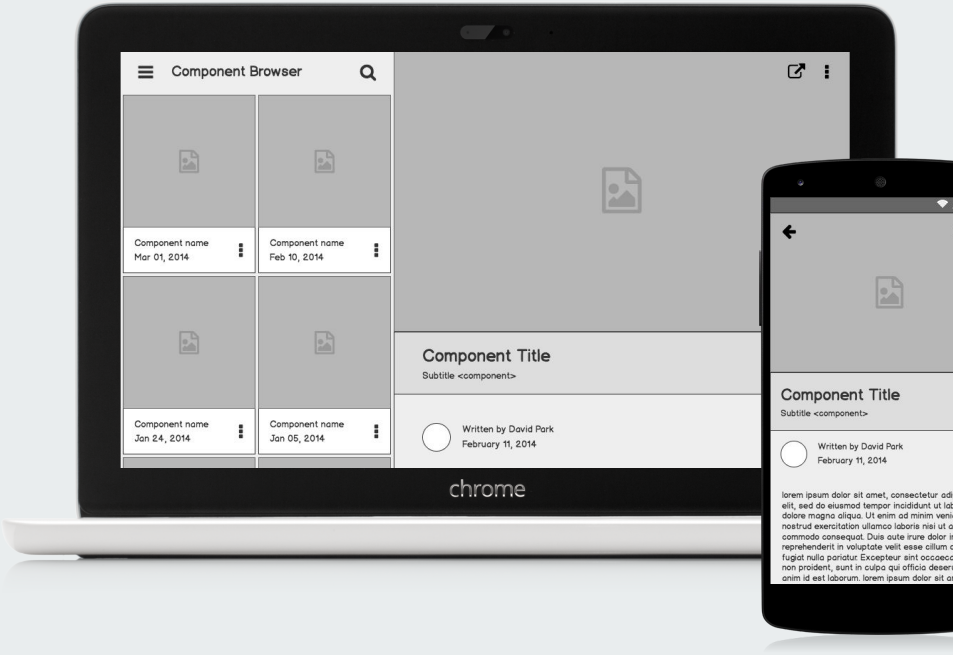
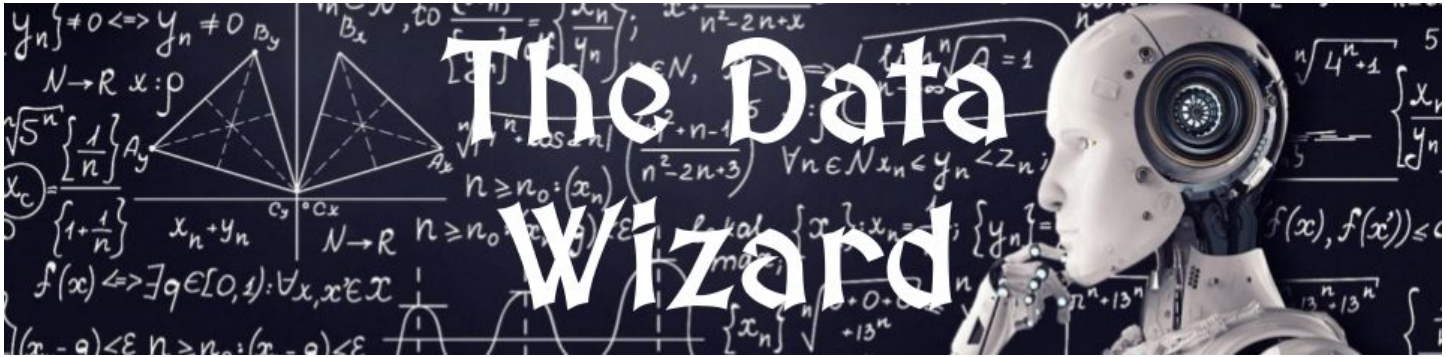
Enterprise Software Engineering Practices: AI/ML Real-World Applications



Lecture by: Brooks Christensen

ABOUT ME



- Master's Degree in Physics from CU Boulder (2020)
- Applied Data Science Certificate from MIT PE Program (2022)
- AI / ML for Business Applications Certificate from UT Austin (2023)
- Nielsen Media Ratings (2021 - 2024)
- Self-employed at Fit Street Global Startup (2024)

Outline

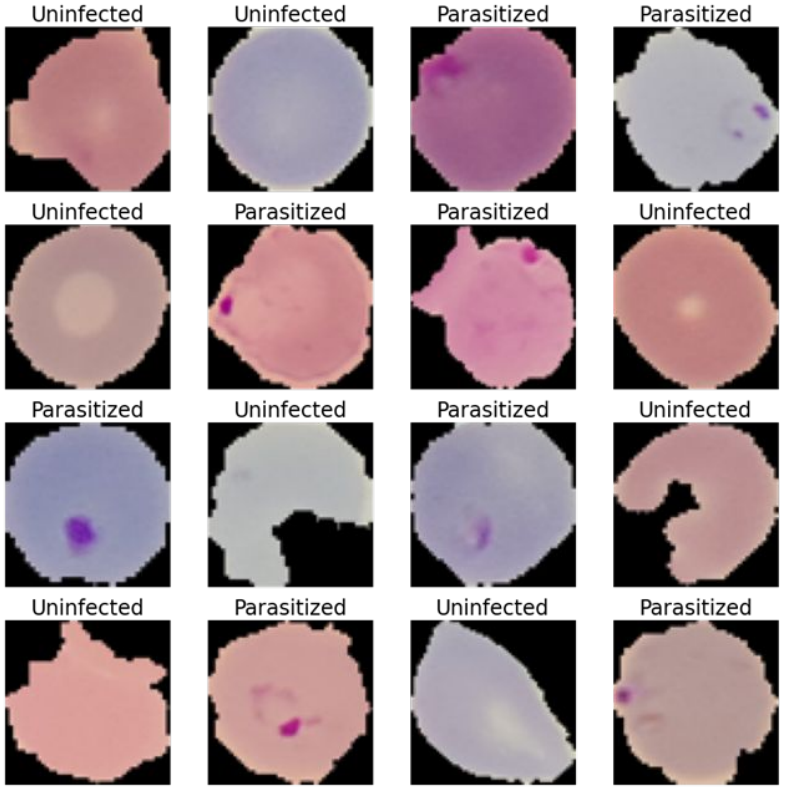
- My Projects
- Intro to AI/ML Applications
- AI/ML Types
- Neural Networks
- Ensembles
- How AI/ML Systems Fail
- Ethical Considerations

“The AI City” - ChatGPT (Oct 1, 2024)

My Projects

Prior Work: AI for Social Good

- Malaria Detection Classifier
  - Uninfected
  - Parasitized
- Blood Sample Images
  - RGB
  - Roughly 150 x 150 Pixels
- CNN Architecture
  - 95%+ Accuracy
- Vision Transformer Architecture

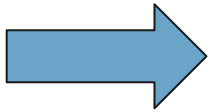


Prior Work: Transition Detection in Media Streams

- Transformer Architecture
- 10,000+ Programs
- Annotated Transitions
- Accuracy Scores: 60% @ 1s, 90%+ @ 15s

Program

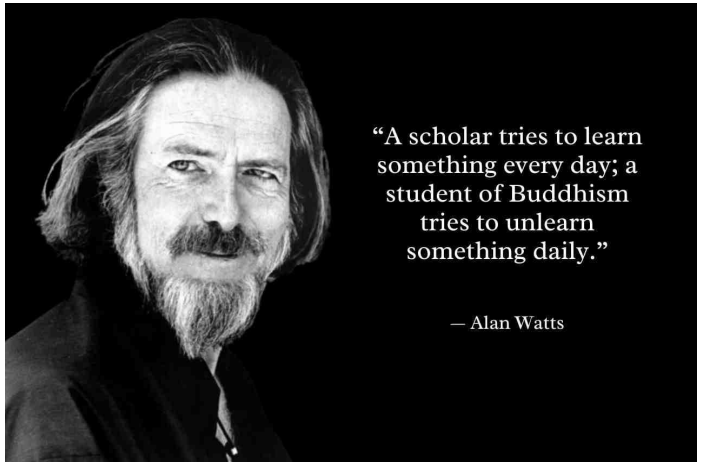
Transition



Advertisement

Current Research: ChatBot Design

- Goal: Speak with Alan Watts
- Three AI Architectures
  - Speech-to-Text (Whisper, OpenAI)
  - Fine-Tuned GPT2 Large Language Model (OpenAI)
  - Text-to-Speech (OpenVoice, MyShell AI)
- Deployment to Production (Future)
  - Kubernetes Cluster
  - Remote Server



Introduction to AI/ML Applications

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, Thomas Brox

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we pr
expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outp
(phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fa

Extreme-Long-short Term Memory for Time-series Prediction

Sida Xing, Feihu Han, Suiyang Khoo

The emergence of Long Short-Term Memory (LSTM) solves the problems of vanishing gradient and exploding gradient in traditional Recurrent N
of LSTM can help to improve the efficiency in other application areas.
In this paper, we proposed an advanced LSTM algorithm, the Extreme Long Short-Term Memory (E-LSTM), which adds the inverse matrix part c
training rounds, thus reducing the overall training time.
In this research, the E-LSTM model is used for the text prediction task. Experimental results showed that the E-LSTM sometimes takes longer to
traditional LSTM, whilst also improving the training speed and the overall efficiency of the LSTM.

Deep Residual Learning for Image Recognition

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially dee
showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we e
place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers.
The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 2
detection, ImageNet localization, COCO detection, and COCO segmentation.

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llior

The dominant sequence transduction models are based on complex recu
recurrence and convolutions entirely. Experiments on two machine translation tas
the WMT 2014 English-to-French translation task, our model establishes a new single-m
limited training data.

Auto-Encoding Variational Bayes

Diederik P Kingma, Max Welling

How can we perform efficient inference and learning in directed probabilistic model
case. Our contributions are two-fold. First, we show that a reparameterization of th
an approximate inference model (also called a recognition model) to the intractable

Generative Adversarial Networks

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley,

We propose a new framework for estimating generative models via an adversarial process, in wh
making a mistake. This framework corresponds to a minimax two-player game. In the space of ar
Markov chains or unrolled approximate inference networks during either training or generation of

Fashion-MNIST: a Novel Image Dataset for Benchmarking Mach

Han Xiao, Kashif Rasul, Roland Vollgraf

We present Fashion-MNIST, a new dataset comprising of 28x28 grayscale images of 70,000 fashion products from
algorithms, as it shares the same image size, data format and the structure of training and testing splits. The dataset

Recurrent Neural Networks (RNNs):

Robin M. Schmidt

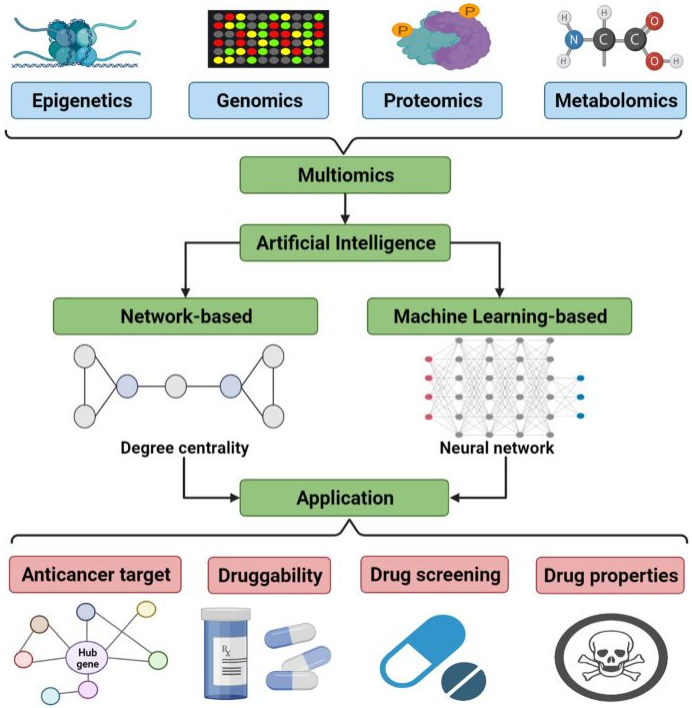
State-of-the-art solutions in the areas of "Language Modelling & G
publications in those areas. In this work we give a short overview
Mechanism" or "Pointer Networks". We also give recommendatior

WaveNet: A Generative Model for Raw Audio

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Ale

This paper introduces WaveNet, a deep neural network for generating raw audio waveforms. The n
speech, it yields state-of-the-art performance, with human listeners rating it as significantly more na
trained to model music, we find that it generates novel and often highly realistic musical fragments.

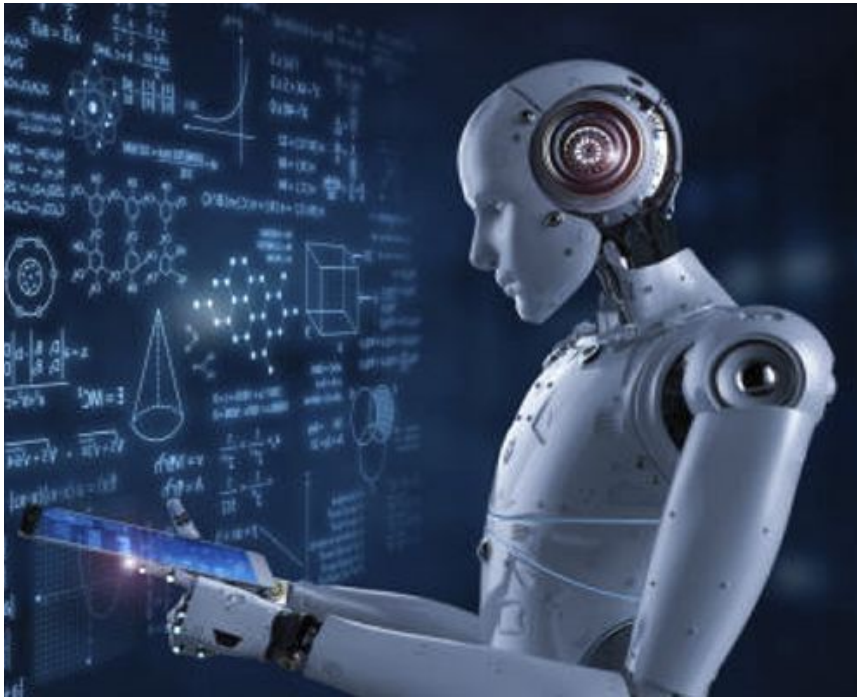
ML Applications



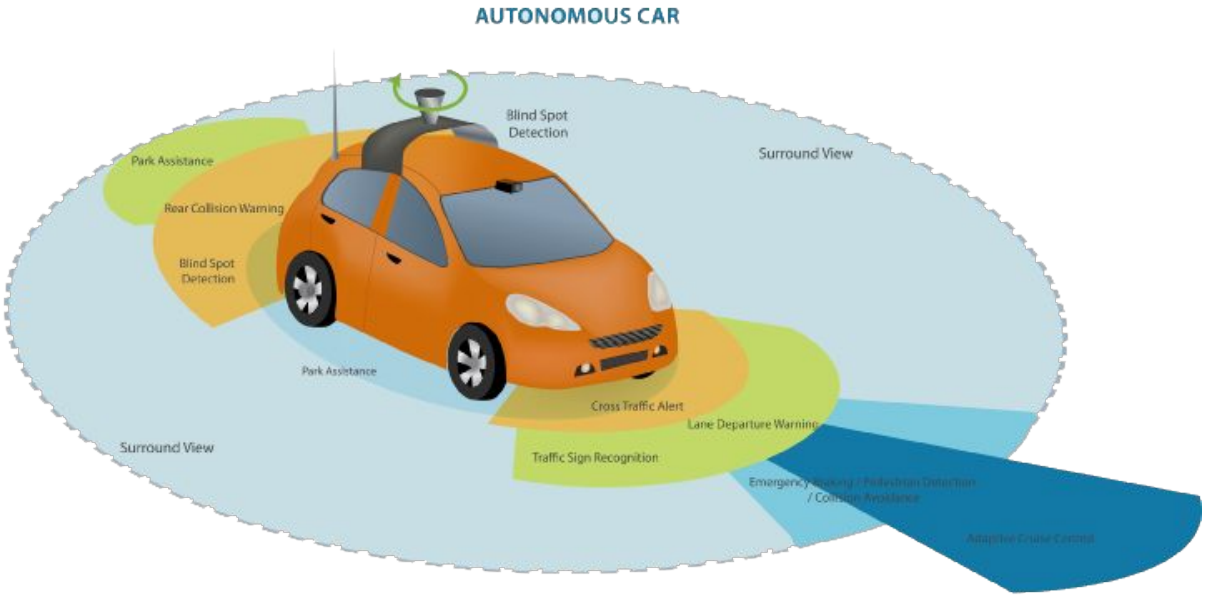
ML Applications

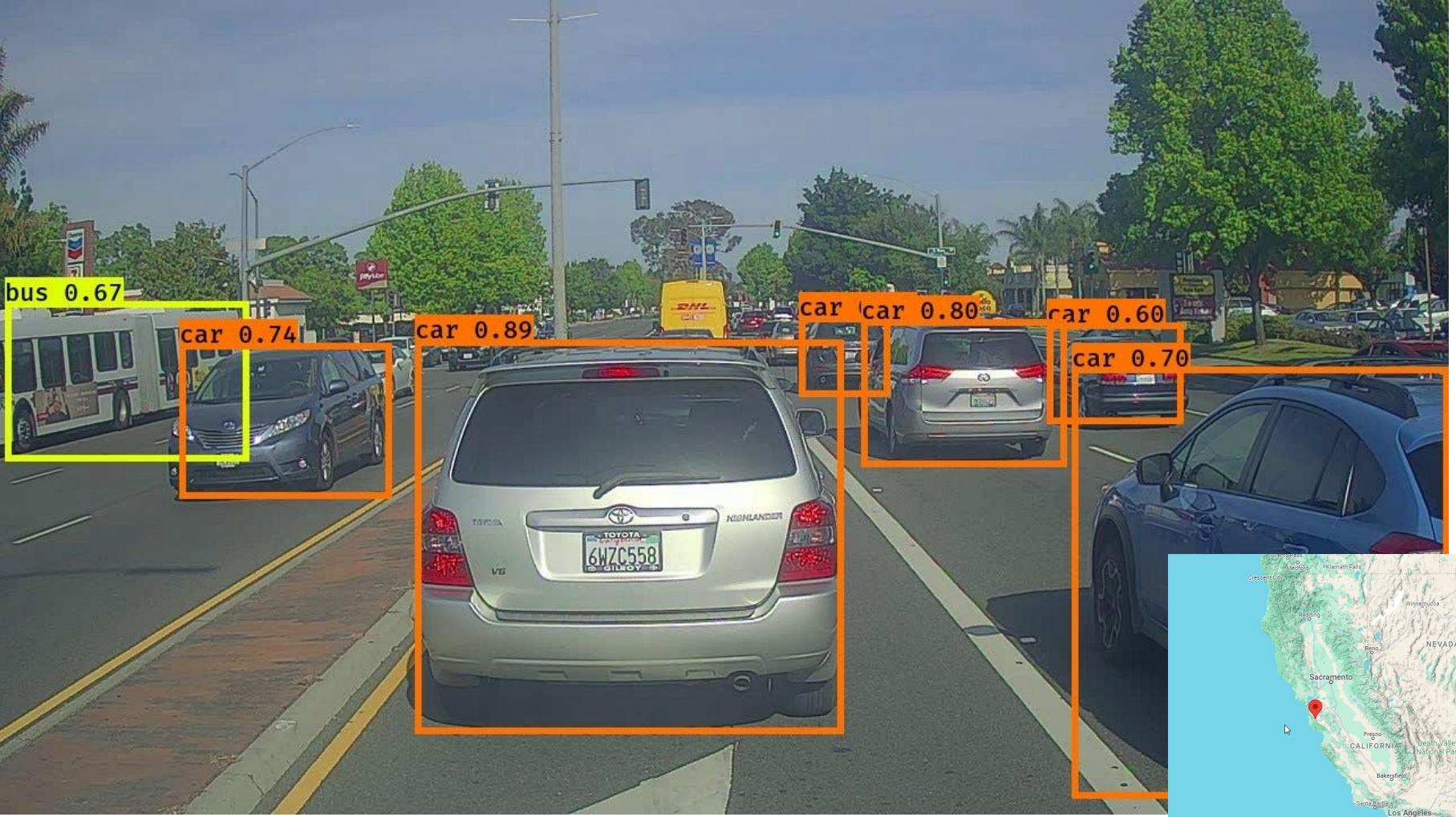
ML Applications

- Large Language Models (ChatGPT, etc)
  - Machine Translation
  - Chatbots
  - Text Summarization
  - Sentiment Analysis
  - Content Creation
  - Knowledge Extraction and Question Answering
  - Personalization
  - Research & Development
  - Retrieval Augmented Generation (RAG)
- Traffic Management
- Food Production
- Environmental Modeling and Sustainability
- Creativity and Art
- Quantum Machine Learning



ML Applications

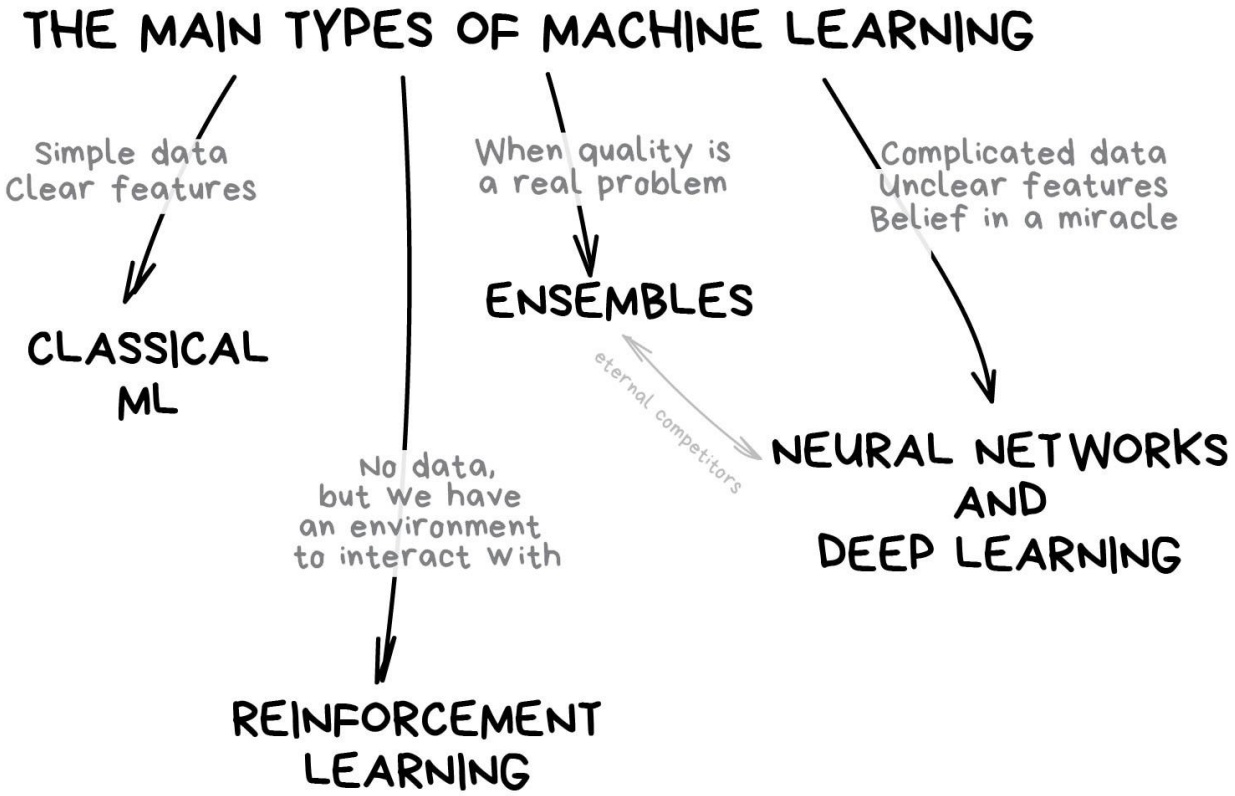


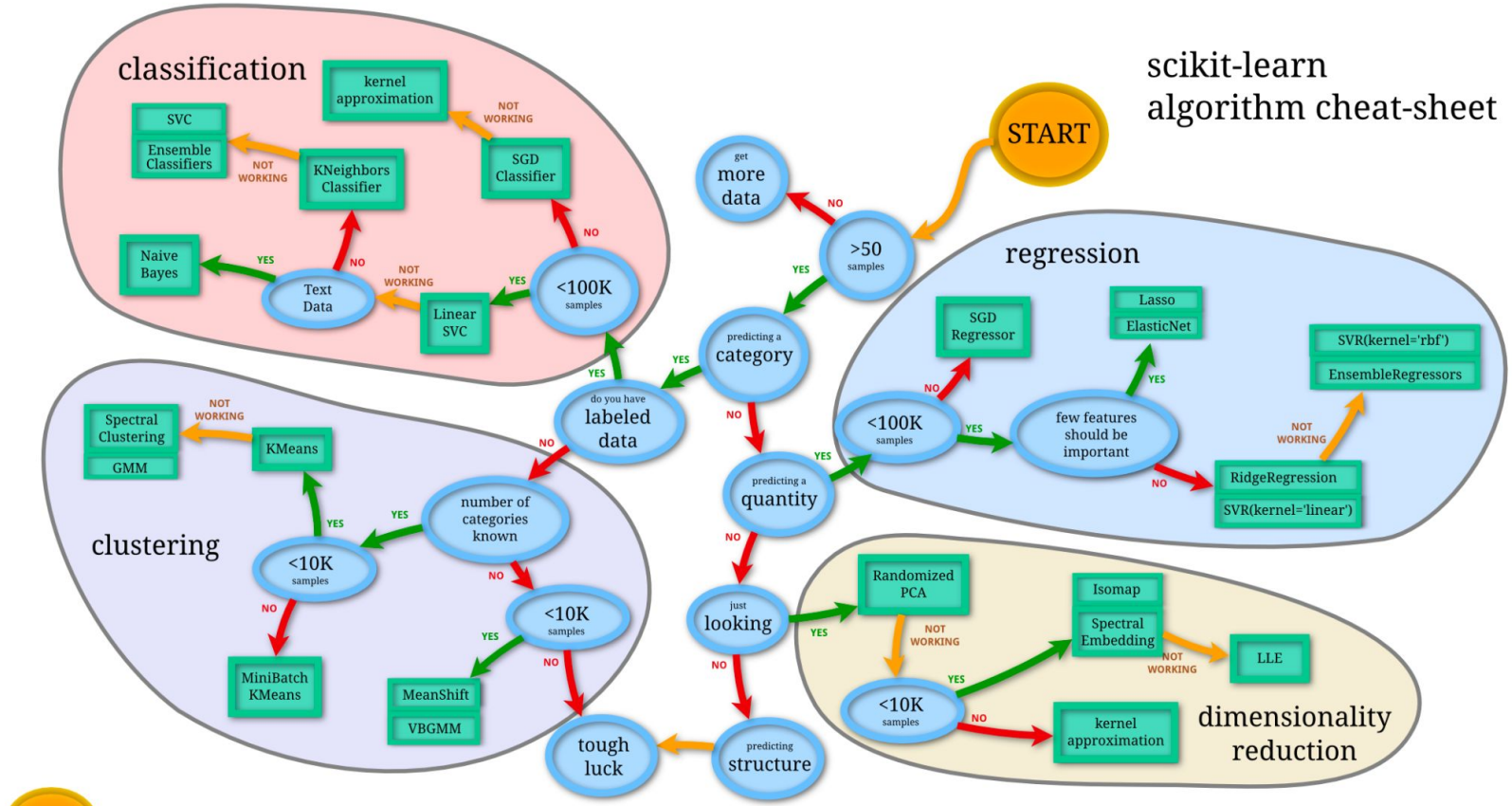

bus 0.67
car 0.74
car 0.89
car
car 0.80
car 0.60
car 0.70
TOYOTA
6WZC558
HIGHLANDER

NEVADA
Sacramento
CALIFORNIA
Las Vegas
Los Angeles
San Diego

AI/ML Types

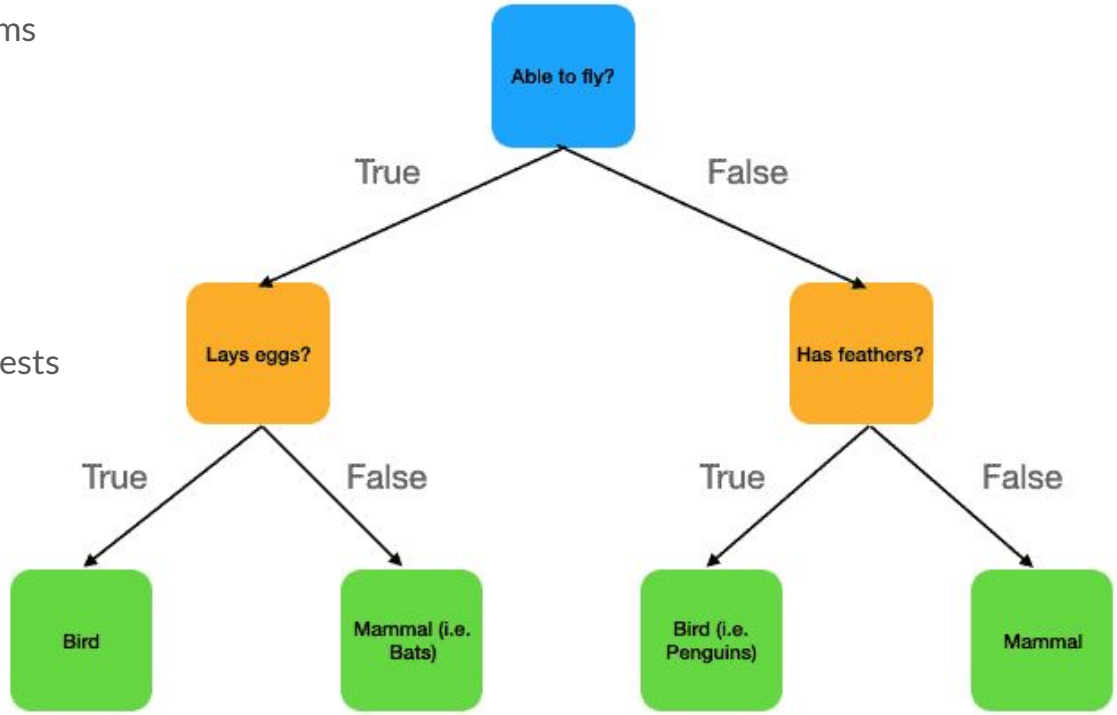
Main Types of ML



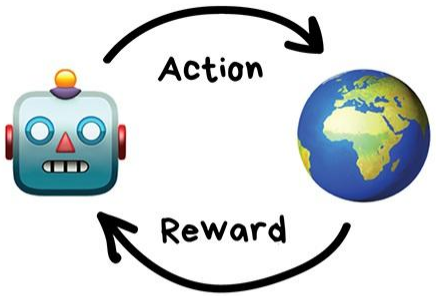

scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
classification
kernel approximation
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
KNeighbors Classifier
SGD Classifier
NO
Naive Bayes
YES
Text Data
NO
NOT WORKING
Linear SVC
YES
<100K samples
NO
predicting a category
YES
do you have labeled data
YES
NO
regression
SGD Regressor
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
NO
<100K samples
YES
few features should be important
YES
NOT WORKING
NO
RidgeRegression
SVR(kernel='linear')
predicting a quantity
YES
NO
clustering
Spectral Clustering
GMM
NOT WORKING
KMeans
YES
number of categories known
YES
NO
<10K samples
NO
MiniBatch KMeans
<10K samples
YES
NO
MeanShift
VBGMM
just looking
YES
NO
Randomized PCA
NOT WORKING
Isomap
Spectral Embedding
NOT WORKING
LLE
YES
<10K samples
NO
kernel approximation
dimensionality reduction
tough luck
predicting structure
Back
scikit learn

Decision Trees

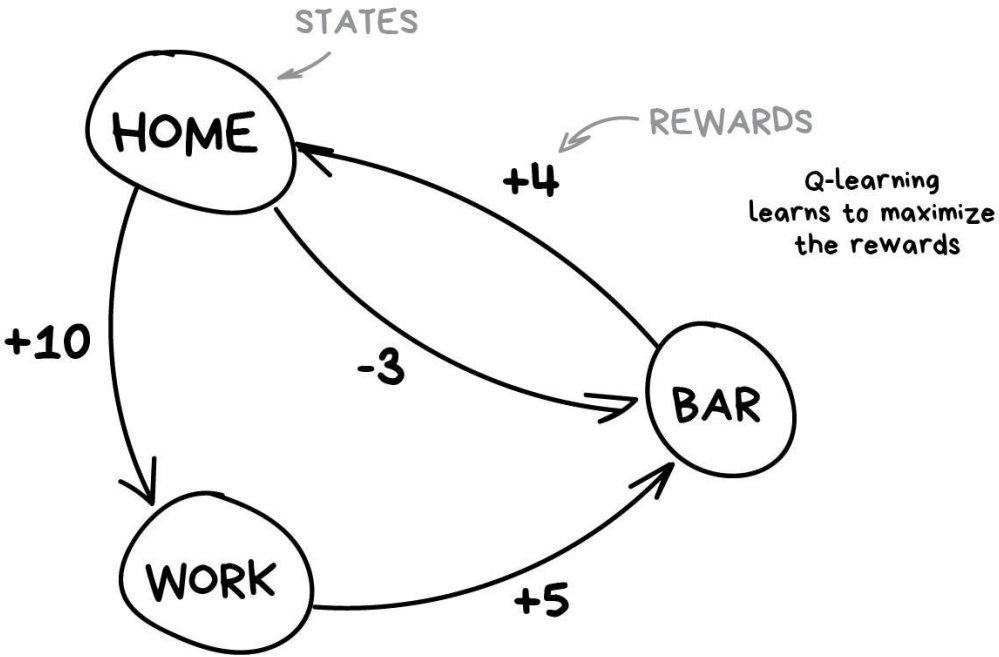
- Classical ML Technique
- Practical Solutions for many Problems
- Useful for Tabular Data
- Not a Neural Network
- Classification or Regression
- Typically Better Performance in Forests



Reinforcement Learning



Reinforcement Learning



ROUTINE MARKOV PROCESS

MICROSOFT / WEB / TL;DR

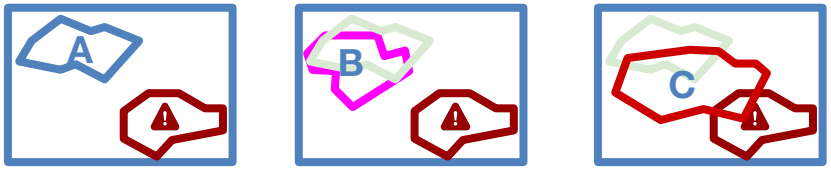
Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

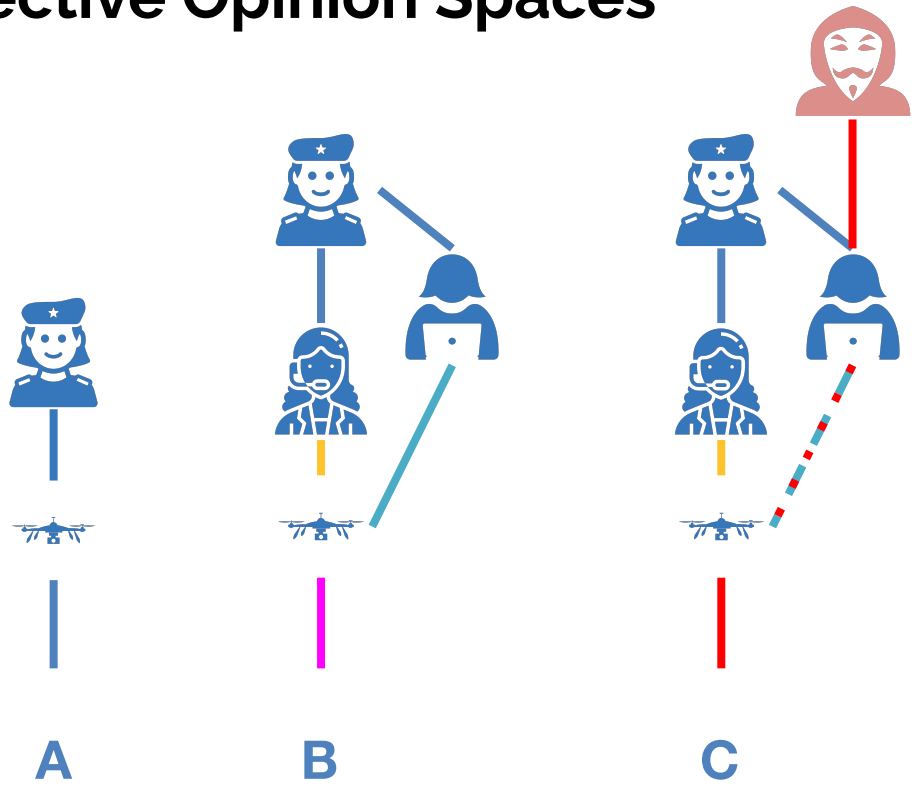
AI Loyalty: Aligning on Subjective Opinion Spaces

Intent is complicated.

The operator, even if faithful, introduces elements of their own intent, even if simply executing the Commander's Intent to the best of their ability. This is because they operate within limits of homotopy and homology of their shared common ground.

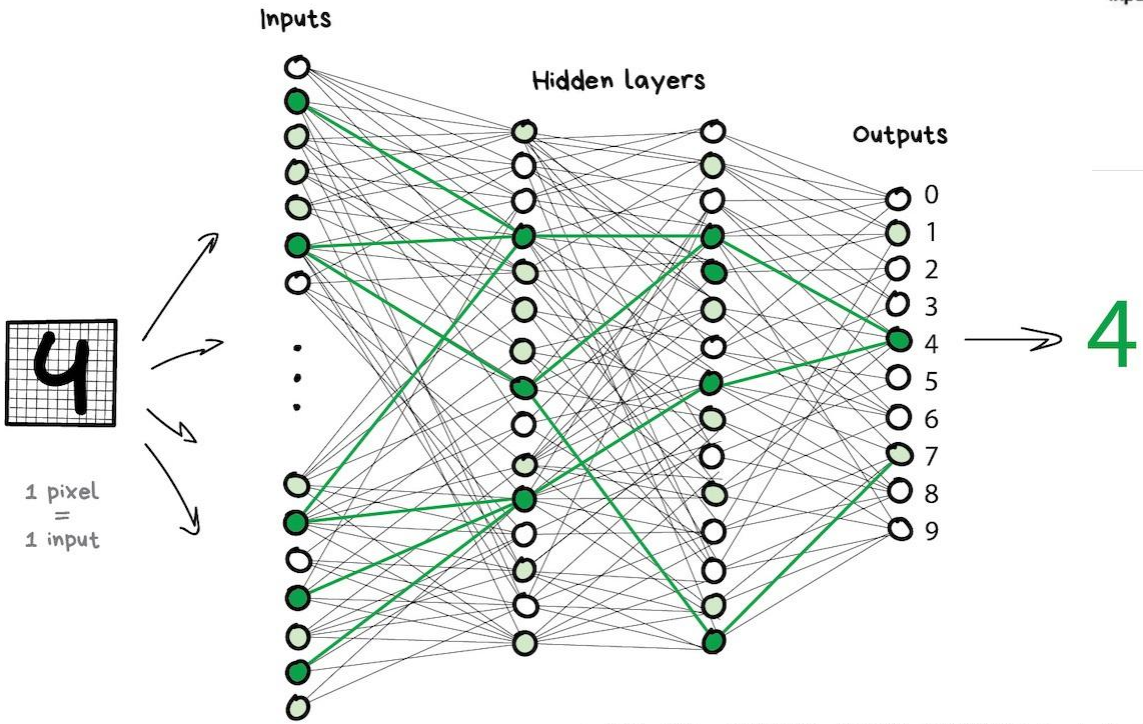
Adversaries need not be directly involved in operation. They can poison data sets, libraries, and generate other upstream threats to an otherwise trustworthy system, introducing critical errors in intention.

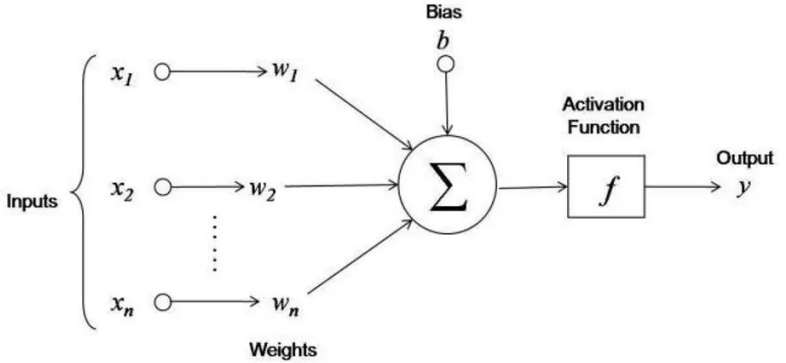




Neural Networks

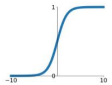
Neural Networks - MLP



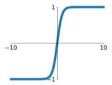


Activation Functions

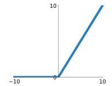
Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$



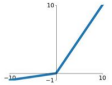
tanh
$\tanh(x)$



ReLU
$\max(0, x)$

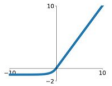


Leaky ReLU
$\max(0.1x, x)$



Maxout
$\max(w_1^T x + b_1, w_2^T x + b_2)$

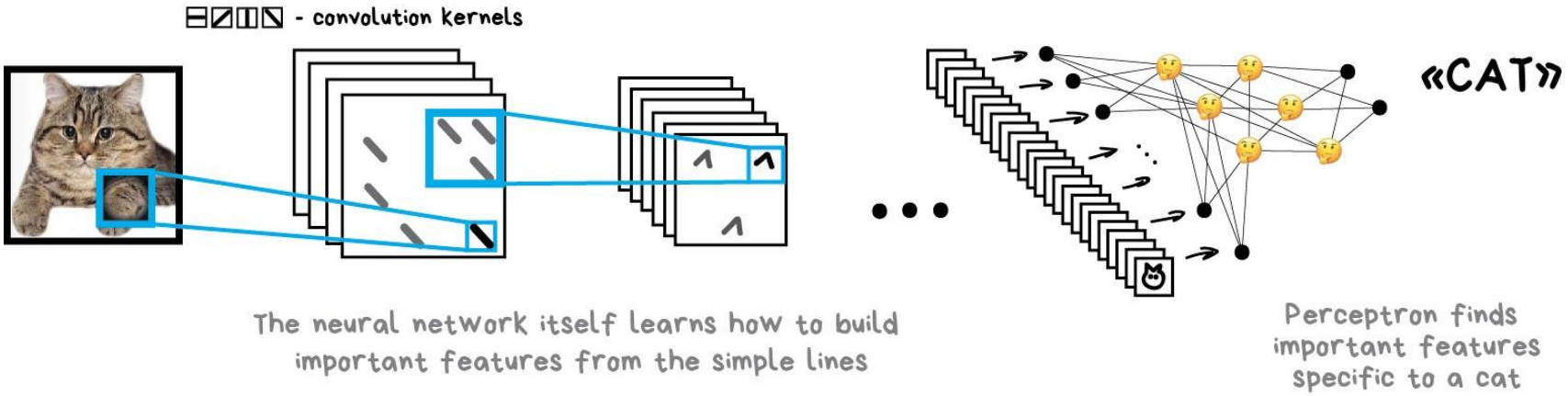
ELU
$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



Cross-Entropy Loss Function

$$H(P^* | P) = -\sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTIRBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTIRBUTION}}$$

Neural Networks - CNN

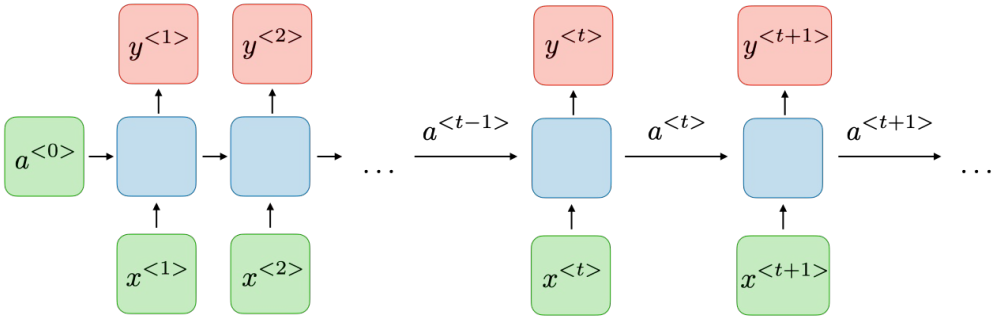


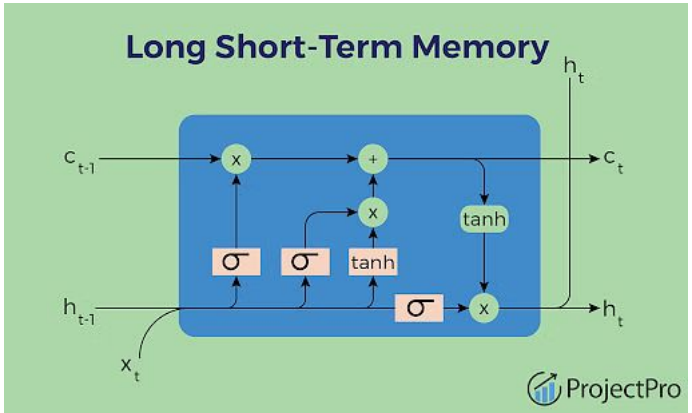
CONVOLUTIONAL NEURAL NETWORK (CNN)

Neural Networks - RNN

- Useful for Sequential Input Processing
- Single Hidden Unit
- Many Architecture Variants
- Activation Function Updated each Iteration
- Vanishing Gradient Problem
- LSTM uses Gating to Address VGP

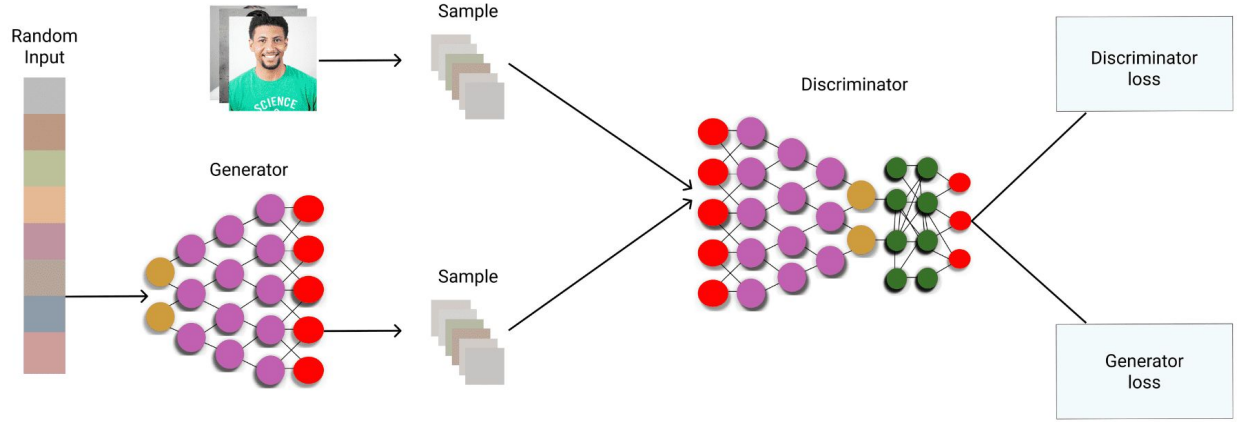
Traditional RNN





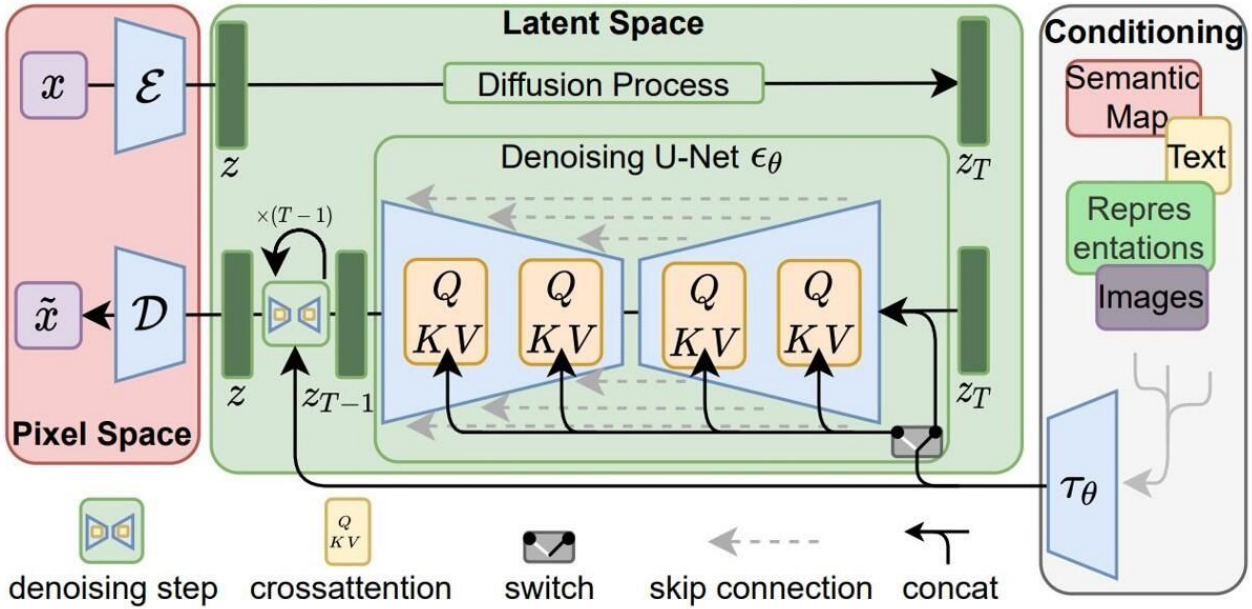
Neural Networks - GAN

- Discriminator and Generator
- Student/Teacher Relationship
- Stability Issues



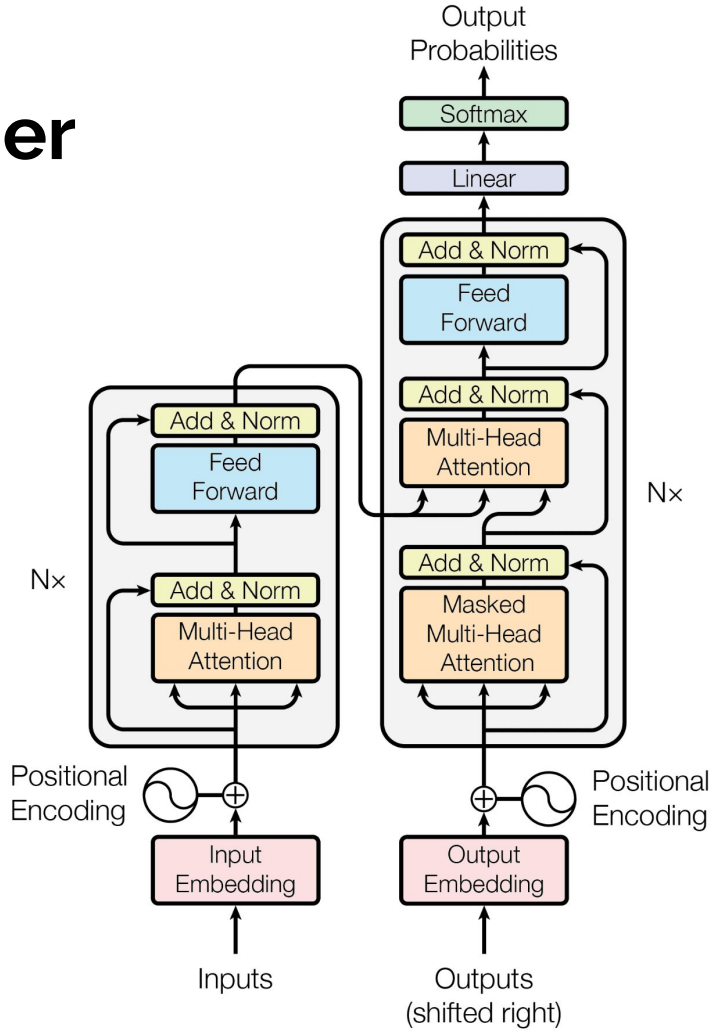
Neural Networks - Stable Diffusion

- More Stable than GAN
- Most Current Image Generation Systems use SD-type Algorithms



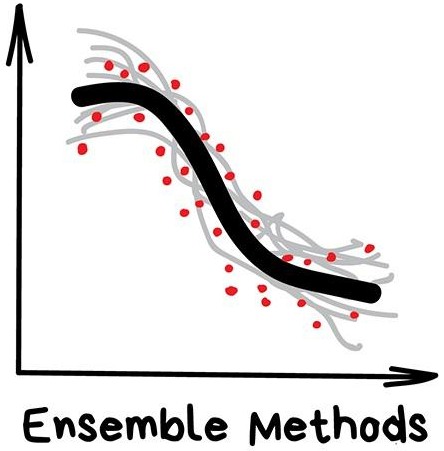
Neural Networks - Transformer

- “Attention is All You Need”, Vaswani, Et. al (2017)
- GPT Architecture
- Key, Query, Value
- Encoder/Decoder
- Positional Encoding
- Masking
- Multimodal

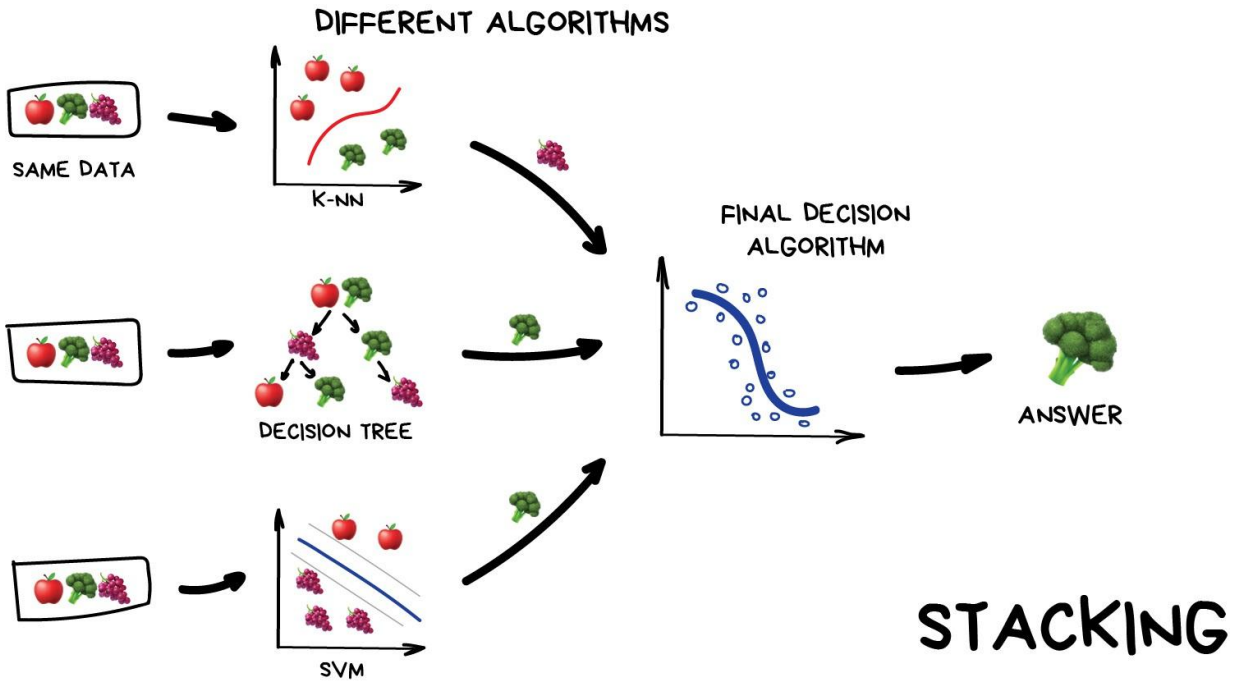


Ensembles

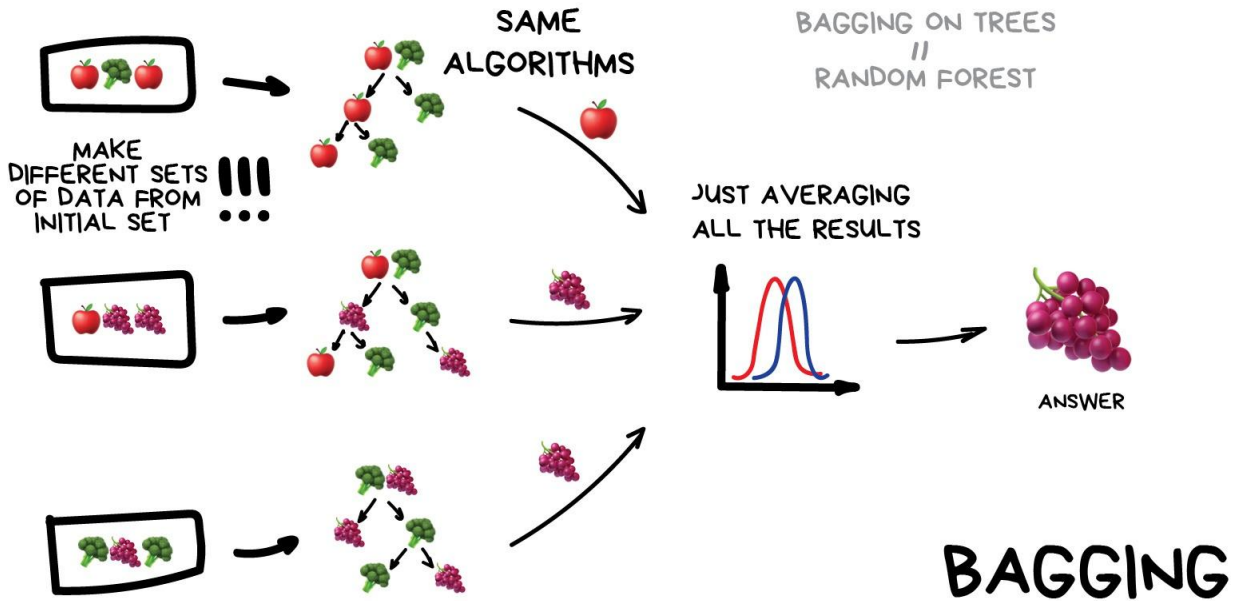
Ensembles



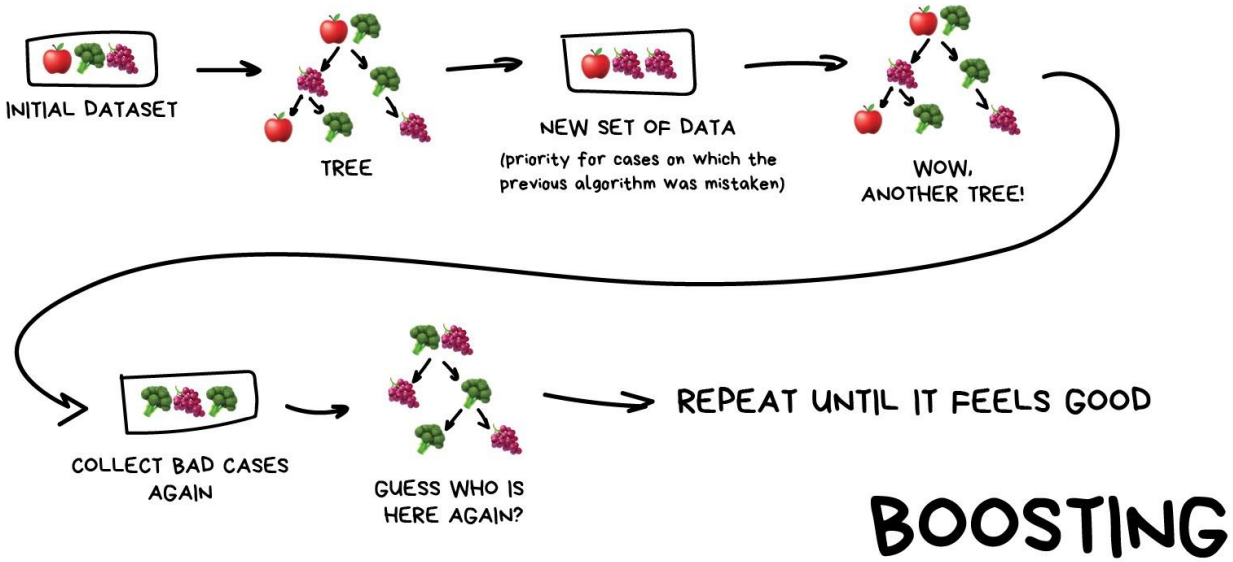
Ensembles


DIFFERENT ALGORITHMS
SAME DATA
K-NN
DECISION TREE
SVM
FINAL DECISION ALGORITHM
ANSWER
STACKING

Ensembles


SAME
ALGORITHMS
BAGGING ON TREES
=
RANDOM FOREST
MAKE
DIFFERENT SETS
OF DATA FROM
INITIAL SET
!!!
JUST AVERAGING
ALL THE RESULTS
ANSWER
BAGGING

Ensembles


INITIAL DATASET
TREE
NEW SET OF DATA
(priority for cases on which the previous algorithm was mistaken)
WOW, ANOTHER TREE!
REPEAT UNTIL IT FEELS GOOD
COLLECT BAD CASES AGAIN
GUESS WHO IS HERE AGAIN?
BOOSTING

How AI/ML Systems Fail

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

What went wrong

Many of the problems that were uncovered are linked to the poor quality of the data that researchers used to develop their tools.

https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/

Traditional Software System Failures

- **Dependency failure –** A third party package no longer maintained
- **Hardware failure –** Hard-drive failure, CPU overheating, losing network access, etc
- **Downtime or Crashing –** Backend infrastructure outage

TECH

Dead Roombas, stranded packages and delayed exams: How the AWS outage wreaked havoc across the U.S.

PUBLISHED THU, DEC 9 2021·8:00 AM EST | UPDATED THU, DEC 9 2021·10:51 AM EST

https://www.cnbc.com/2021/12/09/how-the-aws-outage-wreaked-havoc-across-the-us.html

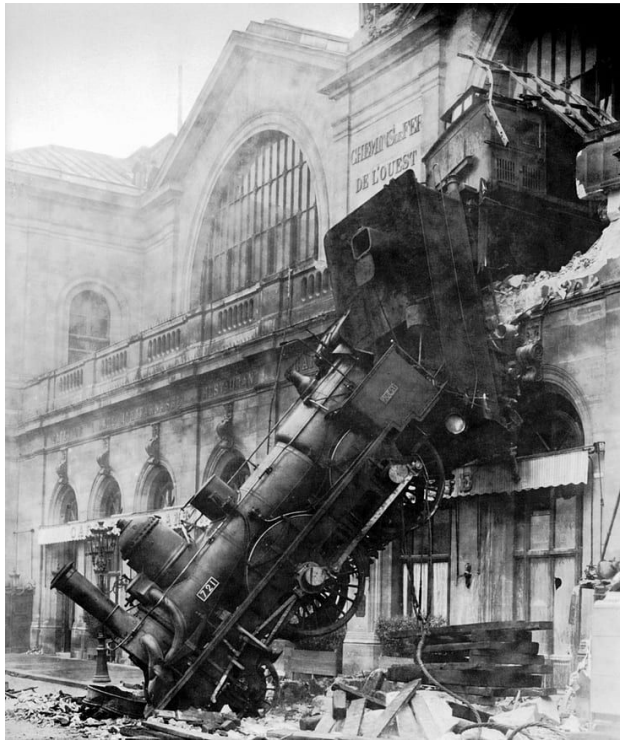
Raw Data Issues

- Sampling Errors
  - Convenience
  - Snowball
  - Judgement
  - Quota
- Imputation
- Data Leakage
- Duplicate Values
- Distribution Shifts

ML-Specific Failures

They account for a smaller portion of failures but are **harder to detect and fix**:

- **Edge cases**
- **Data Distribution Shifts**
- **Degenerate feedback loops**



ML-Specific Failures – Edge Cases

ML-Specific Failures – Data Distribution Shifts

Consider a model estimating the likelihood of pets being adopted in a shelter. Let's call X the characteristics of a pet, Y whether its adopted or not, and P(Y|X) the likelihood of a pet being adopted

- **Covariate shift –** P(X) changes, P(Y|X) remains

A shelter takes in more older pets, but the likelihood of adoption remains the same by age.

- **Label shift –** P(Y) changes, P(X|Y) remains

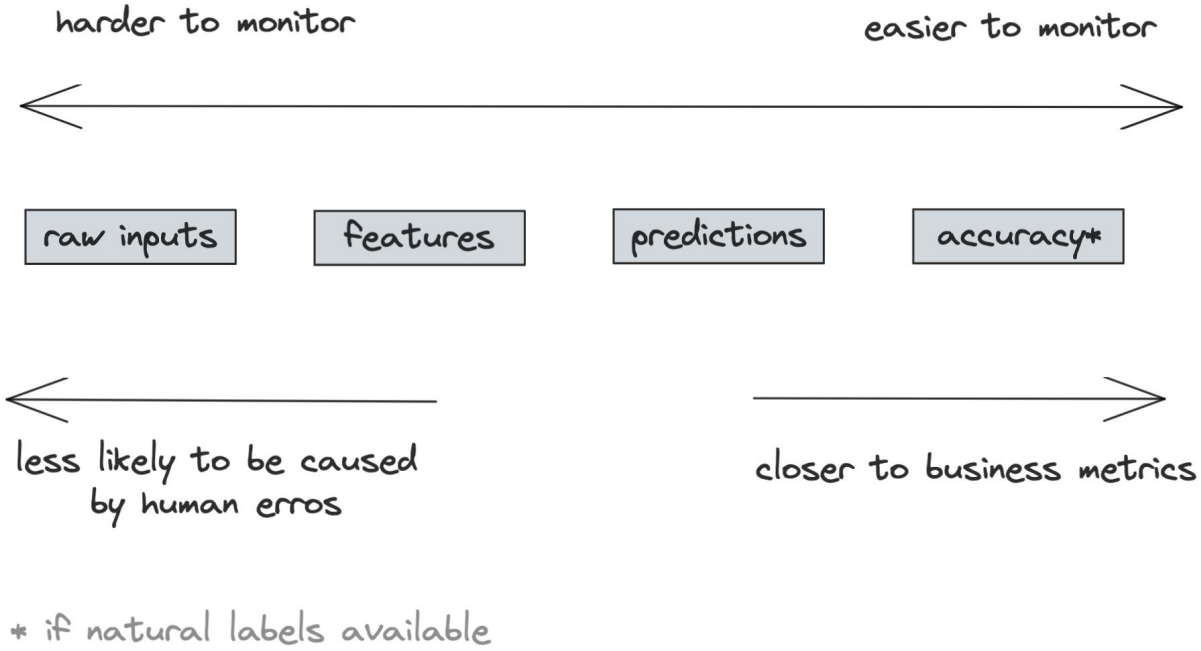
People want to adopt more pets, regardless of age.

- **Concept drift –** P(Y|X) changes, P(X) remains

Society decides it wants to adopt more older pets, but the age of pets in the shelter remains the same.

ML-Specific Failures – Detect Data Distribution Shifts

Monitor data all along the inference pipeline over time

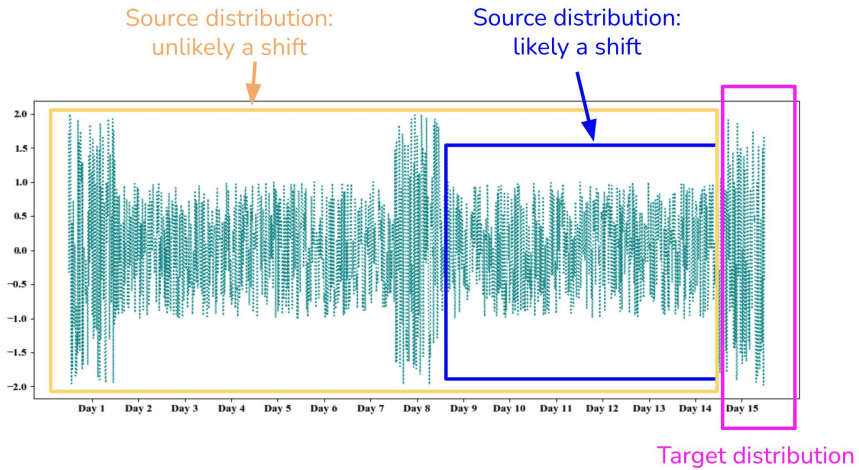


https://huyenchip.com/2022/02/07/data-distribution-shifts-and-monitoring.html

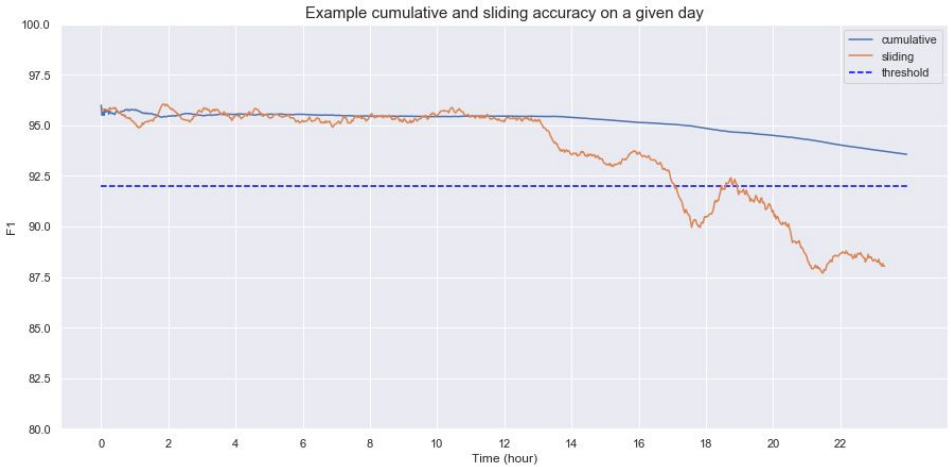
ML-Specific Failures – Detect Data Distribution Shifts

A couple of monitoring pitfalls:

Consider seasonal variations



Cumulative statistics hide sudden shifts



https://huyenchip.com/2022/02/07/data-distribution-shifts-and-monitoring.html

ML-Specific Failures – Addressing Data Distribution Shifts

- **Prevent shifting by initially training on massive dataset**

  Rarely possible in the industry due to cost and time constraints.

- **Adapt a trained model to a new target distribution without new data**

  Very complex and not widely adopted due to lack of research.

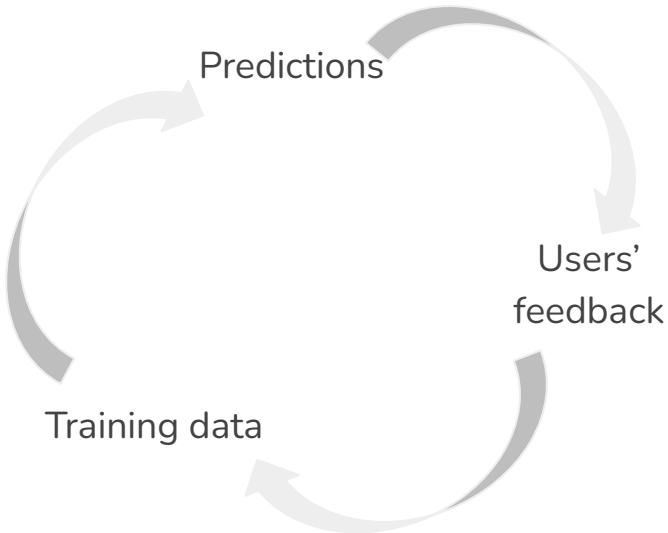
- **Retrain the model with newer data**

  Usually done in the industry due to an easier access to data thanks to the existing inference pipeline; it’s easier to correct predictions than annotate from scratch.

  But watch-out for degenerate feedback loops!

ML-Specific Failures – Degenerate Feedback Loops

- Predictions influence the feedback, which is then used to extract labels to train the next iteration of the model

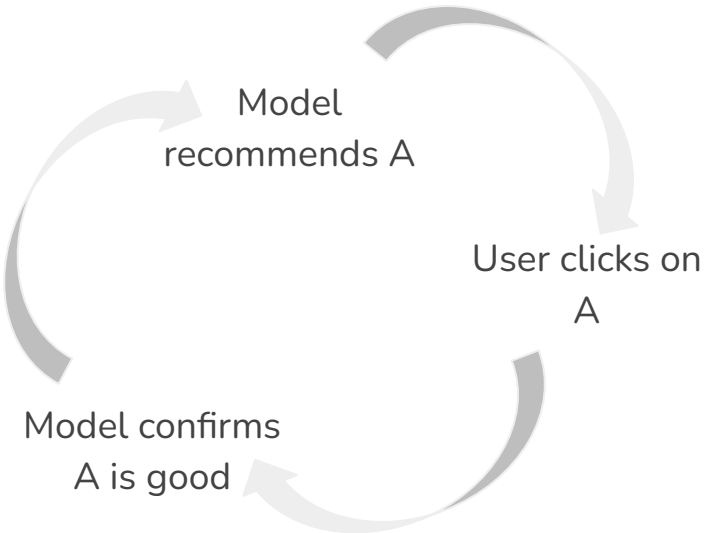


Only arise once models are in production: hard to detect during development!

https://docs.google.com/presentation/d/1tuCIbk9Pye-RK1xqiiZXPzT8IlgDUL6CqBkFSYZXkbY/edit

ML-Specific Failures – Degenerate Feedback Loops

- Originally, A is ranked marginally higher than B: model recommends A;
- After a while, A is ranked much higher than B



Only arise once models are in production: hard to detect during development!

https://docs.google.com/presentation/d/1tuCIbk9Pye-RK1xqiiZXPzT8IlgDUL6CqBkFSYZXkbY/edit

Ethics in AI

General Ethical Considerations

- Bias and Fairness
- Transparency and Explainability
- Privacy and Data Security
- Accountability
- Autonomy and Control
- Job Displacement and Economic Impact
- Weaponization of AI
- Environmental Impact

Case Study 1: Automated Grader’s Biases

- Summer 2020 - Pandemic
- UK A-level exams
- Failures:
  - Setting the Wrong Objective
  - Insufficient Evaluation
  - Lack of Transparency

Case Study 2: The Danger of “Anonymized” Data

- Strava Fitness App (2018)
- PII was “Anonymized”
- Identified US Military Patrol Routes
- “opt-out” vs “opt-in” Confusion
- Further Potential for PII Misuse

Acknowledgement

- Vaswani, A. "Attention is all you need." *Advances in Neural Information Processing Systems* (2017).
- Huyen, Chip. *Designing machine learning systems*. "O'Reilly Media, Inc.", 2022.
- ChatGPT

"AI Dreams" - ChatGPT (Oct 1, 2024)