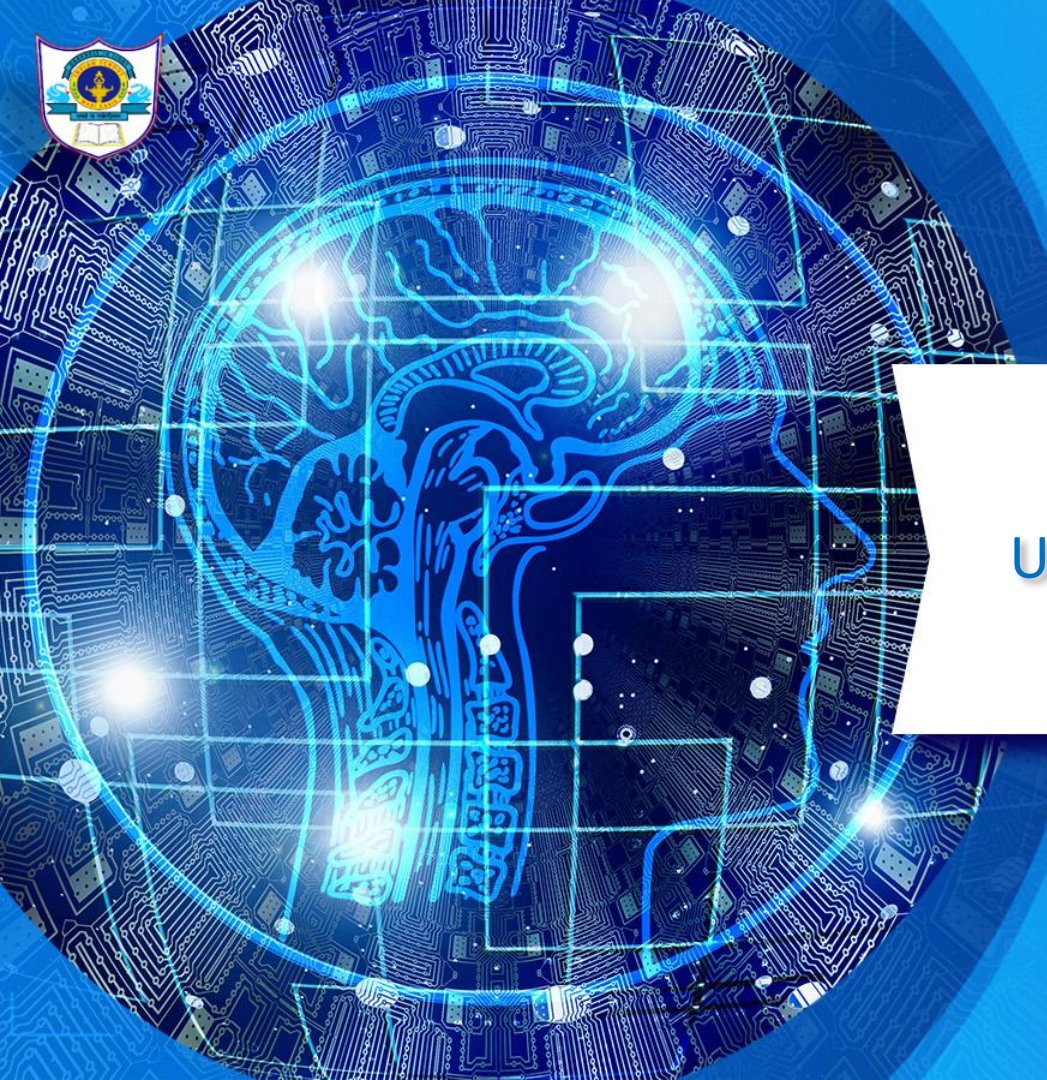


ARTIFICIAL INTELLIGENCE

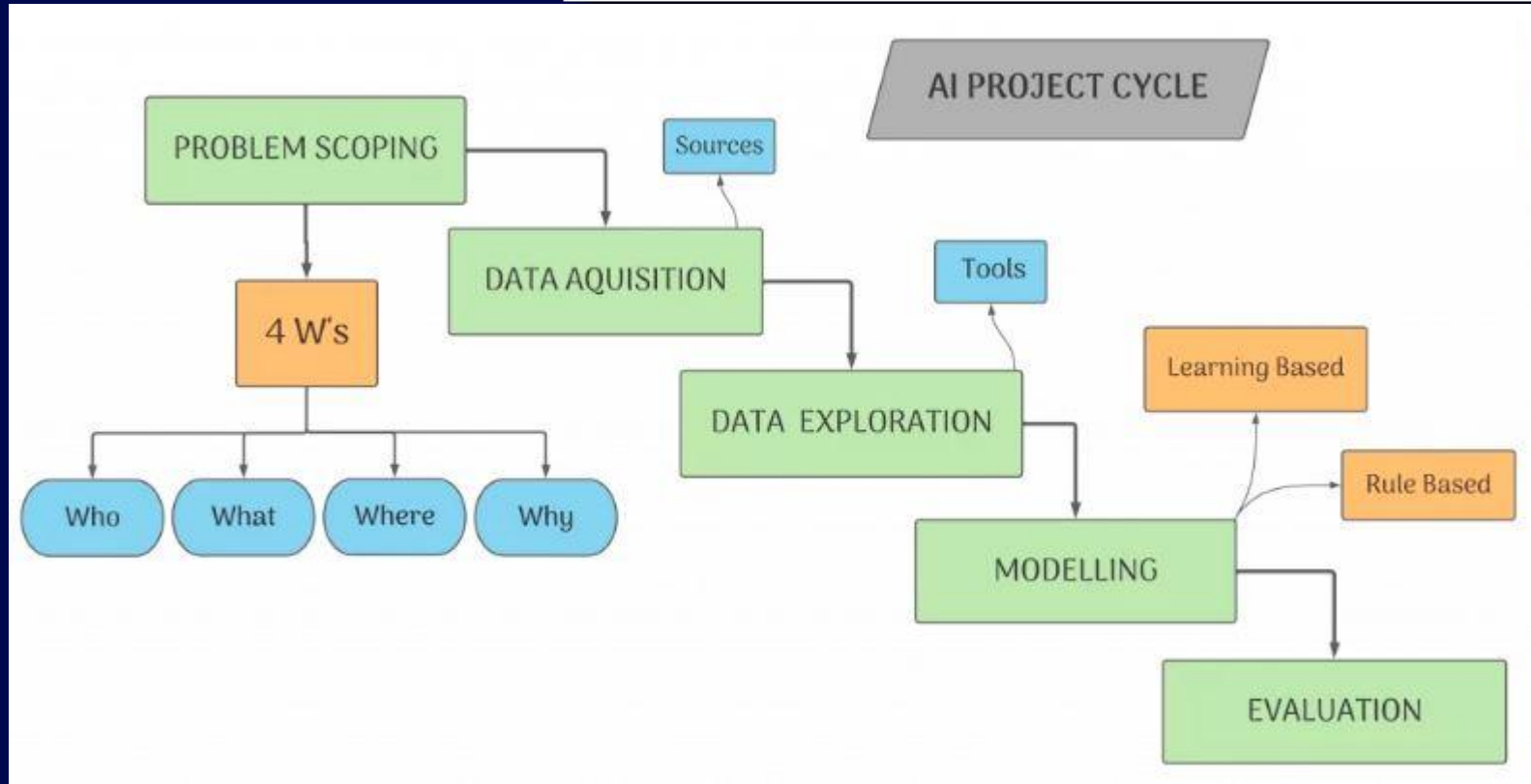
CLASS – X (417)

Unit 1: Revisiting AI Project Cycle & Ethical Frameworks for AI

AI PROJECT CYCLE

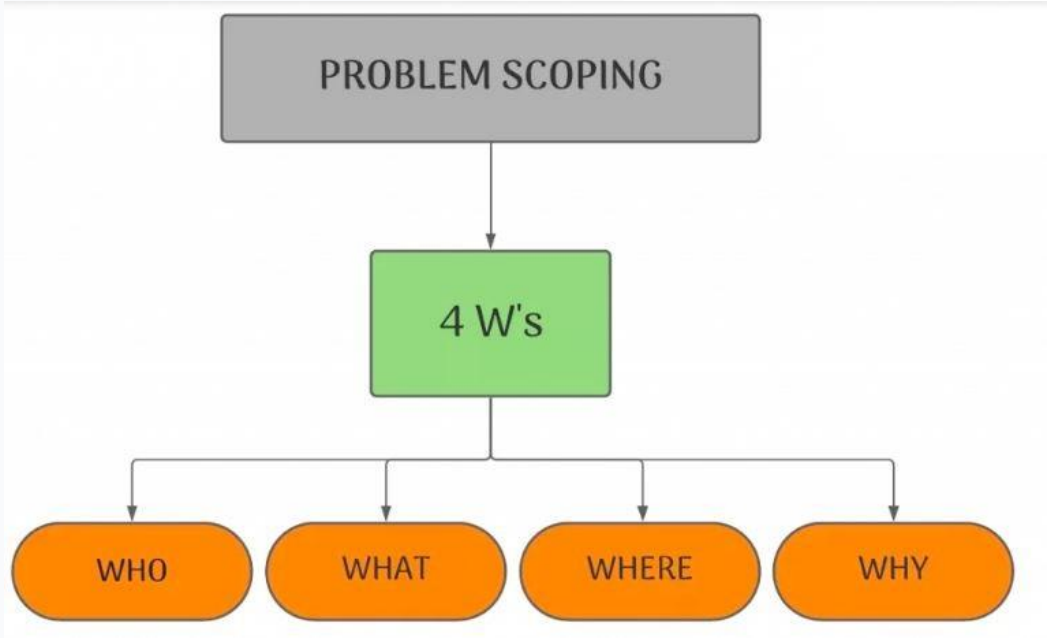
- Life cycle of AI project
- **Project Cycle** is a step-by-step **process to solve** problems using **proven scientific methods** and drawing inferences about them.
- AI PROJECT CYCLE STEPS
    1. Problem Scoping
    2. Data Acquisition
    3. Data Exploration
    4. Modeling
    5. Evaluation
    6. Deployment

AI PROJECT CYCLE



Problem Scoping

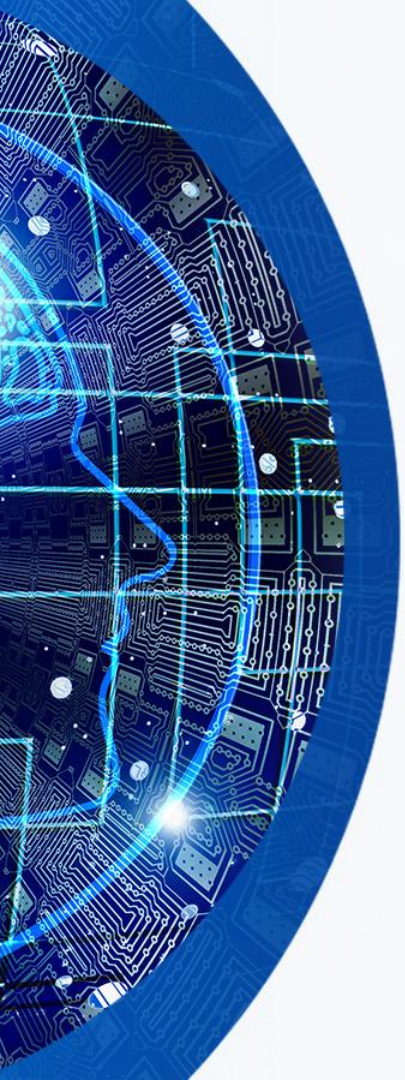
- Problem identification and have a vision to solve it.



Problem Scoping
4Ws in Problem Scoping

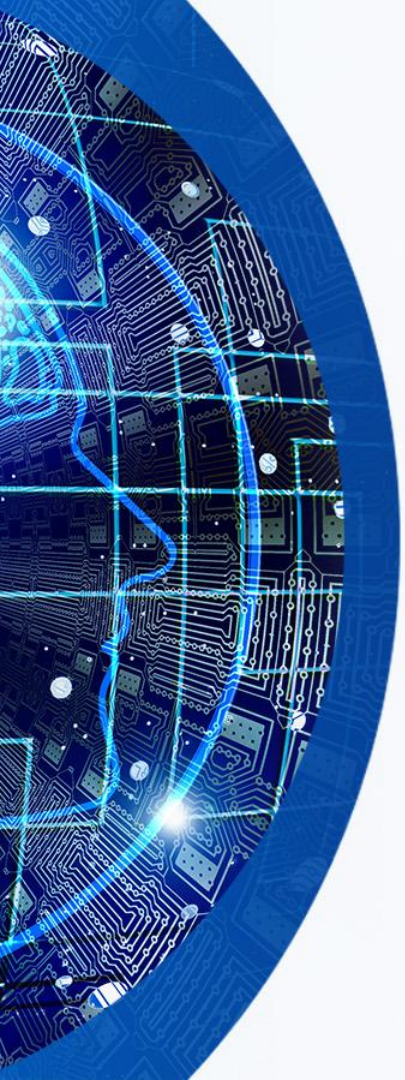
- Who? - Refers that who is facing a problem and who are the stakeholders of the problem.
- What? - Refers to what is the problem and how you know about the problem.
- Where? - It is related to the context or situation or location of the problem.
- Why? - Refers to why we need to solve the problem and what are the benefits to the stakeholders after solving the problem.

The final outcome of problem scoping is the problem statement template.



The problem statement template

- When the above 4Ws are completely filled you need to prepare a summary of these 4Ws.
- This summary is known as the problem statement template.
- This template explains all the key points in a single template.
- So if the same problem arises in the future this statement helps to resolve it easily.



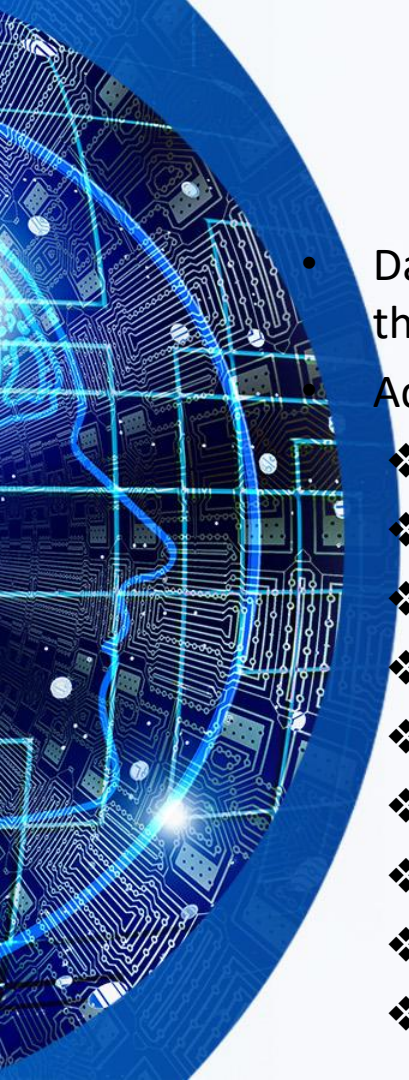
Data Acquisition

- You need to acquire data which will become the base of your project as it will help you understand what the parameters that are related to problem scoping are.
- You go for data acquisition by collecting data from various reliable and authentic sources.

1**. Training Data:** The collected data through the system is known as training data. In other words the input given by the user in the system can be considered as training data.

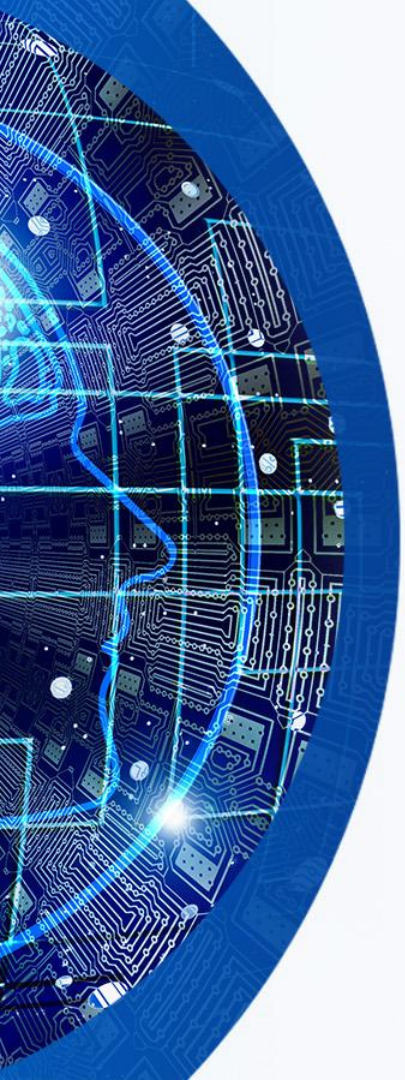
2. **Testing Data:** The result data set or processed data is known as testing data. In other words, the output of the data is known as testing data.

- Eg: to make an Artificially Intelligent system which can predict the salary of any employee based on his previous salaries.
- The previous salary data here is known as **Training Data** while the next salary prediction data set is known as the **Testing Data**
- **For any AI project to be efficient, the training data should be authentic and relevant to the problem statement scoped. .**



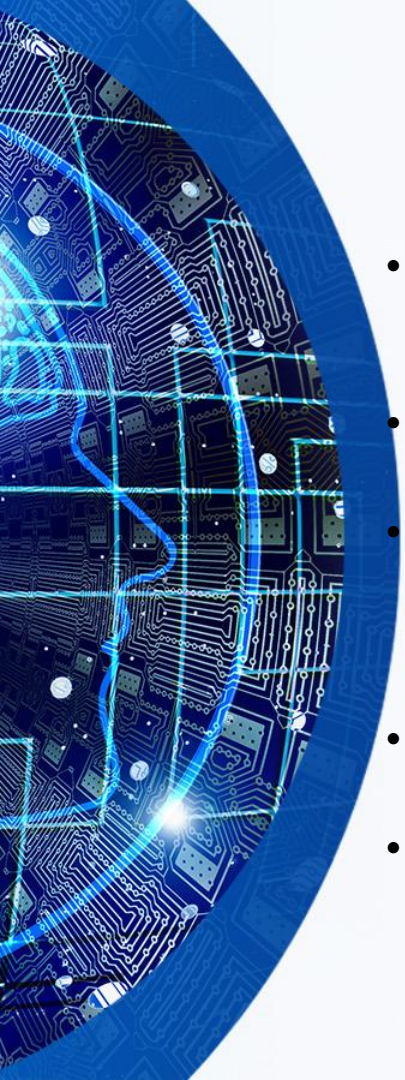
Data Exploration

- Data Exploration refers to the techniques and tools used to visualize data through complex statistical methods.
- Advantages of Data Visualization
  - ❖ A better understanding of data
  - ❖ Provides insights into data
  - ❖ Allows user interaction
  - ❖ Provide real-time analysis
  - ❖ Help to make decisions
  - ❖ Reduces complexity of data
  - ❖ Provides the relationships and patterns contained within data
  - ❖ Define a strategy for your data model
  - ❖ Provides an effective way of communication among users



Visual representations

- Bar Graph
- Line Graph
- Scatter Plot
- Pie chart
- Tree Diagram
- Time table
- Venn Diagrams
- Brainstorm
- Histogram



Modelling

- AI Modelling refers to developing algorithms, also called models which can be trained to get intelligent outputs. That is, writing codes to make a machine artificially intelligent.
- The graphical representation makes the data understandable for humans as we can discover trends and patterns out of it.
- But when it comes to machine accessing and analysing data, it needs the data in the most basic form of numbers (which is binary – 0s and 1s) and when it comes to discovering patterns and trends in data, the machine goes for mathematical representations of the same.
- The ability to mathematically describe the relationship between parameters is the heart of every AI model.
- Thus, whenever we talk about developing AI models, it is the mathematical approach towards analysing data which we refer to.

EVALUATION

Introduction

In the Evaluation stage, we will explore different methods of evaluating an AI model. Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future

What is evaluation?

Evaluation is the process of understanding the reliability of any AI model, based on outputs by feeding test dataset into the model and comparing with actual answers. There can be different Evaluation techniques, depending of the type and purpose of the model. Remember that It's not recommended to use the data we used to build the model to evaluate it. This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set. This is known as overfitting.

What is deployment?

Deployment as the final stage in the AI project cycle where the AI model or solution is implemented in a real-world scenario. Key Steps in Deployment Process the key steps involved in the deployment process: a. Testing and validation of the AI model b. Integration of the model with existing systems c. Monitoring and maintenance of the deployed model. Some examples of successful AI projects that have been deployed in various industries, such as self driving cars, medical diagnosis systems, and chatbots.

AI can be used on Mobile Apps, Website Apps, etc.

Introduction to AI Domains

1. Statistical Data

- Statistical Data is a domain of AI related to data systems and processes, in which the system collects numerous data, maintains data sets and derives meaning/sense out of them.
- The information extracted through statistical data can be used to make a decision about it.

Example of Statistical Data -Price Comparison Websites

These websites are being driven by lots and lots of data. If you have ever used these websites, you would know, the convenience of comparing the price of a product from multiple vendors in one place. PriceGrabber, PriceRunner, Junglee, Shopzilla, DealTime are some examples of price comparison websites. Nowadays, price comparison websites can be found in almost every domain such as technology, hospitality, automobiles, durables, apparel, etc.

2. Computer Vision

- Computer Vision, abbreviated as CV, is a domain of AI that depicts the capability of a machine to get and analyse visual information and afterwards predict some decisions about it. The entire process involves image acquiring, screening, analysing, identifying and extracting information. This extensive processing helps computers to understand any visual content and act on it accordingly.
- In computer vision, Input to machines can be photographs, videos and pictures from thermal or infrared sensors, indicators and different sources.
- Computer vision-related projects translate digital visual data into descriptions. This data is then turned into computer-readable language to aid the decision-making process. The main objective of this domain of AI is to teach machines to collect information from pixels.

Examples of Computer Vision

1. Agricultural Monitoring Computer vision is employed in agriculture for crop monitoring, pest detection, and yield estimation. Drones with cameras capture aerial images of farmland, which are then analysed to assess crop health and optimize farming practices.

2. Surveillance Systems

Computer vision is used in surveillance systems to monitor public spaces, buildings, and borders. It can detect suspicious activities, track individuals or vehicles, and provide real-time alerts to security personnel.

3. Natural Language Processing

- Natural Language Processing, abbreviated as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. Natural language refers to language that is spoken and written by people, and natural language processing (NLP) attempts to extract information from the spoken and written word using algorithms.
- The ultimate objective of NLP is to read, decipher, understand, and make sense of human languages in a valuable manner

Examples of Natural Language Processing

1. Email filters

Email filters are one of the most basic and initial applications of NLP online. It started with spam filters, uncovering certain words or phrases that signal a spam message.

2.Machine Translation

NLP is used in machine translation systems like Google Translate and Microsoft Translator to automatically translate text from one language to another. These systems analyze the structure and semantics of sentences in the source language and generate equivalent translations in the target language.

Ethical Frameworks for AI

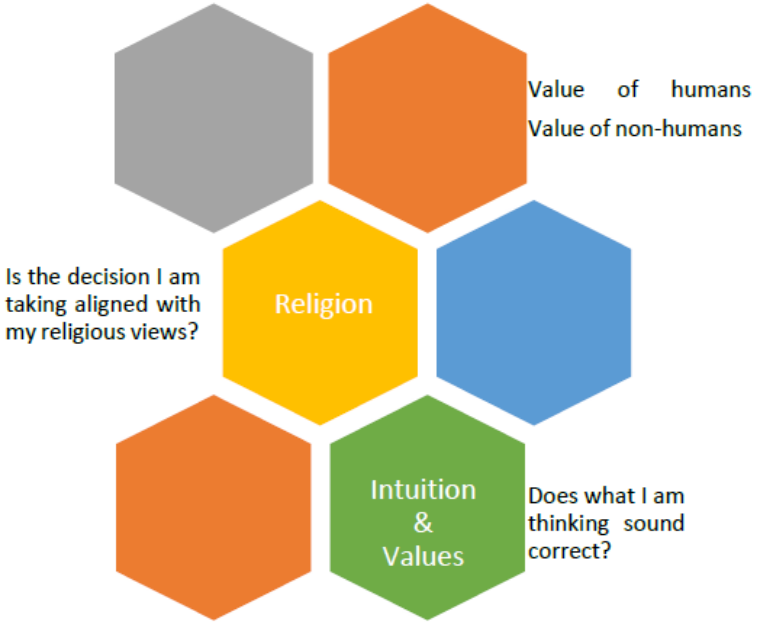
Frameworks

Frameworks are a set of steps that help us in solving problems. It provides a step-by-step guide for solving problems in an organized manner. Moreover, frameworks offer a structured approach to problem-solving, ensuring that all relevant factors and considerations are taken into account.

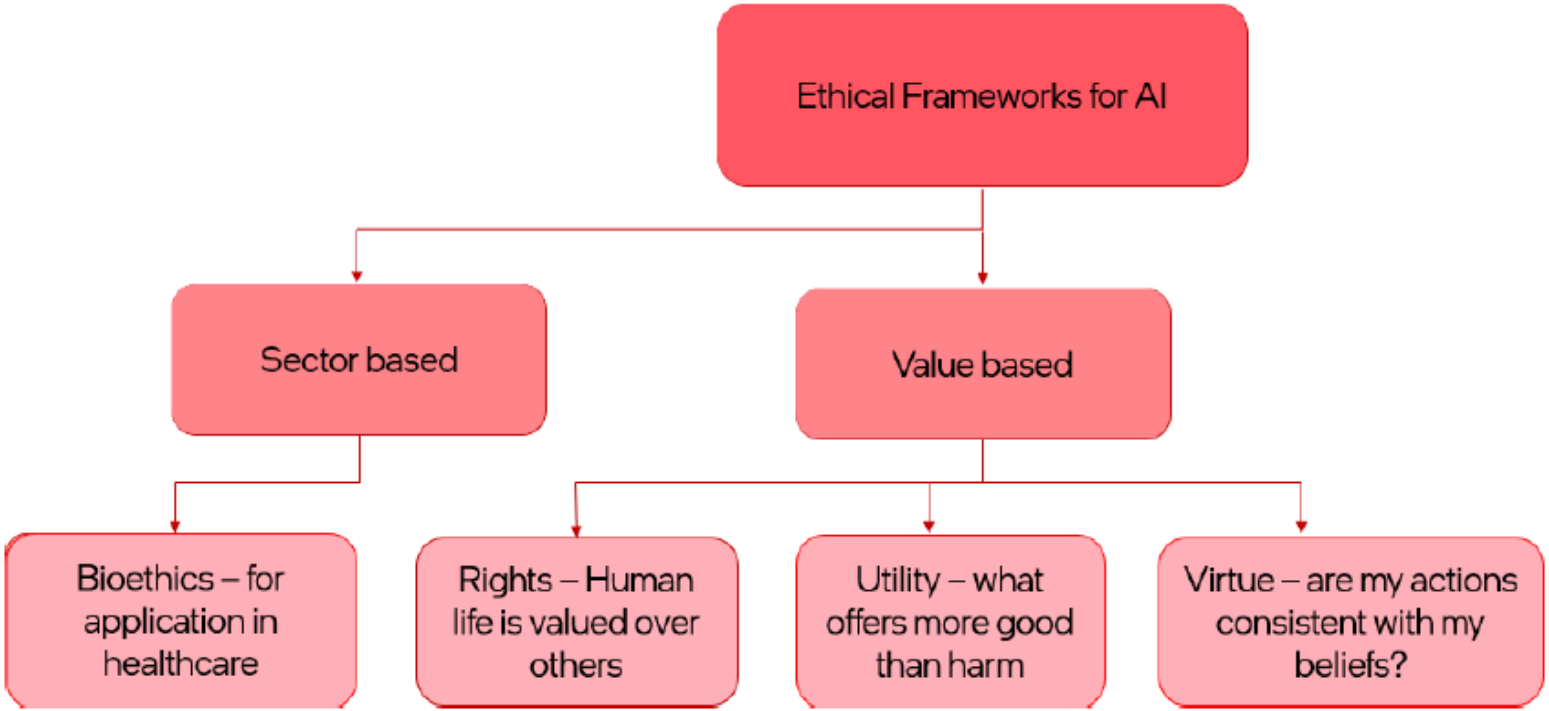
Ethical Frameworks

- Ethical frameworks are frameworks which help us ensure that the choices we make do not cause unintended harm.
- Ethical frameworks provide a systematic approach to navigating complex moral dilemmas by considering various ethical principles and perspectives.
- By utilizing ethical frameworks, individuals and organizations can make well- informed decisions that align with their values and promote positive outcomes for all stakeholders involved.

- **Why do we need Ethical Frameworks for AI?**
- Bias could result in unwanted outcomes in AI solutions. Think of the hiring algorithm which is biased against women applicants.AI is essentially being used as a decision-making/ influencing tool. As such we need to ensure that AI makes morally acceptable recommendations
- Ethical frameworks ensure that AI makes morally acceptable choices. If we use ethical frameworks while building our AI solutions, we can avoid unintended outcomes, even before they take place.
- **3 factors** which knowingly or unknowingly influence decision-making are: Religion, values and Intuitions
- **Visit the website – https://www.my-goodness.net/**



Types of Ethical Frameworks



ETHICAL FRAMEWORKS

Ethical frameworks for AI can be categorized into **two main** types: **sector-based and value- based frameworks.**

1. Sector-based Frameworks:

These are frameworks tailored to specific sectors or industries. In the context of AI, one common sector-based framework is Bioethics, which focuses on ethical considerations in healthcare. It addresses issues such as patient privacy, data security, and the ethical use of AI in medical decision-making. Sector-based ethical frameworks may also apply to domains such as finance, education, transportation, agriculture, governance, and law enforcement.

Bioethics

Bioethics is an ethical framework used in healthcare and life sciences. It deals with ethical issues related to health, medicine, and biological sciences, ensuring that AI applications in healthcare adhere to ethical standards and considerations.

Principles of bioethics:

Respect for Autonomy.

Do not harm.

Ensure maximum benefit for all.

Give justice.

Sector-based Frameworks:

1. Respect for autonomy: Enabling users to be fully aware of decision-making. E.g., users of an AI algorithm should know how it functions. The data that models were trained on, and used to make decisions, should be reproducible and accessible to the patients. In the event of performance concerns, model predictions and data labels should be released.

2. Do not harm: Harm to anyone (be it human or non–human) must be avoided at all costs. If no choice is available path of least harm must be always chosen.

Promote well-being, minimize harm, and ensure that benefits and harms are distributed among stakeholders in a just manner.

The AI algorithm must be trained on data sets that equitably reduce harm for all, not just harm for some groups.

In this instance, patients from other regions excluding western part who were less ill would receive more intensive care from doctors than patients who actually require help.

This algorithm, if implemented, would actively harm patients belong to western region by inappropriately recommending healthcare resource allocation.

SECTOR-BASED FRAMEWORKS

3. Maximum benefit: Not only should we avoid harm our actions must focus on providing the maximum benefit possible. The solution should be held to clinical practice standards, not merely technological ethics standards.
It should go beyond nonmaleficence and strive for beneficence.
The data we use for training must be unbiased.

4. Justice: All benefits and burdens of a particular choice must be distributed in a justified manner across people irrespective of their background. Solution development requires concerted and in-depth knowledge of social structures at play that result in issues like racism and sexism (a few types of societal biases).
The solution needs to be aware of social determinants of healthcare and actively work against those structures.

VALUE-BASED FRAMEWORKS

2. Value-based Frameworks: Value-based frameworks focus on fundamental ethical principles and values guiding decision- making. It reflects the different moral philosophies that inform ethical reasoning. Value-based frameworks are concerned with assessing the moral worth of actions and guiding ethical behaviour. They can be further classified into three categories:

i. Rights-based: Prioritizes the protection of human rights and dignity, valuing human life over other considerations. It emphasizes the importance of respecting individual autonomy, dignity, and freedoms. **In the context of AI, this could involve ensuring that AI systems do not violate human rights or discriminate against certain groups.**

ii. Utility-based: Evaluates actions based on the principle of maximizing utility or overall good, aiming to achieve outcomes that offer the greatest benefit and minimize harm. It seeks to maximize overall utility or benefit for the greatest number of people. **In AI, this might involve weighing the potential benefits of AI applications against the risks they pose to society, such as job displacement or privacy concerns.**.

iii. Virtue-based: This framework focuses on the character and intentions of the individuals involved in decision-making. It asks whether the actions of individuals or organizations align with virtuous principles such as honesty, compassion, and integrity. **In the context of AI, virtue ethics could involve considering whether developers, users, and regulators uphold ethical values throughout the AI lifecycle**.

“**Non-maleficence**” refers to the ethical principle of avoiding causing harm or negative consequences. It emphasizes the obligation to minimize harm as much as possible and prioritize actions that prevent harm to individuals, communities, or the environment.

“**Maleficence**” refers to the concept of intentionally causing harm or wrongdoing.

“**Beneficence**" refers to the ethical principle of promoting and maximizing the well-being and welfare of individuals and society. It emphasizes taking actions that produce positive outcomes and contribute to the overall good, ensuring that the greatest benefit is achieved for all stakeholders involved.

THANK YOU