13th August '24

$$\text{Regulation} \propto \frac{1}{\text{innovation}} \times$$

$$\propto \text{innovation} \times$$

- IT companies not considered 'companies' in India. They are considered as shops.

- curious case of 'Google ethics council'
  - Realized in USA, that too much power in hands of few big companies.
  - IT companies should be strictly regulated.
  - Google ethics council ⇒ checks the scope of product in the ethics realm.
  - Council produced a paper "Stochastic Parrot" (Readings)
  - The entire council was fired.
  - NO company can be self-regulated; however companies aim to increase shareholder value.
    ⇓
    either innovate or exploit.

- Use of ethics language to prevent ethical & regulation. is called ethics washing.

Technosolutionism and reification

- 19th century victorian Britain
- streets full of garbage
- we go there using a time machine
- what do we observe?
  - women not equal; considered as property
  - we feel the need to liberate the people there
- another time portal opens; in comes Elon Musk.
- Introduces Pressure cooker; claims to solve the problem and goes.
  - Doesn't solve; women are still expected to cook
  - Just makes exploitation efficient
- Any societal problem cannot be made by making the existing structural issue faster.
- Power doesn't change society; efficiency not a solution

Eg Less Doctors in India; Technosolutionism says make an online app to increase connectivity.

- Reification
  - suppose, this a cooking course.
  - Easiest way to teach for instructor; Teach ~~best~~ the most popular dishes.
  - Final exam based on cooking those dishes.
    ~~Does~~ Is the topper the best cook in India?
  - Problems
    - By choosing these dishes, we make a false reality of cooking.

reification

- widespread in Machine Learning.

AI as scapegrat, or the machine never "understand" or "does" policy

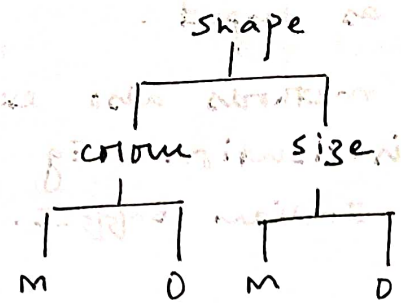
- copy striking in Youtube done using ML.
- may make some error.
- AI by itself not creating problem. It is a tool. People with the power to deploy them are the cause.

Machine Learning

- curse of Dimensionality → More are the number of dimensions, more complex does the underlying math and computation becomes.
- Cross validation  Training Valid - Take every part as validation data.
- Active Learning → use old model to filter out the new incoming data; New images with high error need to be used.
- Oracle generates synthetic testing use-cases.
- Explainability → Why is model giving the output it is giving. Decision Trees ✓ Neural Net ✗

16th August '24

- supervised ML → classification (Discrete values)
  → Regression (Continous values)
- Annotations? → manually (Using Annotators)
  - Need annotator agreement; may not be sufficient but necessary for Data quality
- Bias variance Trade off → underfitting vs Overfitting

- Decision Trees
  - Mango & oranges example
  - 

  Go through features in order of their usefulness

- Bayesian classifier
  - Uses Probability Theory
  - simplest model assumes that features don't affect each other. (Naive-Bayes classifier

- K-Nearest Neighbours
  - Algorithm works by finding k nearest neighbours to a given data point (based on some distance metric)
  - The class of the data point is then determined by the majority vote of the k-neighbours.

- The above 3 methods assume the linearity of features with respect to the labels.

- For Non-linear Relations, the following can be used
  - Neural Networks
  - Support Vector Machines (SVMs)
    - Uses the concept of Kernels
    - less complex than NN
    - can only do binary classification (can be modified to support multiple classification problems)

Unsupervised ML (No annotation)

- clustering (Grouping the data)
  - Needs number of groups as input.
  - Non-parametric clustering methods also exist which use Bayesian distribution. Eg Chinese restaurants and Indian Buffet.

Topic Modelling
- Eg you are given 100,000 digitized question papers. Task is divide them thematically.
- Uses statistical distribution of words across documents and uses this distribution in a single document to get the ratio of different themes in the document.

- Neural Networks
  - Auto-encoders
  - Generative Adversarial Networks (GAN)
    - One part creates synthetic data; second part is tasked to figure out whether data is synthetic or not.

What is coding?
- Began with Machine Level Instructions
- Assembly Language was then developed as a wrapper around machine Language.
- High level language add another layer of abstraction above the Assembly Language. Eg Python.

compiler and interpreters

Translates in one go	Translates line by line.

Python Tutorial

- '#' for comments
- keywords cannot be used identifiers
- print() for displaying output.
- input() for taking in input
- feature overloading ⇒ can have ~~same~~ different functions in different instances

Eg 3 + 3 = 6

"3" + "3" = '33"