Department of
Information Technology

RJITC04
ARTIFICIAL INTELLIGENCE

Day 1/ Session 2
Ethics and Applications of AI

Prof. Aditi Prajapati
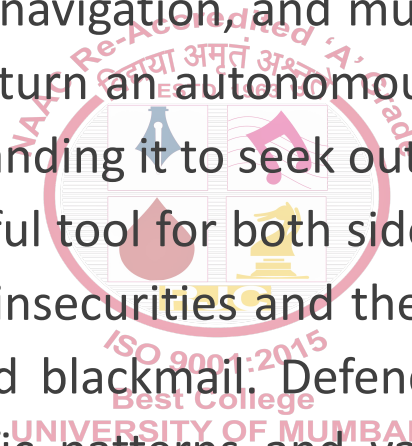Department of Information Technology

09-12-2024

Content Overview

- Ethics in AI

- Applications of AI

Introduction

- AI is a **dual use technology**: AI technologies that have peaceful applications such as flight control, visual tracking, mapping, navigation, and multiagent planning, can easily be applied to military purposes. It is easy to turn an autonomous quadcopter into a weapon simply by attaching an explosive and commanding it to seek out a target.

- Machine learning can be a powerful tool for both sides in the cybersecurity battle. Attackers can use automation to probe for insecurities and they can apply reinforcement learning for phishing attempts and automated blackmail. Defenders can use unsupervised learning to detect anomalous incoming traffic patterns and various machine learning techniques to detect fraud
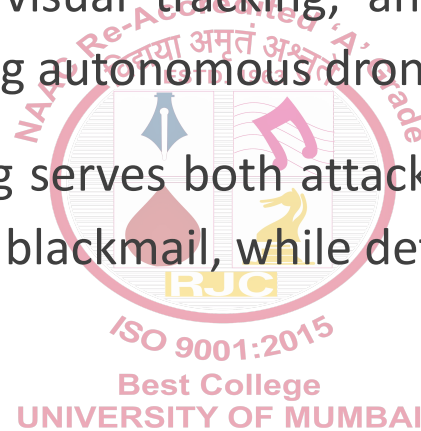
Introduction

AI is a dual-use technology with applications in both peaceful and military domains. Technologies such as flight control, visual tracking, and multi-agent planning can be easily repurposed for warfare, such as turning autonomous drones into weapons.

In cybersecurity, machine learning serves both attackers and defenders: attackers use it for automated exploitation, phishing, and blackmail, while defenders leverage it to detect anomalies, fraud, and suspicious traffic patterns.

Ethics of AI

The gist is that every organization that creates AI technology, and everyone in the organization, has a responsibility to make sure the technology contributes to good, not harm. The most commonly-cited principles are:

1. Ensure Safety
2. Ensure Fairness
3. Respect Privacy
4. Promote Collaboration
5. Provide Transparency
6. Limit Harmful uses of AI
7. Uphold human rights and value
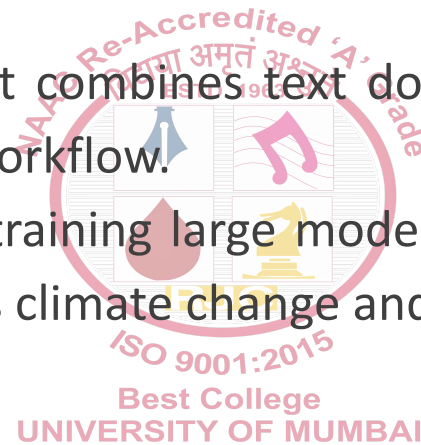8. Minimize environmental risks

Continue…

- The 2011 U.S. Department of Defense (DOD) roadmap says: "For the foreseeable future, decisions over the use of force [by autonomous systems] and the choice of which individual targets to engage with lethal force will be retained under human control." The primary reason for this policy is practical: "autonomous systems are not reliable enough to be trusted with military decisions."

- AI engineers should be clear on what uses of surveillance are compatible with human rights, and decline to work on applications that are incompatible.

- People with similar likelihood to pay back the loan should have the same expected utility, regardless of the class they belong to.

- To earn trust, any engineered systems must go through a **verification and validation** (V&V) process. Verification means that the product satisfies the specifications. Validation means ensuring that the specifications actually meet the needs of the user and other affected parties.

- When an AI system turns you down for a loan, you deserve an explanation - explainable AI (XAI).

- Business process automation that combines text documents and structured data to make business decisions and improve workflow.

- Address the carbon footprint of training large models and promote the use of AI to tackle environmental challenges, such as climate change and resource optimization.
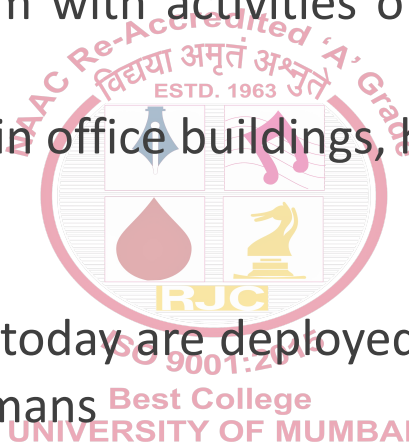
Prof. Aditi Prajapati
Department of Information Technology

Applications of AI

- HOMECARE: Robots have started to enter the home to care for older adults and people with motor impairments, assisting them with activities of daily living and enabling them to live more independently.
- SERVICES: Mobile robots help out in office buildings, hotels, and hospitals.
- ENTERTAINMENT
- AUTONOMOUS CARS
- INDUSTRY: The majority of robots today are deployed in factories, automating tasks that are difficult, dangerous, or dull for humans
- DRONES: Agriculture, wildlife photography
- SPORTS

Prof. Aditi Prajapati
Department of Information Technology

Group Discussion

- AI benefits and drawbacks

- Where AI is used and can be used?

- Use of Robots in real life

- Will AI cause unemployment?

- AI in Healthcare (Benefits, Risk and Challenges)

- AI in Education (Benefits and Challenges)

- AI in Entertainment and Media (Benefits and Challenges)

Prof. Aditi Prajapati
Department of Information Technology

Reference

Chapter 27 of Artificial Intelligence: A Modern Approach, Fourth Edition, Stuart J. Russell and Peter Norvig, Pearson, [2021].

Prof. Aditi Prajapati
Department of Information Technology

THANK YOU

Prof. Aditi Prajapati
Department of Information Technology