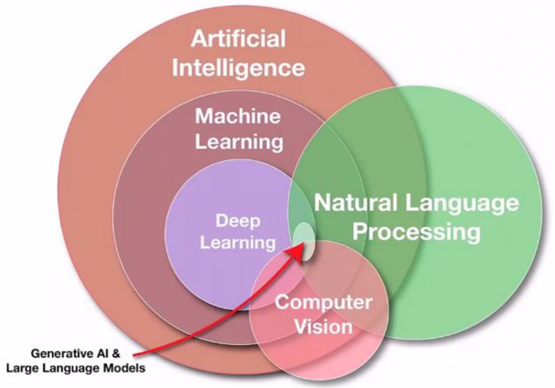
A Comprehensive Introduction to Artificial Intelligence

Zacharie GUIBERT, CQF, FRM

25 March 2025

Definition



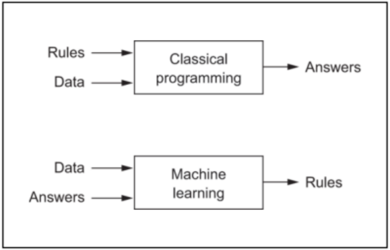
- **Intelligence**: Ability to process information to make informed future decisions
- **Artificial Intelligence**: When done by a machine

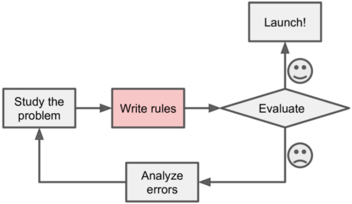
Definition

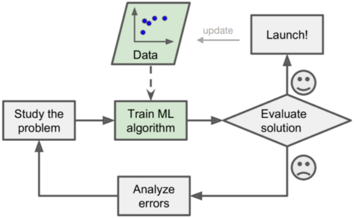
- **Artificial Intelligence**: Tasks normally requiring human intelligence - perception, reasoning, and decision-making
- **Machine Learning**: Algorithms and models that can learn from data and make predictions or decisions based on that data
- **Deep Learning**: Multiple layers of interconnected nodes (neurons) that process and transform data, inspired by the structure and function of the human brain. Automatic learning of hierarchical representations of data, allowing them to extract complex features and patterns.
- **Natural Language Processing**: Comprehend, interpret, and interact with human language. Bridge between human communication and machine intelligence, enabling computers to understand the nuances of language and respond meaningfully.
- **Computer vision**: Perceive and comprehend visual information from the real world, much like human vision.
- **Generative AI**: Create new and innovative content - images, music, text, and even simulate human-like conversations

What is (not) Machine Learning

- Classical modeling: low need of data - simulations
- Machine learning: data consuming - learns from data
- Key similarity : Model and predict
- Key difference: Importance of the feedback loop in ML

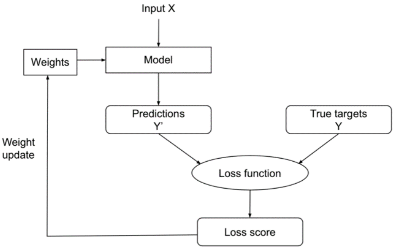






Learning process

Optimization problem to learn a useful representation of the input data



- X (features): anything - fundamental / technical / sentiment indicators, news heatmap, CDS, yield curves, tabular data, ...
- Model (learning algorithm) : KNN, SVM, Random Forest, Neutral Networks, ...
- Weight update : Minimization of the loss score until global minima reached

Find rules (weights) which allow the closest possible solution to the observed

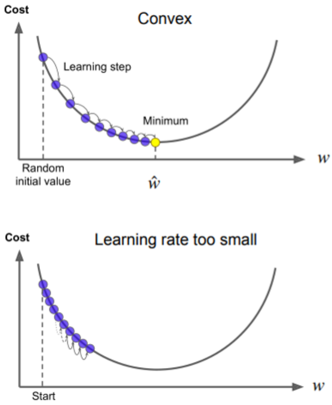
Loss Function

- Measure of how well the model predicts VS the actual target
- Loss function must be convex so that the model converges towards a unique solution (global minima).

REGRESSION	**CLASSIFICATION**
Quantitative prediction: *Stock price in 10 days*	Qualitative prediction: *Up or down trend in 10 days*
$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$\text{BCE} = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i)$ $+ (1 - y_i)\log(1-\hat{y}_i)$

Gradient descent

Descent step: $W_{new} \leftarrow W_{old} - \eta \frac{\partial J(W)}{\partial W}$, where η is the learning rate

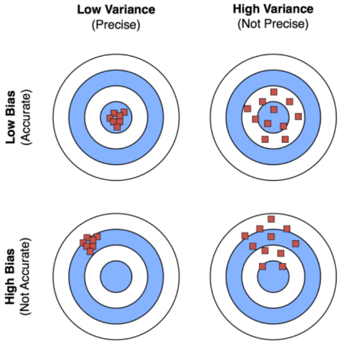


1. How does the machine learn?
   - Introduction
   - Learning process
   - Learning challenges
2. Machine Learning paradigms
3. From AI to Generative AI
4. Summary

Bias and Variance

Bias: Error introduced by approximating a real-world problem by a simplified model. High bias doesn't capture underlying patterns well (underfits).

Variance: Error introduced by the model sensitivity to small fluctuations in the dataset. High variance fits the training data too closely, e.g. its noise and outliers (overfits)

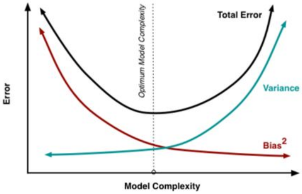


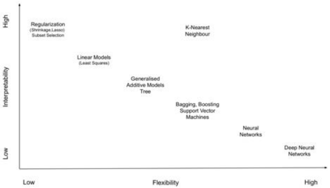
Bias Variance trade-off

- Estimation of $Y = f(X) + \epsilon$ by $\hat{f}(X)$ with a ML approach
- Expected squared prediction error $Err(S) = E[(Y - \hat{Y})^2]$ (1)
- Integration of $E[\hat{y}]$ in (1):

$$Err(S) = (E[\hat{y}] - y)^2 + E[(\hat{y} - E[\hat{y}])^2] + \sigma_e^2$$

$$Err(S) = Biais^2 + Variance + Irreductible\ Error$$



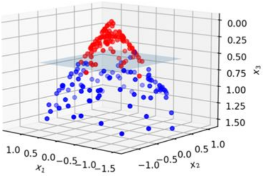


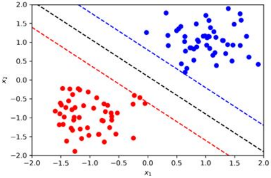
Increase (decrease) in model complexity increases (decreases) its variance and reduces (increases) its bias.

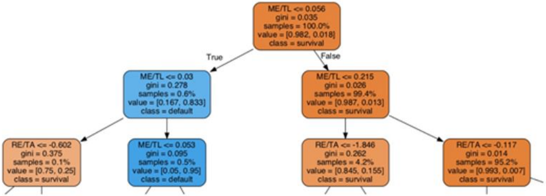
Data representation

ML $\sim$ Statistical learning $\sim$ Learning useful representation of input data

Learning $\sim$ Automatic search for better representation

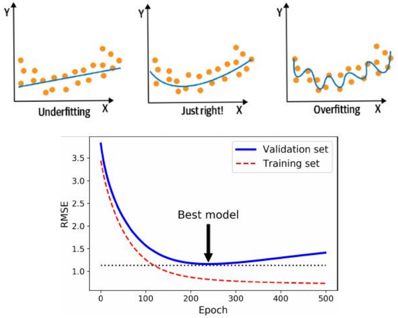






Under / Over fitting

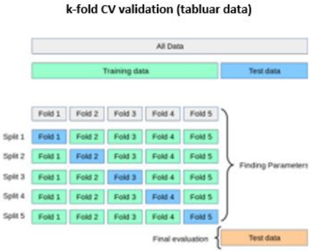
- Underfitting: Model has no capacity to fully learn from the data
- Overfitting: Model does not generalize well with data (eg memorize)

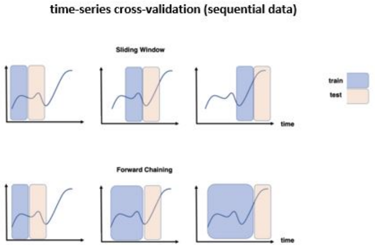


Techniques to limit overfitting: Early stopping, reduce number of features, regularization, cross-validation, etc

Cross validation

Draw sample from a training set and refit a model on each sample (resampling)

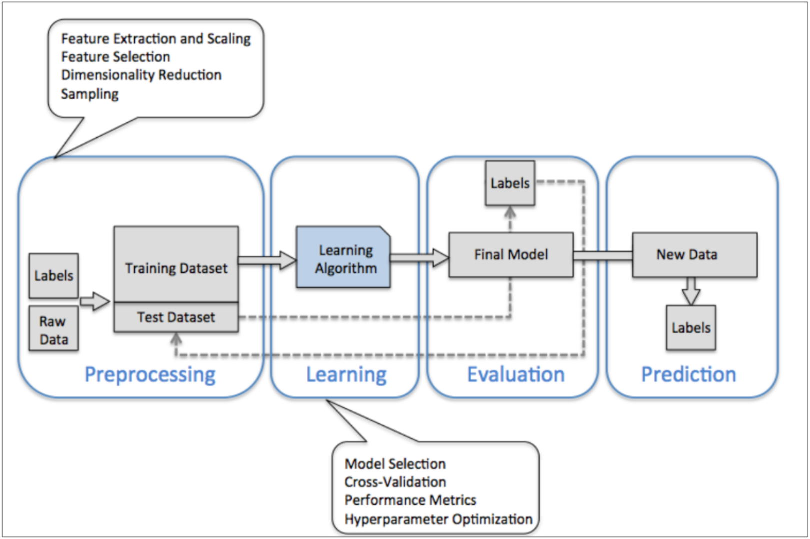




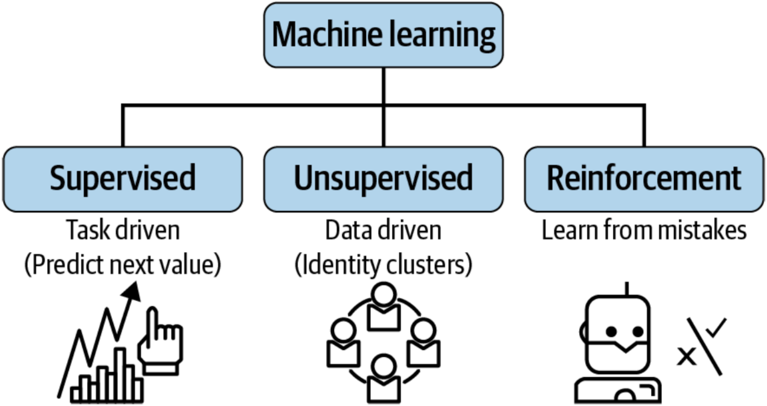
Financial time series:

- Beware of look-ahead bias (data leakage)
- Sequential in nature, characterised by the correlation between observations
- Classical CV assumes samples are iid, hence poor estimates for TS

Predictive Modeling Process

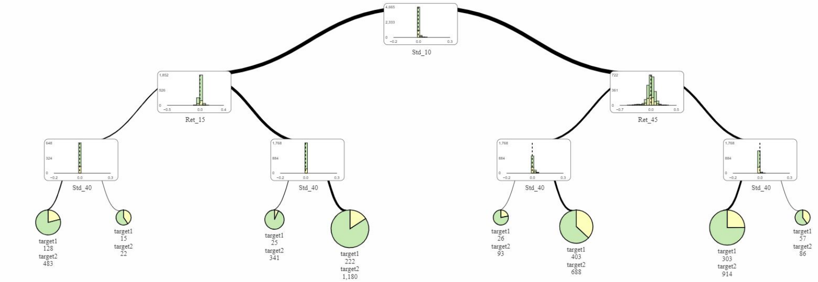


Machine Learning paradigms



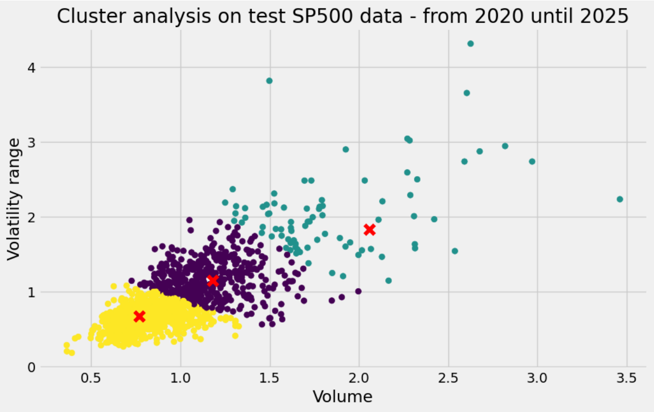
Supervised Machine Learning

Use of labeled data to identify underlying patterns and relationships between input features (X) and outputs (Y), and create a model that can predict correct outputs on new real-world data.



Unsupervised Machine Learning

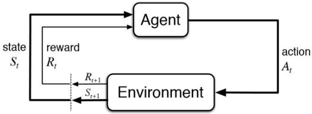
Use of unlabeled data to identify the underlying patterns and structure within the data, without explicit guidance.



Reinforcement Learning

Model learns by interacting with an environment, receiving rewards or penalties based on its actions, as it learns their consequences from taking them.

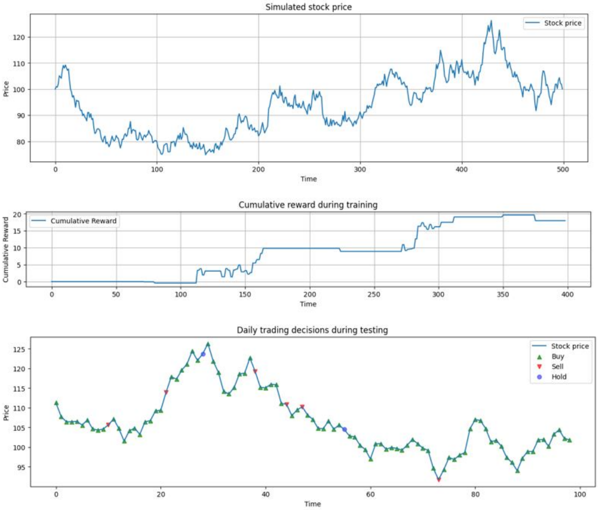
- Game or environment is not explicitly programmed - trial and error learning
- Learns what actions to take so as to maximize reward / minimize penalty.



Balance between exploiting and exploring mandatory to keep learning

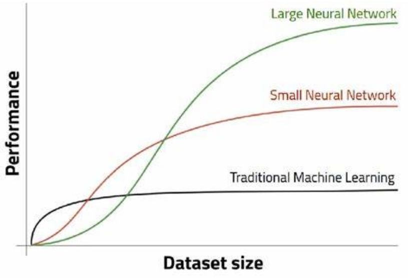
- On one hand, model must take advantage of/exploit everything it has learned in order to make the best choice (eg maximize the reward)
- But it won' t learn anything unless it has done plenty of exploration beforehand.

Reinforcement Learning

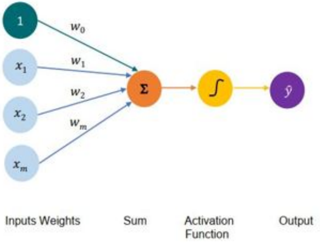


Limitations of Machine Learning

- 1. ML limiting in a multi-dimensional framework (many features)
- 2. ML restrictive as amount of data increases
- 3. ML requires features selection while DL extracts the data and its patterns



Neural Networks



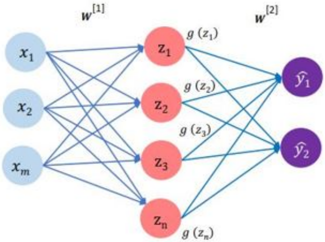
Linear combination of inputs

Output

$$\hat{y} = g\left(w_0 + \sum_{j=1}^{m} x_j w_j\right)$$

Non-linear Activation Function

Bias



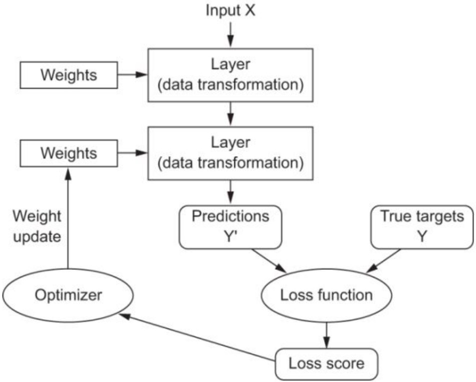
$$z_i = w_{0,i}^{[1]} + \sum_{j=1}^{m} x_j w_{j,i}^{[1]}$$

$$\hat{y}_i = g\left(w_{0,i}^{[2]} + \sum_{j=1}^{n} z_j w_{j,i}^{[2]}\right)$$

Learning process

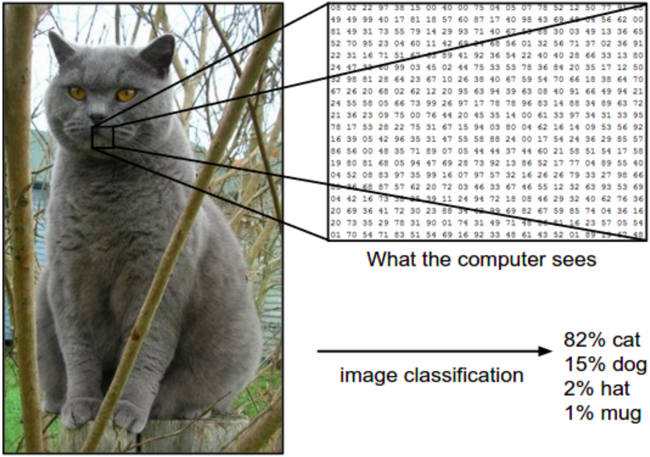
Training a neural network $\Leftrightarrow$ Finding the network weights that achieve the lowest loss

$$W^* = argmin_W \frac{1}{n} \sum_{i=1}^{n} L(f(x^{[i]}; W), y^{[i]})$$



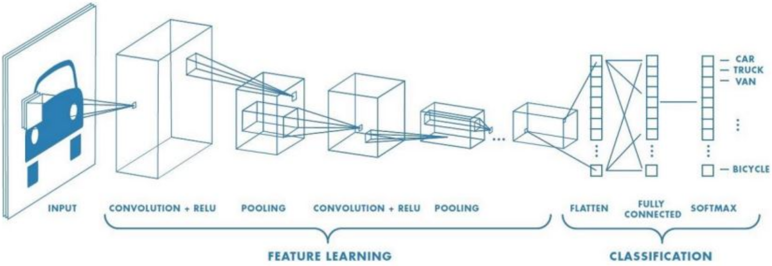
3 types of Neural Networks: 1. CNN

How the computer sees!



3 types of Neural Networks: 1. CNN

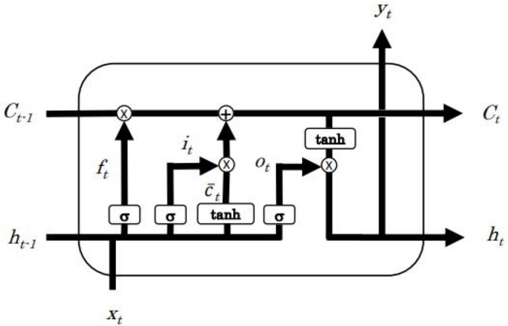
Convolutional Neutral Network



3 types of Neural Networks: 2. RNN / LSTM

Recurrent NN Long-Short Term Memory

- Apply NN to problems involving sequential processing of data / identifying patterns
- Sequence data comes in many forms : text, audio, video and financial time series
- Modeling to predict the next sequence of events (word, sound, time series)



$$f_t = \sigma\,(W_f . [h_{t-1}, x_t] + bf)$$

$$i_t = \sigma\,(W_i\, [h_{t-1}, x_t] + bi)$$

$$\bar{c}_t = tanh\,(W_c\, [h_{t-1}, x_t] + b_c)$$

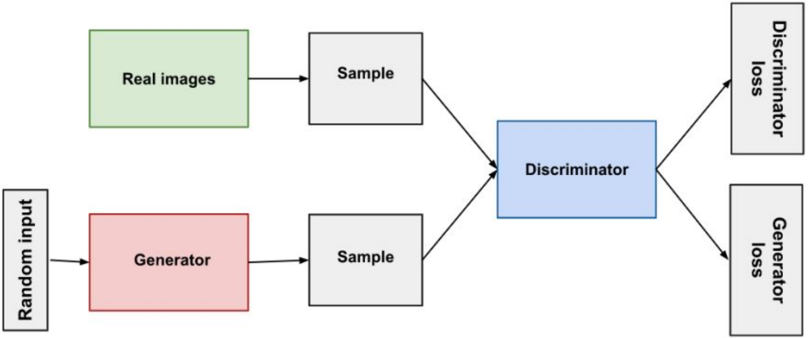
$$C_t = f_t * Ct_{-1} + it * \bar{c}_t$$

$$o_t = \sigma\,(W_o\, [h_{t-1}, x_t] + bo)$$

$$h_t = o_t * tanh\,(C_t)$$

3 types of Neural Networks: 3. GAN

Generative Adversial Networks

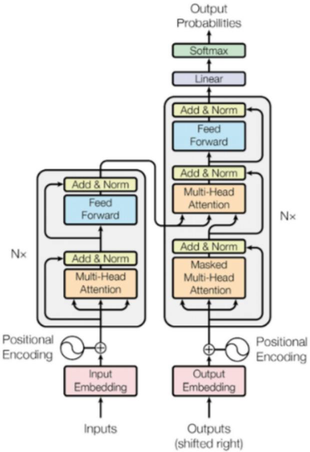


Transformers

LLM output relates to a probability. Each response is different (softmax function)!

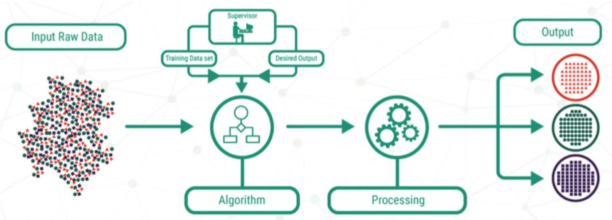
Transformer

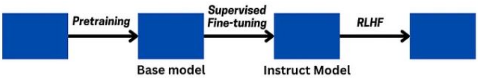
Attention Is All You Need



Large Language Models

- **Pre-training, unsupervised**: model learns on a massive corpus of text data
- **Fine-tuning**: (self) supervised learning to optimize model's responses

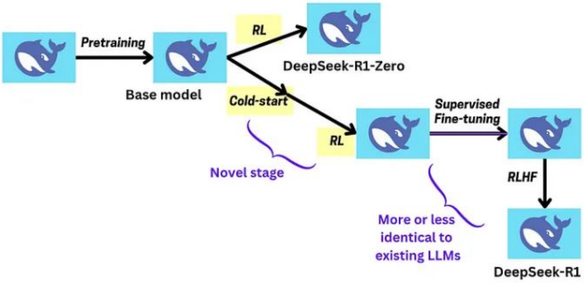




Several limitations: Extensive human-labeled data, time-intensive and costly (collecting + annotating data), limited scalability to new domains, potential biases in training data

DeepSeek breakthrough

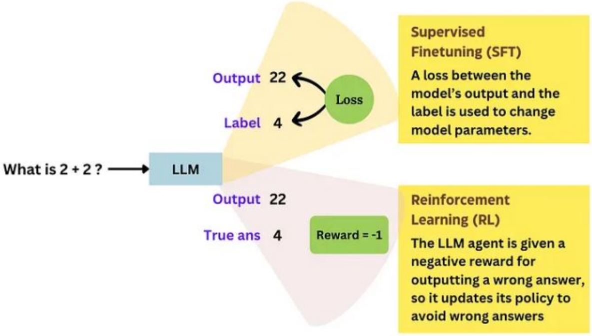
- Learns by trial and error, refining responses without human intervention
- Develops reasoning capabilities naturally instead of being spoon-fed answers, leading to more robust decision-making. Not human but-rule-based reward system, allowing it to develop reasoning capabilities and context



- Self-corrects over time, reducing need for ongoing dataset updates

RL v/s SFT in LLM training

- **Supervised Fine-Tuning**: tends to memorize the training data, model fails to answer questions outside the scope of the specific training data

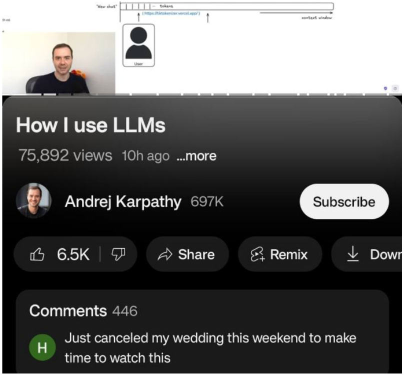


- **Reinforcement Learning**: tends to generate policies that help the agent learn the general concepts rather than memorize the training data

How to use LLM effectively!

2-hour tutorial from Open AI co-founder on how to increase productivity with LLMs:

Choose appropriate LLM, create Task-Specific Custom GPTs, etc



Summary

- **Model & predict** : Learning some useful representation of the input data
- **In front of an AI problem, ask yourself**:
  - Data sources + history?
  - Supervised, unsupervised or reinforcement learning?
  - Training process? Cross-validation, etc
- **Gen AI**:
  - Deep learning architecture to learn complex data representations
  - Traditional LLM struggle to contextualize and generalize
  - Next generation of Gen AI = Artificial General Intelligence = Graal!

References

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013), An Introduction to Statistical Learning
- Francois Chollet (2017), Deep Learning with Python
- Aurelien Geron (2019), Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow
- Kannan Singaravelu (2022), Practical Machine Learning
- Scott Fortmann-Roe (2012) , Understanding the Bias-Variance Tradeoff
- Stefan Jansen (2022), Machine Learning for Algorithmic Trading
- Standford University, Massachusetts Institute of Technology, Technische Universität München, Notes on Artificial Intelligence
- Andrej Karpathy (2015), The Unreasonable Effectiveness of Recurrent Neural Networks