

TYPES OF ETHICAL FRAMEWORK SUMMARY NOTES

Title: Ethical Frameworks in AI

◆ Definition

An **ethical framework** is a system of principles and guidelines that helps individuals or organizations determine what is **right or wrong** and guides them in **moral decision-making** and **ethical actions**.

◆ Relevance to AI

- AI affects many spheres of life: medicine, business, climate, manufacturing, etc.
- While **AI is useful**, it can also make **mistakes**, be **unfair**, or **harm people**.
- Thus, ethical rules are necessary to ensure AI acts **fairly, safely, and helpfully**.

◆ What Are Ethical Frameworks of AI?

The **rules around the use of ethics in AI decision-making** are called **ethical frameworks of AI**.

AI Ethics and Ethical Principles of AI

◆ AI Ethics

- A **set of rules and guidelines**.
- Ensures AI acts in a **good, fair, and legal** manner.
- AI should:
 - **Help** people, not harm them.
 - Be **clear** and follow **laws**.

◆ Key Ethical Principles

1. AI should **respect human rights**.
 2. AI should be **fair and safe**.
 3. AI should **not replace humans** in important decisions.
-

✚ Note (Highlighted)

The rules built around the use of ethics in AI decision-making are known as **ethical frameworks of AI**.

1. Consequentialism Framework

◆ Definition:

- Morality of an action is based on its consequences.

- The right action is the one that produces the most happiness or least harm for the greatest number of people.

◆ Consequentialist AI ensures:

- AI actions maximize positive results and minimize harm.
- A self-driving car should choose the less harmful option during an accident.
- Governments may impose lockdowns via AI to protect public health.

◆ *Examples of Consequentialism in AI:*

- AI in finance to prevent market crashes.
 - AI in healthcare to allocate resources and save lives.
 - AI in traffic systems to reduce accidents and optimize fuel use.
-

2. Utilitarianism Framework

(A subtype of Consequentialism)

◆ Definition:

- Actions are ethical if they maximize overall happiness or well-being for the greatest number.
- Moral value depends on consequences.

◆ Utilitarian AI ensures:

- AI decisions should benefit the most people.
- AI should prioritize treatments that help the most patients.
- Governments can use AI for universal healthcare systems.

◆ *Examples of Utilitarianism in AI:*

- AI for environmental impact reduction through sustainability.
- AI in education, prioritizing programs benefiting most students.
- AI in productivity, improving output while reducing economic disruptions.

3. Deontological Ethics Framework (Duty-Based Ethics)

- Based on **rules, duties, and obligations**, not outcomes.
- Ensures AI:
 - Does **not harm humans**.
 - Acts with **honesty, fairness, and truthfulness**.
- AI should **not discriminate** and must follow **anti-discrimination laws**.
- AI systems must **respect privacy and human rights**.

◆ *Examples of Applications:*

- AI in **healthcare**, applying equal rules regardless of background.
 - AI in **hiring**, ensuring fair treatment and avoiding bias.
-

4. Virtue Ethics Framework

- Focuses on **virtues** (character traits) like **honesty**, **kindness**, **empathy**, **courage**, and **fairness**.
- Encourages **ethical character**, not just rule-following.

◆ *Virtue Ethics ensures:*

- AI should **act with honesty**, **kindness**, and **empathy**.
- AI should be **polite and respectful**.
- Should reflect **human values** like **trustworthiness**.

◆ *Examples of Applications:*

- AI that **communicates with empathy**.
 - AI systems that **prevent fake news**.
 - AI in **business** acting ethically, not just for profit.
-

5. Human-Centered AI Framework

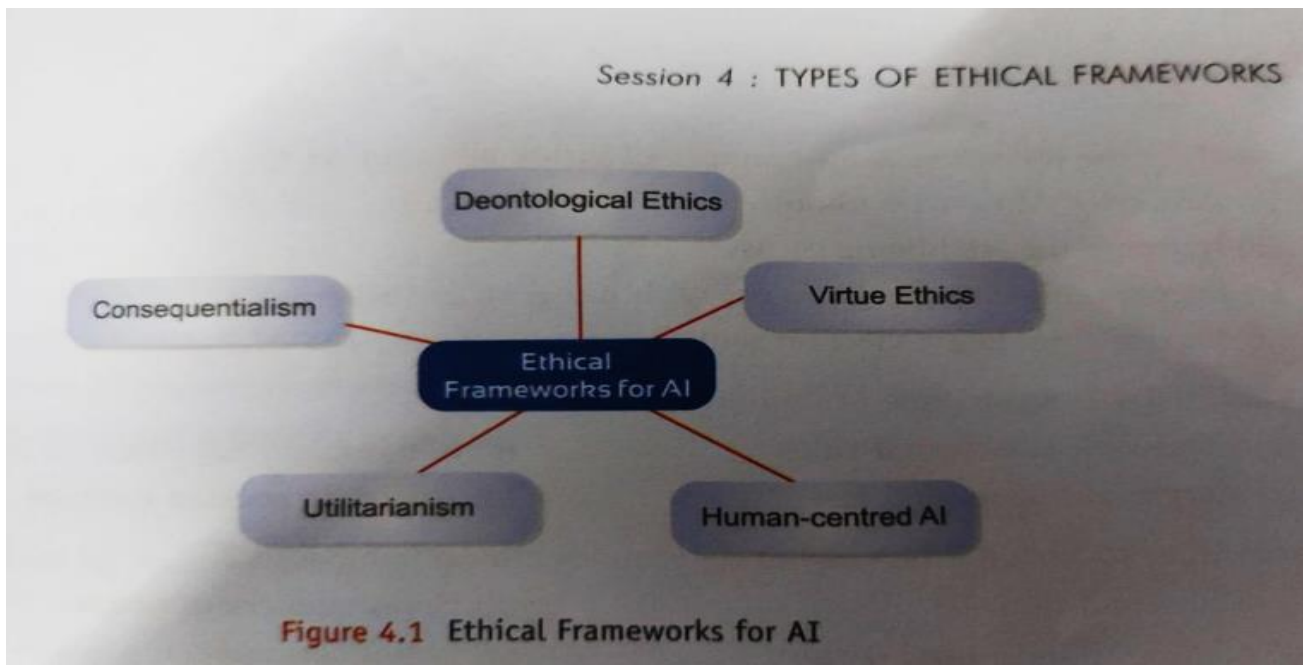
- Emphasizes **human values**, **needs**, and **empathy**.
- Less focused on abstract rules or consequences.

◆ *Human-centered AI ensures:*

- AI must **help** and **respect** humans.
- AI should **work with** humans, not replace them.
- AI should reflect **human interests** and **values**.

◆ *Examples of Applications:*

- AI in **workplaces**, supporting **human tasks**.
- AI in **healthcare**, helping with **better decisions**.
- AI in **governance**, ensuring **transparency** and **inclusion**.



LET US REVISE

- An ethical framework refers to a system of principles and guidelines that helps individuals or organizations determine what is right and wrong, guiding them in making moral decisions and taking ethical actions.
- As per the Deontological Ethics framework, actions are considered ethical if they follow a set of established rules or duties, regardless of the outcome.
- Consequentialism framework emphasizes on the consequences of actions, i.e., an action is deemed ethical based on the outcomes it produces.
- As per the Utilitarianism framework, the actions are ethical if they maximize overall happiness or well-being for the greatest number of people.
- Virtue ethics framework emphasizes on character and moral virtues of an individual.
- Human-centered AI frameworks emphasize human values, need, and empathy rather than abstract rules or general consequences.