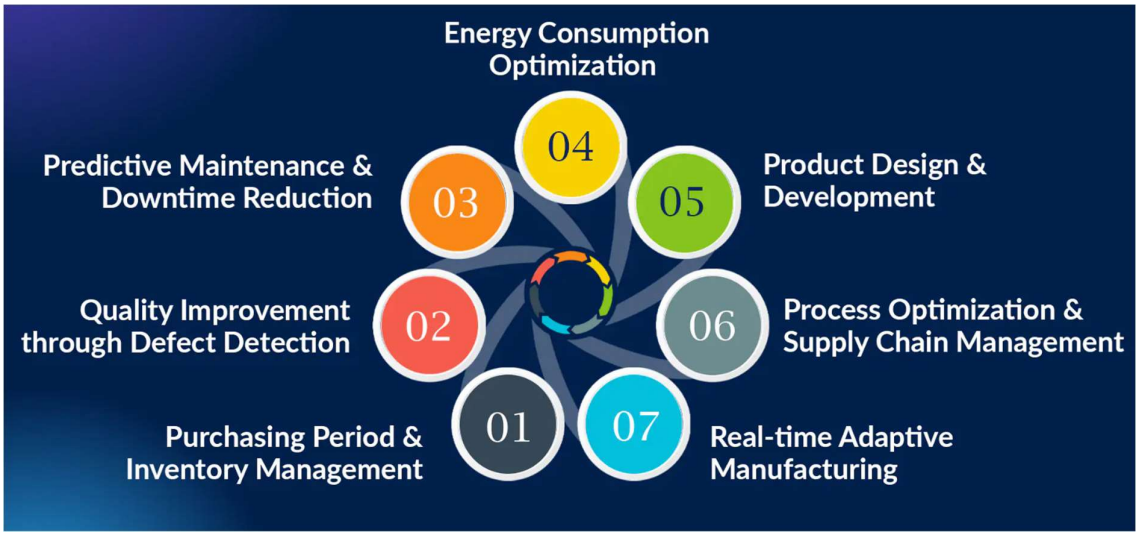
https://github.com/amitkedia007/Financial-Fraud-Detection-Using-LLMs

https://www.kaggle.com/datasets/amitkedia/financial-statement-fraud-data

Fine tuning llm for fraud detection using Azure OpenAI

https://mkonda007.medium.com/fine-tuning-ai-models-creating-a-financial-fraud-detection-assistant-c23505c8d669

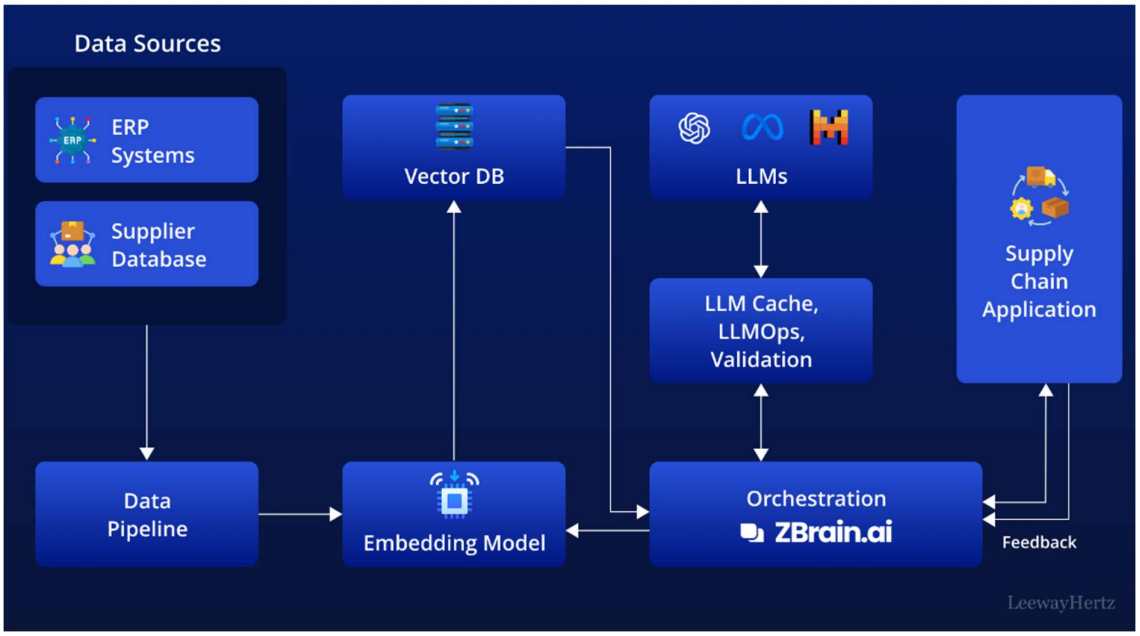
Manufacturing



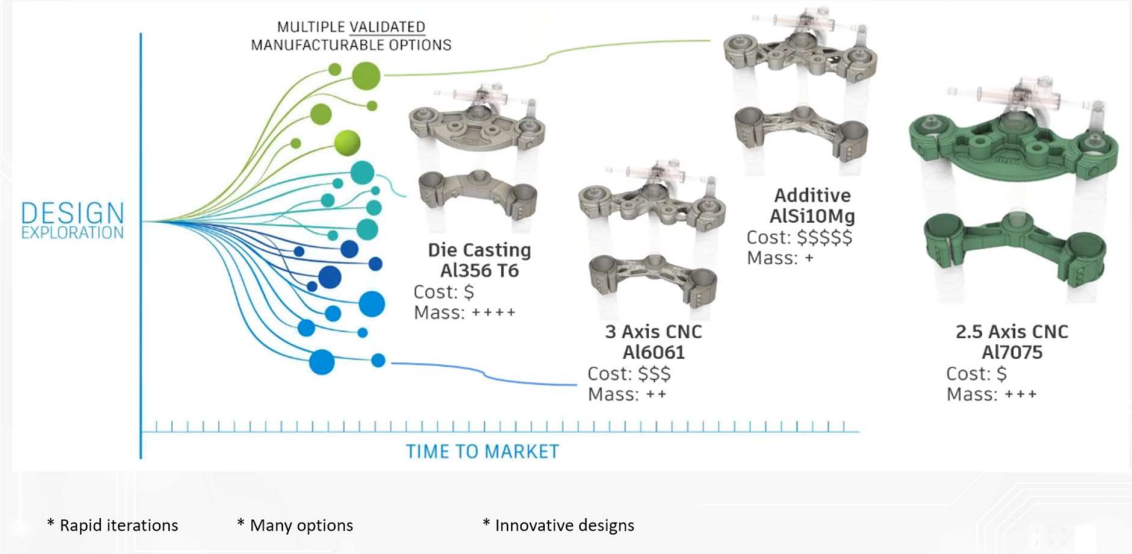
https://www.matellio.com/blog/generative-ai-in-manufacturing-use-cases/

https://nextgeninvent.com/blogs/generative-ai-in-manufacturing-use-cases/

https://www.leewayhertz.com/generative-ai-in-supply-chain/



Generative AI in CAD/CAM, Product Design:



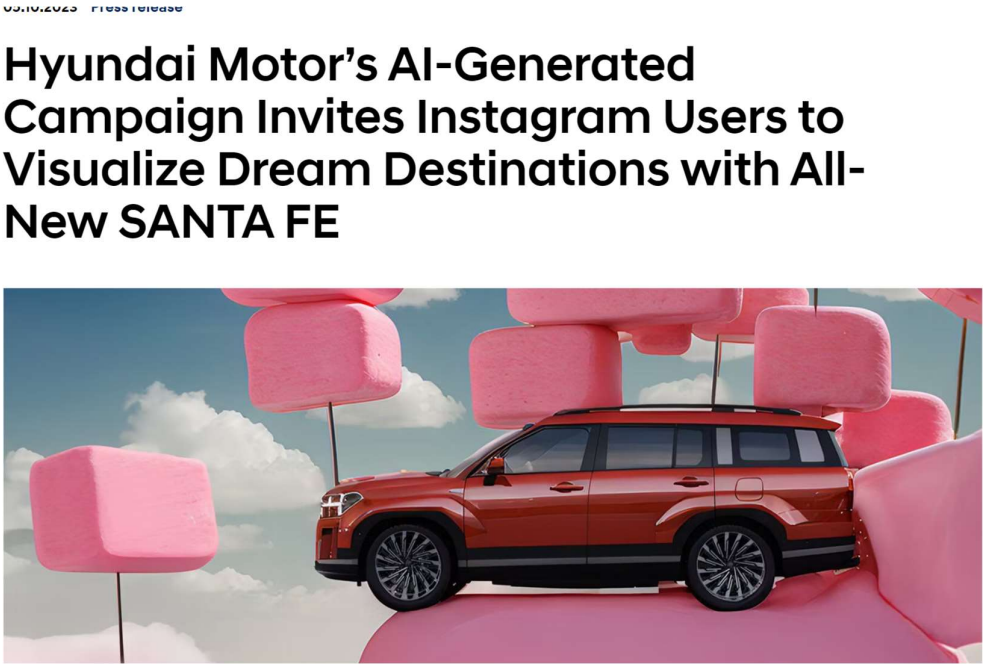
https://www.youtube.com/watch?v=PSSt8wswNJQ

https://www.youtube.com/watch?v=a7nRzOOVgB8

AI and Digital Twin:

https://www.youtube.com/watch?v=g78YHYXXils

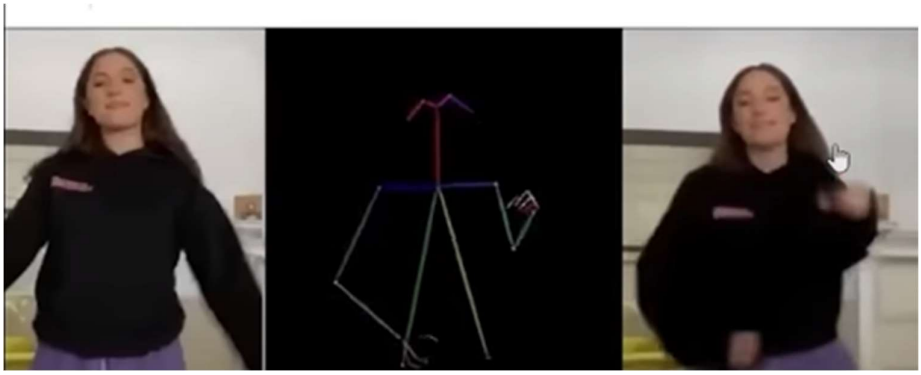
https://aichat.com/mitsubishi-motors-singapore-chatbot/



https://www.hyundai.news/eu/articles/press-releases/santa-fe-ai-generated-campaign-invites-instagram-users-to-dream-destinations.html

Media and Entertainment

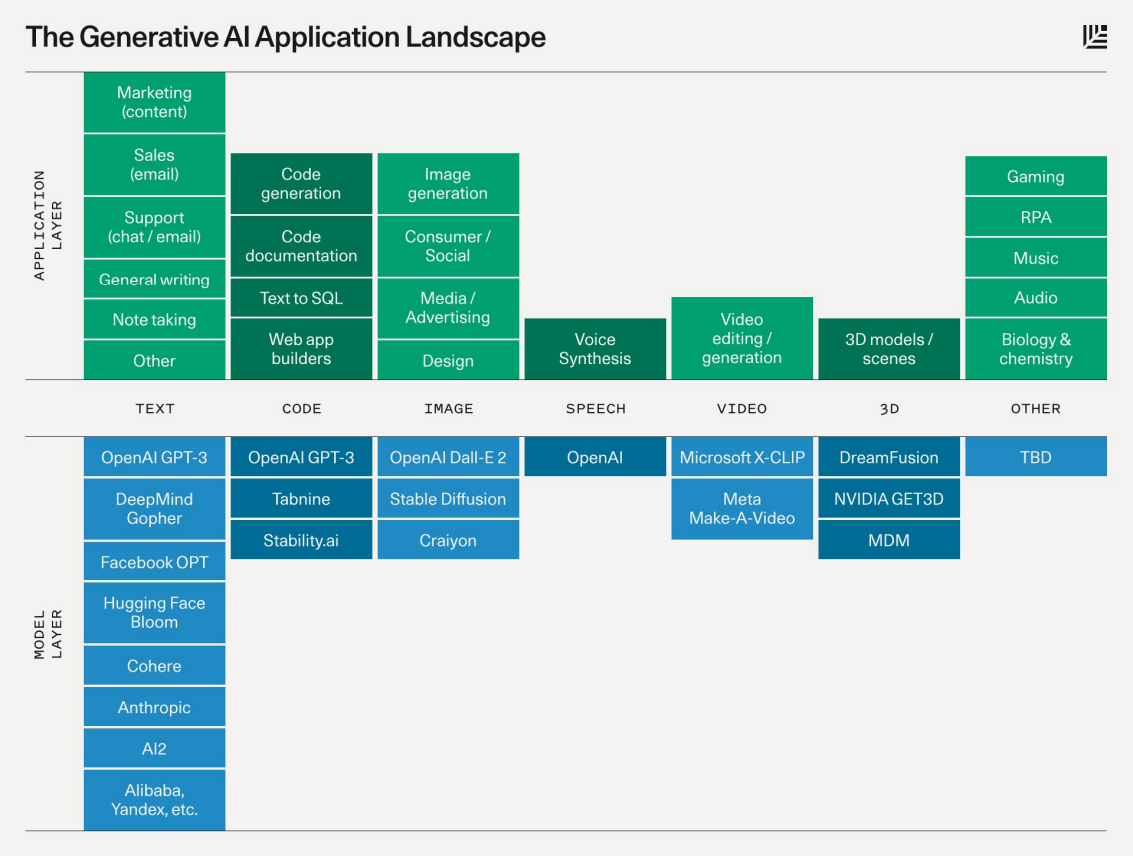
AI can make you dance



Top AI tools for content creation

https://www.instagram.com/awa_k_penn/p/DESWVNSNsiZ/

Popular Gen-AI tools and applications



A good curated list here:

https://hashcollision.substack.com/p/a-primer-on-generative-ai-genai

AI Chatbots:

OpenAI ChatGPT: https://chatgpt.com

Microsoft copilot: https://copilot.microsoft.com

Google Gemini: https://gemini.google.com

Grok from X: https://x.com/i/grok?focus=1

Anthropic Claude (Amazon backed): https://claude.ai

Perplexity AI: https://www.perplexity.ai/

Meta AI: https://www.meta.ai/

Cohere: https://dashboard.cohere.com/

Presentation Slides Generator:

https://gamma.app/

https://www.slidesai.io/

https://pptgenai.com/

Text to video:

https://invideo.io

https://sora.com/login?next=%2Flibrary

https://help.openai.com/en/articles/9957612-generating-videos-on-sora

Sora demo:

https://openai.com/index/sora/

AI Studios:

https://app.aistudios.com/dashboard

Summarize PDF , chat with pdf, generate slide deck from pdf

https://chatpdf.com

https://popai.pro

https://www.presentations.ai/tools/pdf-to-ppt

https://getcoralai.com

Excel AI, Dashboards with chat

Arcwise:

https://arcwise.app/

Google AI Studio:

https://aistudio.google.com/live

See demo at:

https://www.youtube.com/watch?v=3cvczHJSRNs

Excel Dashboard AI:

https://www.exceldashboard.ai/

Bricks:

https://www.thebricks.com/?utm_source=bricks&utm_medium=resources&utm_campaign=how-to-create-a-sales-dashboard-with-ai

https://www.thebricks.com/resources/how-to-create-a-sales-dashboard-with-ai

Demo:

https://www.youtube.com/watch?v=-tl-oxY-DAE

https://www.youtube.com/watch?v=1vz6EkUhbOg

Marketing and Sales

Apollo: https://www.apollo.io/

Hubspot: 10 best AI tools in 2024:

https://blog.hubspot.com/sales/ai-sales-tools

A long list of AI tools for Sales:

https://www.coldiq.com/ai-sales-tools

Image Generation:

Midjourney: https://www.midjourney.com/explore?tab=top

Conversational AI and Agents/Audio:
Eleven Labs:

https://elevenlabs.io/app/conversational-ai

Murf:

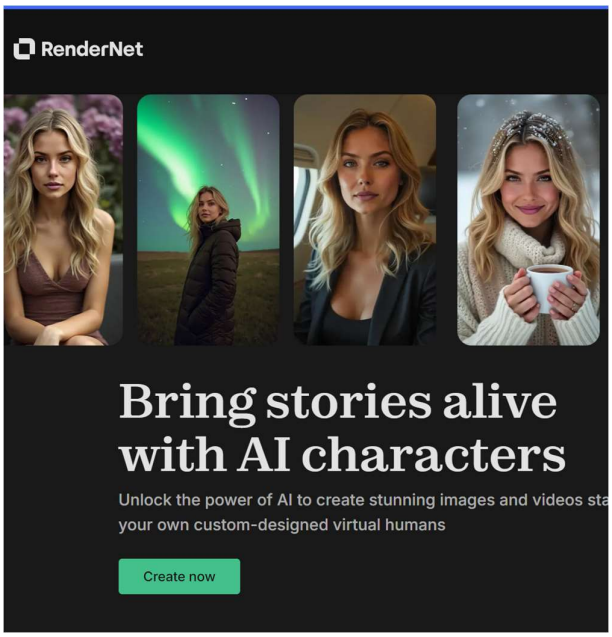
https://murf.ai

Vision Agent

https://va.landing.ai/agent

RenderNet - create stunning images and videos starring your own custom-designed virtual humans, including voice generation

https://rendernet.ai/



Articles, Blogs, Newsletters, SM Handles

Good sites and blogs for latest in AI:

https://lablab.ai/blog

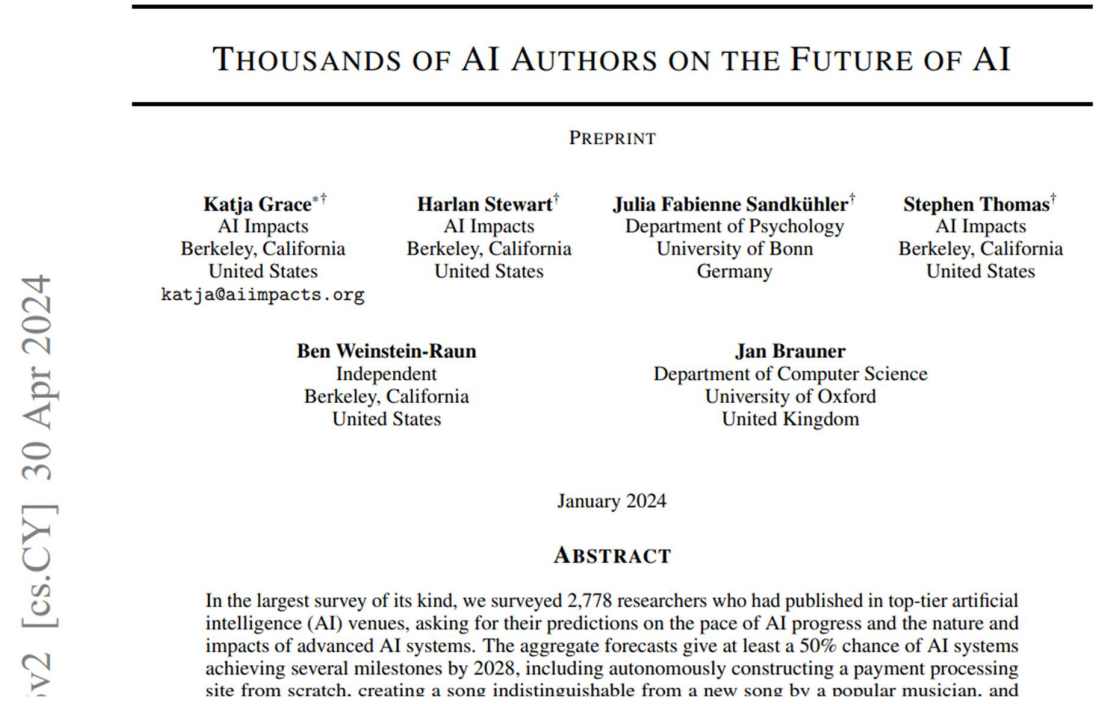
https://www.thedeepview.co/

https://arxiv.org/pdf/2401.02843

https://medium.com/@akash.kesrwani99/understanding-next-token-prediction-concept-to-code-1st-part-7054dabda347

Image embeddings:

https://encord.com/blog/image-embeddings-to-improve-model-performance/



THOUSANDS OF AI AUTHORS ON THE FUTURE OF AI

PREPRINT

Katja Grace[*†]
AI Impacts
Berkeley, California
United States
katja@aiimpacts.org

Harlan Stewart[†]
AI Impacts
Berkeley, California
United States

Julia Fabienne Sandkühler[†]
Department of Psychology
University of Bonn
Germany

Stephen Thomas[†]
AI Impacts
Berkeley, California
United States

Ben Weinstein-Raun
Independent
Berkeley, California
United States

Jan Brauner
Department of Computer Science
University of Oxford
United Kingdom

January 2024

ABSTRACT

In the largest survey of its kind, we surveyed 2,778 researchers who had published in top-tier artificial intelligence (AI) venues, asking for their predictions on the pace of AI progress and the nature and impacts of advanced AI systems. The aggregate forecasts give at least a 50% chance of AI systems achieving several milestones by 2028, including autonomously constructing a payment processing site from scratch, creating a song indistinguishable from a new song by a popular musician, and

https://hashcollision.substack.com/p/a-primer-on-generative-ai-genai

https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#industry-impacts

https://www.deeplearning.ai/

Measuring ROI of Generative AI

https://www.gartner.com/en/articles/take-this-view-to-assess-roi-for-generative-ai#:~:text=Generative%20AI%20promises%20unprecedented%20productivity,Quick%20wins
https://embracingenigmas.substack.com/p/next-token-prediction-is-a-fundamental

Notable websites on AI:

datacamp.com
deeplearning.ai
towardsai.net

analyticsvidya

Notable Instagram Handles on AI:

Name	URL	Description
Aiverse	https://www.instagram.com/aiversepage/	Shares insights and updates on artificial intelligence and its applications.
Artificial Intelligence News	https://www.instagram.com/artificialintelligencenews.in/	Provides the latest news and developments in AI and machine learning.
AI Explaining	https://www.instagram.com/aiexplaining/	Breaks down complex AI concepts into understandable content for enthusiasts.
NVIDIA AI	https://www.instagram.com/nvidiaai/	Showcases advancements and applications of AI technology by NVIDIA.
Microsoft Research	https://www.instagram.com/microsoftresearch/	Highlights research and innovations in AI and other computing fields by Microsoft.
Amazon Science	https://www.instagram.com/amazon.science/	Shares Amazon's scientific research and developments in AI and related areas.
MIT Technology Review	https://www.instagram.com/technologyreview/	Covers the latest in technology, including AI breakthroughs and analyses.
Deep Learning Nerds	https://www.instagram.com/deeplearningnerds/	Focuses on AI, data science, and machine learning content for enthusiasts and professionals.
Best of AI	https://www.instagram.com/bestofaidaily/	Curates top AI-related content, news, and innovations from around the world.
AI Art Universe	https://www.instagram.com/aiartuniverse/	Showcases AI-generated art and explores the intersection of creativity and technology.
AI Trends	https://www.instagram.com/aitrends/	Provides updates on the latest trends and developments in artificial intelligence.
AI Daily	https://www.instagram.com/aidaily/	Shares daily news, insights, and articles related to AI and machine learning.
AI Innovations	https://www.instagram.com/ai_innovations/	Focuses on innovative applications and breakthroughs in the field of AI.
AI Hub	https://www.instagram.com/aihub/	A community page sharing AI news, research, and technological advancements.

Notable Twitter (X) handles on AI:

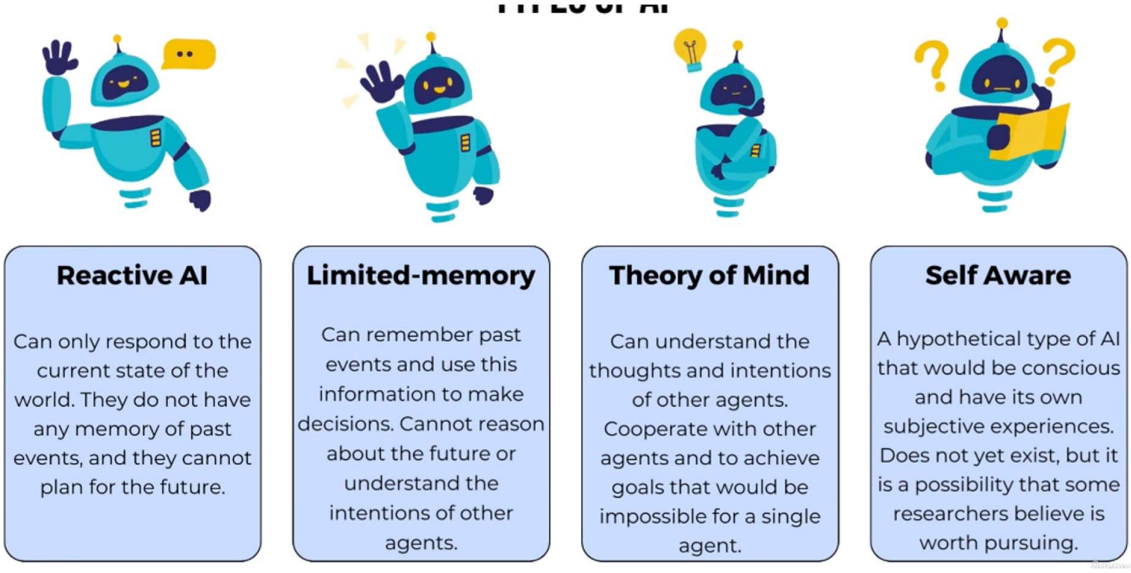
X Handle	URL	Description
Tut_ml	https://x.com/tut_ml	ML Blogs
LangChainAI	https://x.com/LangChainAI	Building applications with LLMs through composability.
Ethan Mollick	https://x.com/emollick	Professor at Wharton, sharing insights on AI, innovation, and education.
Omar Sar0	https://x.com/omarsar0	AI researcher and enthusiast, focusing on machine learning advancements.
Rob Feers	https://x.com/rfeers	Data scientist sharing thoughts on AI, data analysis, and technology trends.
DataCamp	https://x.com/DataCamp	Learn data skills online at your own pace.
KDnuggets	https://x.com/KDnuggets	Leading site on AI, Analytics, Big Data, Data Mining, and Data Science.
iScienceLuvr	https://x.com/iScienceLuvr	Enthusiast sharing the latest in AI research and machine learning developments.
Shubham Saboo	https://x.com/Saboo_Shubham_	AI practitioner and educator, providing tutorials and insights on machine learning.
Victor Dey	https://x.com/victor_explore	AI journalist exploring the latest trends and breakthroughs in artificial intelligence.
Sam Altman	https://x.com/sama	CEO of OpenAI, sharing insights on AI development and its implications.
Greg Brockman	https://x.com/gdb	Co-founder and Chairman of OpenAI, discussing AI advancements and research.
Demis Hassabis	https://x.com/demishassabis	CEO of DeepMind, focusing on AI research and its applications.
Andrew Ng	https://x.com/AndrewYNg	Co-founder of Coursera and AI researcher, offering insights into machine learning and education.
Allie K. Miller	https://x.com/alliekmiller	AI leader and investor, sharing knowledge on machine learning and AI startups.

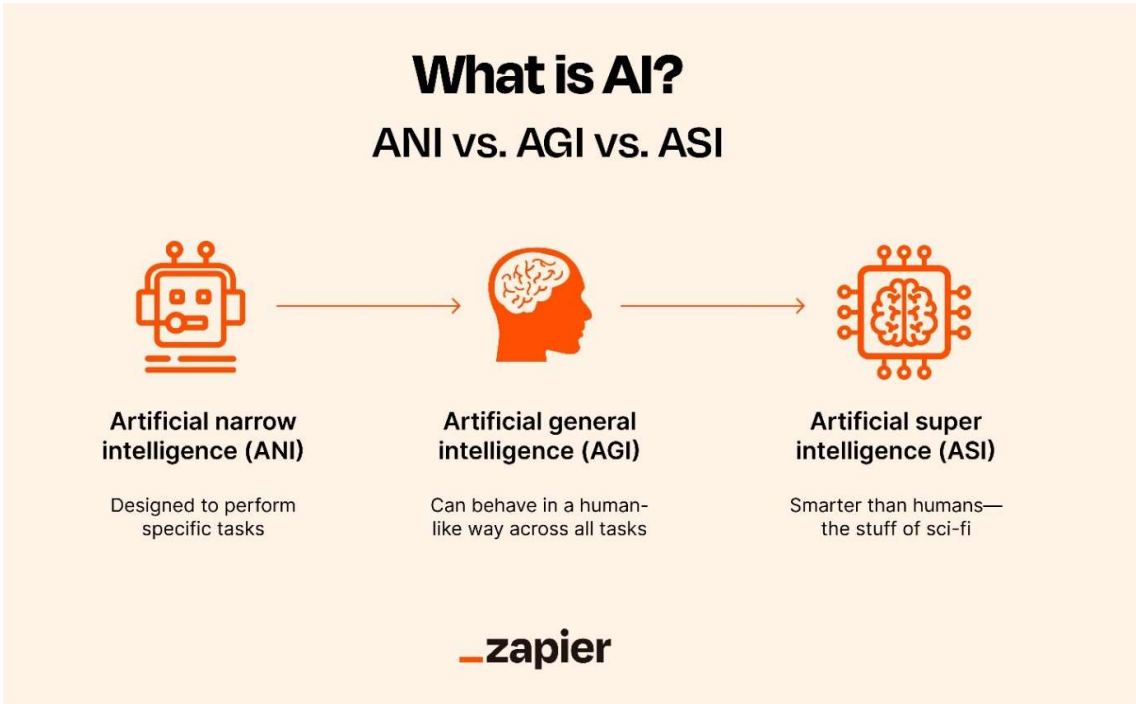
Notable YouTube channels on AI:

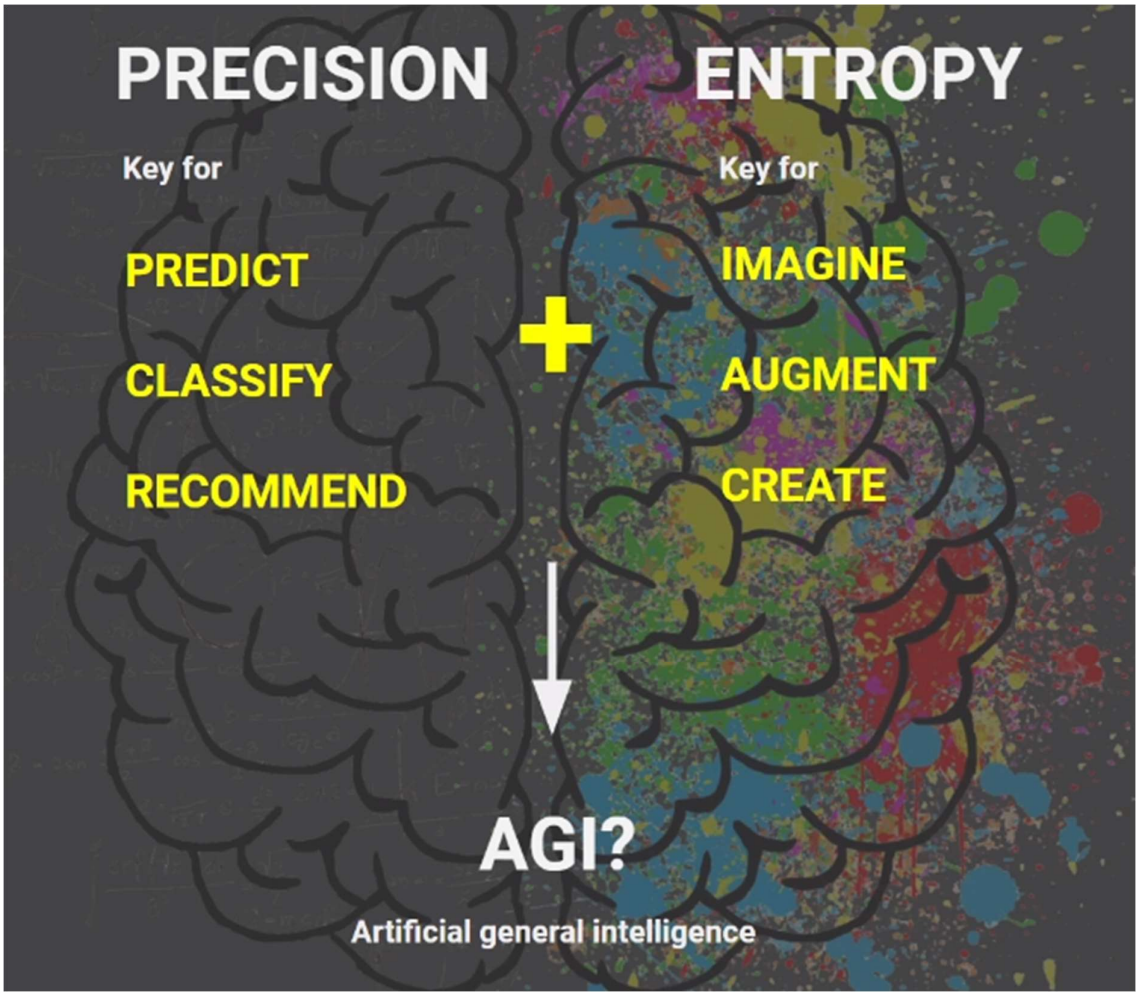
Name	Description	URL
Infinite Codes	A channel dedicated to coding tutorials and software development insights.	https://www.youtube.com/@InfiniteCodes_
AI Papers Academy	Focuses on discussing and analyzing academic papers in the field of artificial intelligence.	https://www.youtube.com/@aipapersacademy
GPT for Work	Provides tutorials on integrating GPT models into Excel, Word, Sheets, and Docs.	https://www.youtube.com/@gptforwork
StatQuest	Simplifies complex statistical concepts with clear and engaging explanations.	https://www.youtube.com/@statquest
Grant Sanderson	Creator of 3Blue1Brown, offering deep dives into mathematical	https://www.youtube.com/@GrantSanderson

	concepts with visual explanations.	
AI Revolution	Explores the latest advancements and trends in artificial intelligence and machine learning.	https://www.youtube.com/@airevolutionx
Vector Institute	Shares research and insights from the Vector Institute, focusing on AI and machine learning.	https://www.youtube.com/@vectorinstituteai
The AI Nexus	Discusses AI developments, tools, and their applications across various industries.	https://www.youtube.com/@TheAINexusOfficial
Matthew Berman	Offers content on AI, technology, and their intersection with society.	https://www.youtube.com/@matthew_berman
AI Thought Leaders	Features interviews and discussions with leading experts in the field of artificial intelligence.	https://www.youtube.com/@AIThoughtLeaders
3Blue1Brown	Visualizes mathematical concepts, making them accessible and engaging for a broad audience.	https://www.youtube.com/@3blue1brown
TheAIGRID	Focuses on AI research, tutorials, and the latest news in the AI community.	https://www.youtube.com/@TheAiGrid
Simplilearn	Provides online training videos on various topics, including AI, data science, and programming.	https://www.youtube.com/@SimplilearnOfficial
Kore.ai	Specializes in AI-powered chatbots and conversational AI platform insights.	https://www.youtube.com/@Koreai
AI Uncovered	Delves into AI technologies, their applications, and implications in the modern world.	https://www.youtube.com/@AI.Uncovered
Hume AI	Explores the intersection of AI and human emotions, focusing on empathetic AI development.	https://www.youtube.com/@hume_ai
The AI Hacker	Provides tutorials and insights on AI programming, hacking, and development techniques.	https://www.youtube.com/@theaihacker777
Arseny Shatokhin	Shares content on AI, machine learning, and related technological advancements.	https://www.youtube.com/@vrsen
AI News	Offers the latest news and updates in the world of artificial intelligence.	https://www.youtube.com/@AINewsOfficial
David Shapiro	Discusses AI development, programming, and the ethical considerations surrounding AI.	https://www.youtube.com/@DaveShap

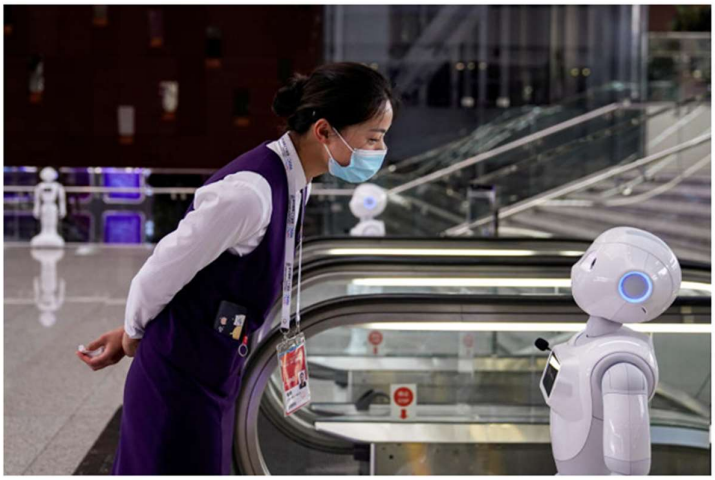
What does the future hold?

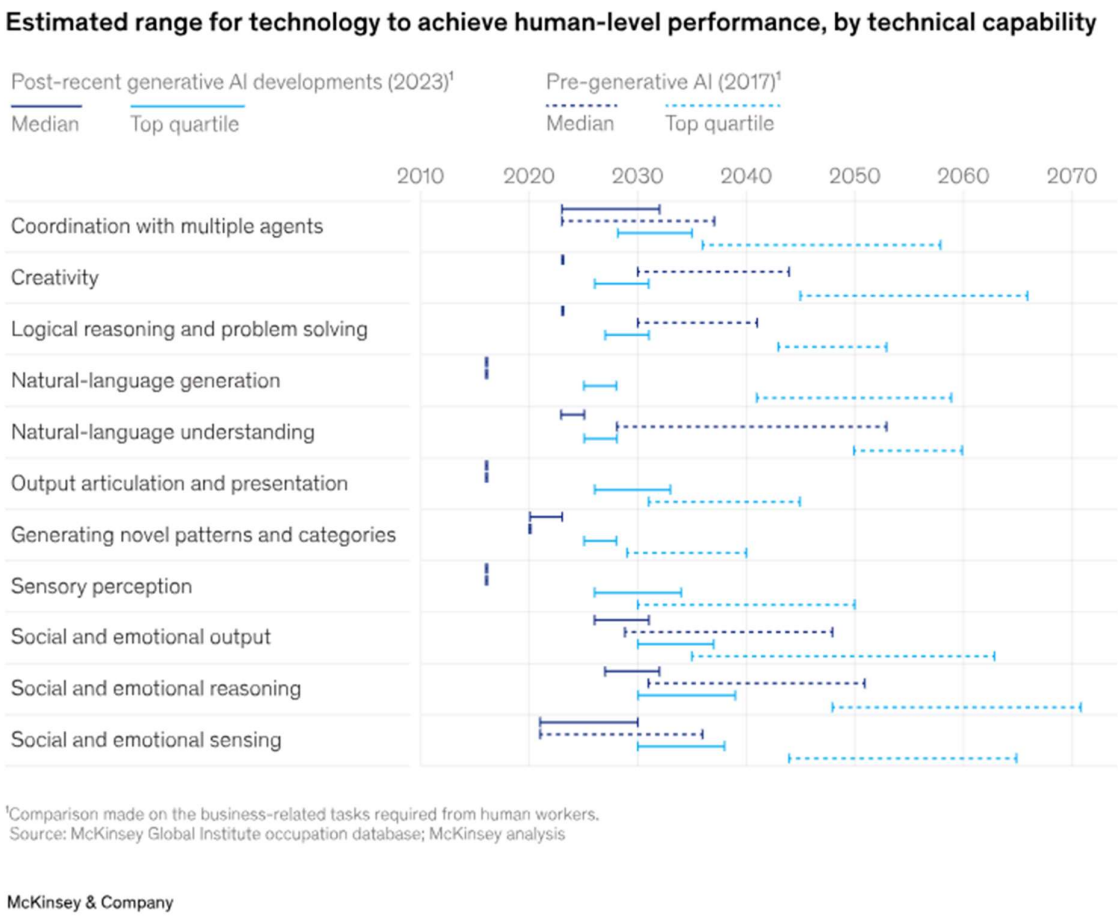




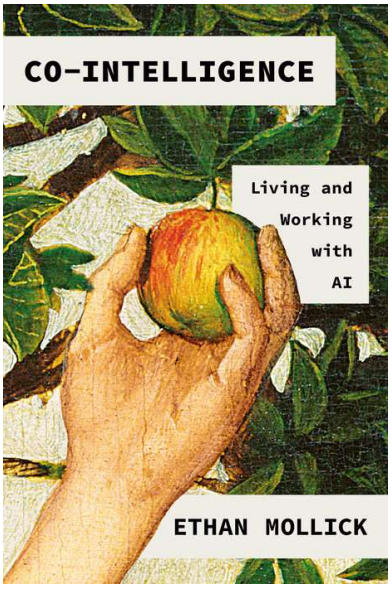


Sentience refers to the capacity to experience feelings and sensations, a concept traditionally associated with living beings.

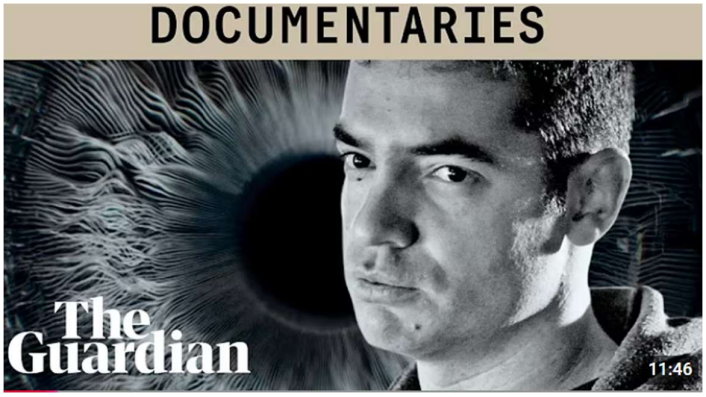




https://www.penguinrandomhouse.com/books/741805/co-intelligence-by-ethan-mollick/?ref=PRH410E2C567AF



Ilya interview with the GUARDIAN:



https://www.youtube.com/watch?v=9iqn1HhFJ6c

THANK YOU!!!

Prepared By: Vijay Agrawal

Connect on LinkedIn: https://www.linkedin.com/in/agrawalvijay/

Disclaimers:

1) These notes are meant as supplemental reading for participants in the author's Gen AI training program.
2) They are compendium of author's own notes and snippets from a variety of publicly available sources. Citations and source credits provided where applicable.
3) Author does not endorse any of the mentioned tools nor does he have any commercial interests with any of the vendors mentioned
4) There are no explicit or implied warranties – the materials are provided as-is for reference and learning
5) Highly recommended to do own research when applying Gen AI to real life problems as this space is rapidly evolving and new tools and techniques are coming up every day.