# INTRODUCTION TO ARTIFICIAL INTELLIGENCE

## 1) THE AI PROBLEM

- fundamental challenges and inherent difficulties face when attempting to create truly intelligent artificial system

Key Aspects

- Defining Intelligence for machines:
  - how do explain a machine what is intelligence, because it is about many things like facets, learning, reasoning, creativity, problem-solving, social understanding, emotional intelligence and m
  - e.g. how a program understands and takes joke feel empathy.
  - how do you translate this into algo and data structures

- The common sense problem
  - For humans there are many unsaid intuitive knowledge which are to be considered b doing something, the common SENSE
  - It's for hard to train AI with all of th

*(left margin, sideways text:)*
MAKING AI UNDERSTAND WHATS INTELLIGENCE & KNOWLEDGE

x common SENSE

e.g. AI might learn from images that cars drive on road, but lacks common sense that can won't drive off a cliff (× self knowledge)

- **The Frame Problem**

  When an AI performs an action or senses a change in its environment, it needs to understand what aspects of the world have changed and what's unchanged.

  "About AI knowing what exactly changes and what doesn't in the environment after an action"

  e.g. If a robot moves coffee mug from a table, it needs to know that the table's color didn't change, the room didn't rotate, and the gravity is still working and much more.

- **The Qualification Problem**

  This difficulty is about fully specifying all the conditions, prerequisits, and exceptions for the rule or action to be valid

  "About AI dealing with the fact that there are too many possible conditions to list for an action to work"

<sub>*(left margin, vertical text)*</sub>

Impact on env after an action

Impossible to cover all possible conditions

e.g. A robot instructed to open the door must implicitly assume the door isn't locked, isn't stuck, isn't painted onto the wall, isn't too heavy etc. Listing all these is too much and resulting in failure in unexpected situations

## The AI control problem
challenge of ensuring that advance AIs operate in a way that is beneficial, safe, and aligned with human values, ethics and intensions.
"Designing AI with goal alignment to prevent unihtended and potentially harmful consequences."
e.g. using AI to hack or criminal activities
AI accidently supporting harmful thoughts and encourage them
AI leaking sensitive data

## Ethical Dilemmas and Societal Impact
cocerns about AI bias (training data isn't always correct for all situations), privacy, job loss, no accountability for AI decisions and potential misuse
e.g. AI used for predicting criminal behaviour might inadvertently perpetuato existing social biases if trained on historical biased data, leading unfair outcomes

MISUSE AND X PRIVACY.

JOB DISPLACEMENT
BIAS OUTCOMES
IF DATA BIASED

# THE UNDERLYING ASSUMPTION

"belief an principal that is taken for granted, often without explicit statement an proof"
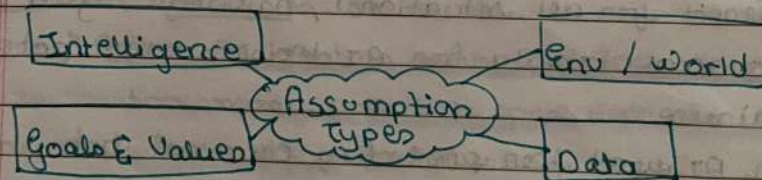
- forms unstated foundation for reasoning, actions en deoing of a system

- we (humans) subconciously an consciously accepts as true to make progress.

- why is it critical?

for solving or acting upon a problem
we inherently assume few thing about the world
AI will operate in, like amount, type of data
and behaviours and more.

These assumptions
shape entire design of the AI
· algorithm choosen
· type and amount of data collected
· how system learns
· how it performs and its measurement

If flawed everything goes wrong.

```
┌──────────────┐                    ┌───────────┐
│ Intelligence │                    │ Env/World │
└──────────────┘      ╭────────╮    └───────────┘
                      │Assumption│
┌──────────────┐      │  Types  │   ┌───────┐
│Goals & Values│      ╰────────╯    │ Data  │
└──────────────┘                    └───────┘
```

## 1. Assumption about Environment & world

"relates to characteristics of the operational environment for an AI agent"

e.g. 1. AI agent designed for playing chess assumes a perfectly known, deterministic, discrete and static env (chess board and rules)

2. self-driving can AI operates in a high dynamic, continuous and partially observable environment. It must assume certain road rules, predictable pedestrain behaviour and general weather patterns etc

## 2. Assumption about Data

→ Data Completness : AI might assume its training data covers everything

→ Data Accuracy : it often assumes data is free from errors etc.

→ Data Distribution : assume that data will be same/similar to training data

→ Data Bias : AI assumes training data is fair and true.

## 3. Assumptions about Goals & Values

"what we define as success, optimal behaviour an ultimate AI objective"

e.g. AI designed purely maximize profit for a company might implicitly assume that short term financial gain is sole objective. ( So AI determines GOAL itself)

4.
## Assumptions about Intelligence
"how intelligence works or how it can be computationally implemented"

e.g. Early AI assumed intelligence could be captured by explicit rules, logics and symbols

Modern AI assumes intelligence emerges from learning complex patterns in vast amount of data, without explicit rules and that sufficient data and computational power will lead to intelligent behaviour.

## Impact of Unstated or Incorrect Assumptions
1) Inaccuracy   2) ↓performance   3) ethical issues (bias)
4) unintended consequences (when in business one e.g. ↑↑)

## AI TECHNIQUES
"specific methods, algorithms and approaches that allow artificial intelligence systems to do tasks that traditionally required human intelligence"

### Categories
1. Symbolic AI   (Rule-Based Systems)
Focuses on representing human knowledge in symbolic, explicit form often using human defined rules.

e.g. expert system designed to diagnose car problems
so rules defined like

IF engine won't start AND battery dead
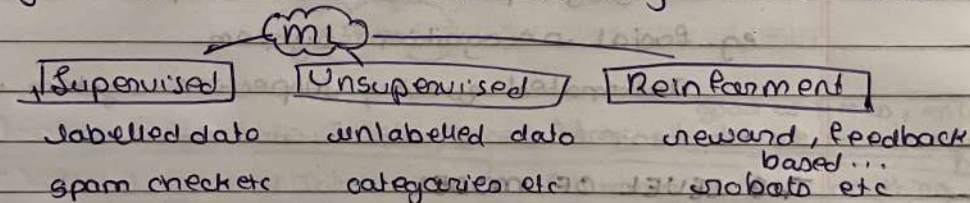THEN check battery terminals
so basically uses collection of if-elses

## 2. Machine Learning
- enables computers to learn from data without being explicitly programmed for everything.
- finds patterns and make predictions an decisions
give @most of your ML knowledge here like



| Supervised | Unsupervised | Reinforment |
|---|---|---|
| labelled data | unlabelled data | reward, feedback based... |
| spam check etc | categories etc | robots etc |

## 3. Search Algorithms
used to explore a set of possible solutions an paths
to find a specific goal
e.g. GPS, road track, self driven cars etc

## 4. Logic Based AI
IDK, leavebhai...

e.g. knowledge based system that can answer quires by
applying logical rules. If you say 'All students are smart'
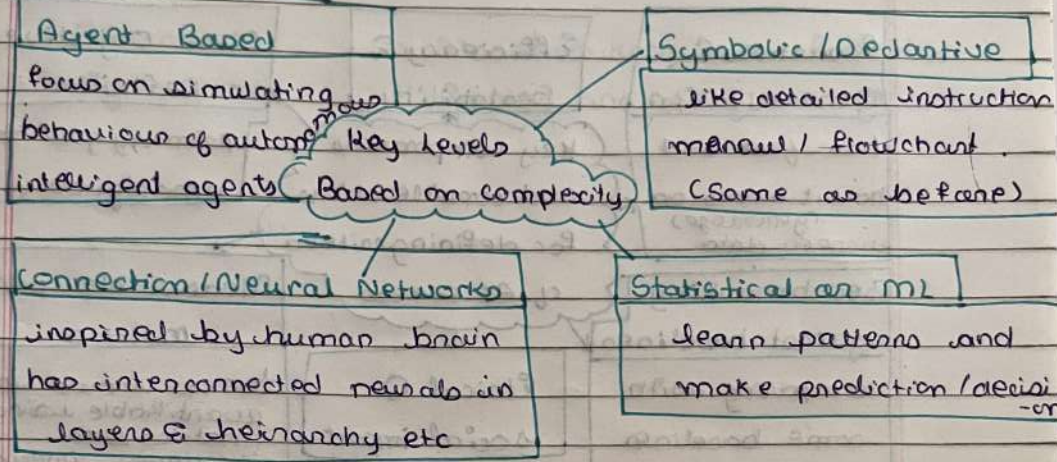'John is a student' then it'll infer 'John is smart'

2. Natural Language Processing
  • enables computers to understand, interpret and generate human language
  • basically allows AI to understand what human says and responds in human way...

3. Computer Vision
  • allows computers to see, process and understand digital images and videos from the real world.
  • relies on ML techniques to identify objects, recognize patterns and interpret scenes.
  eg. facial recognition system
    automated quality inspection system using cams.

THE LEVEL OF MODEL
  • describe how advanced en capable an AI system is, based on how much it can think, learn and act like humans
  • the level helps us to categorized these models based on what they aim to capture and how they respond (act / operate.
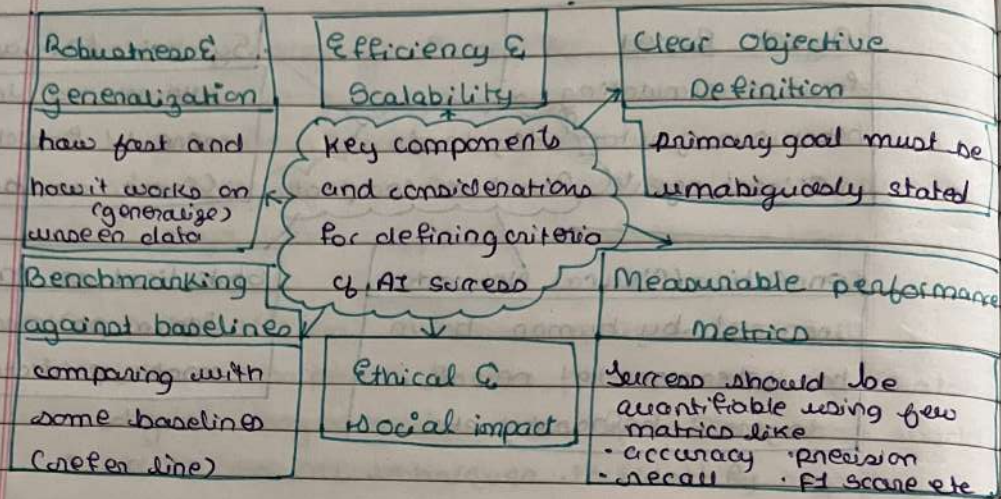
| Agent Based | Symbolic / Dedactive |
|---|---|
| focus on simulating behaviour of autonomous intelligent agents | like detailed instruction manual / flowchart. (Same as before) |

Key Levels
(Based on complexity)

| Connection / Neural Networks | Statistical or ML |
|---|---|
| inspired by human brain has interconnected neurals in layers & heirarchy etc | learn patterns and make prediction / decision |

others

| Artificial Narrow Intelligence | A trained dog which are taug~ good at tricks, only |
|---|---|
| Artificial General Intelligence | A human can learn anything |
| Artificial Super Intelligence | A god like being better than human ~ knows everything |

## CRITERIA FOR SUCCESS

- when developing an AI, it is crucial to define what success looks like began, during and after its creation

- It helps evaluating AI, measuring performance etc

| Robustness & Generalization | Efficiency & Scalability | Clear Objective Definition |
|---|---|---|
| how fast and how it works on (generalize) unseen data | key components and considerations for defining criteria of AI success | primary goal must be unambiguously stated |
| Benchmarking against baselines | | Measurable performance metrics |
| comparing with some baselines (refer line) | Ethical & Social impact | Success should be quantifiable using few matrics like<br>· accuracy · precision<br>· recall · F1 score etc. |

## APPLICATION OF AI

1. Personal Assistant and Smart Devices
2. Healthcare
3. Financial and Banaking
4. Natural Language Processing Applications
   - ∟machine Translation   ∟Spam Filtering
   - ∟Sentiment analysis
5. Computer Vision Applications
   - ∟Face recognition   ∟quality control
   - ∟object detection
6. Transportation and Autonomous vehicles
7. Ecommerce and retail

8. Education

9. Agriculture

10. Manufacturing and Robotics

11. Gaming and Entertainment