Chapter 1: Introductions

1.1 Definitions:

What Is Artificial Intelligence?

- Artificial intelligence refers to computer systems that can perform tasks commonly associated with human cognitive functions — such as interpreting speech, playing games and identifying patterns.

- Typically, AI systems learn how to do so by processing massive amounts of data and looking for patterns to model in their own decision-making.
- In many cases, humans will supervise an AI's learning process, reinforcing good decisions and discouraging bad ones.

- But some AI systems are designed to learn without supervision; for instance, by playing a game over and over until they eventually figure out the rules and how to win.

Strong AI vs. Weak AI

Artificial intelligence is often distinguished between weak AI and strong AI. Weak AI (or narrow AI) refers to AI that automates specific tasks, typically outperforming humans but operating within constraints. Strong AI (or artificial general intelligence) describes AI that can emulate human learning and thinking, though it remains theoretical for now.

Weak AI

- Also called narrow AI, weak AI operates within a limited context and is applied to a narrowly defined problem.

- It often operates just a single task extremely well.

- Common weak AI examples include email inbox spam filters, language translators, website recommendation engines and conversational chatbots.

Strong AI

- Often referred to as artificial general intelligence (AGI) or simply general AI, strong AI describes a system that can solve problems it's never been trained to work on, much like a human can.

- AGI does not actually exist yet.

- For now, it remains the kind of AI we see depicted in popular culture and science fiction.

1.2 Benefits of AI

AI is beneficial for automating repetitive tasks, solving complex problems, reducing human error and much more.

Automating Repetitive Tasks

- Repetitive tasks such as data entry and factory work, as well as customer service conversations, can all be automated using AI technology. This lets humans focus on other priorities.

Solving Complex Problems

- AI's ability to process large amounts of data at once allows it to quickly find patterns and solve complex problems that may be too difficult for humans, such as predicting financial outlooks or optimizing energy solutions.

Improving Customer Experience

- AI can be applied through user personalization, chatbots and automated self-service technologies, making the customer experience more seamless and increasing customer retention for businesses.

Advancing Healthcare and Medicine

- AI works to advance healthcare by accelerating medical diagnoses, drug discovery and development and medical robot implementation throughout hospitals and care centers.

Reducing Human Error

- The ability to quickly identify relationships in data makes AI effective for catching mistakes or anomalies among mounds of digital information, overall reducing human error and ensuring accuracy.

1.3 Disadvantages of AI

While artificial intelligence has its benefits, the technology also comes with risks and potential dangers to consider.

Job Displacement

- AI's abilities to automate processes, generate rapid content and work for long periods of time can mean job displacement for human workers.

Bias and Discrimination

- AI models may be trained on data that reflects biased human decisions, leading to outputs that are biased or discriminatory against certain demographics.

Privacy Concerns

- The data collected and stored by AI systems may be done so without user consent or knowledge, and may even be accessed by unauthorized individuals in the case of a data breach.

Ethical Concerns

- AI systems may be developed in a manner that isn't transparent, inclusive or sustainable, resulting in a lack of explanation for potentially harmful AI decisions as well as a negative impact on users and businesses.

Environmental Costs

- Large-scale AI systems can require a substantial amount of energy to operate and process data, which increases carbon emissions and water consumption.

1.4 Artificial Intelligence Examples

Specific examples of AI include:

Generative AI Tools

- Generative AI tools, sometimes referred to as chatbots — including ChatGPT, Gemini, Claude and Grok — use artificial intelligence to produce

written content in a range of formats, from essays to code and answers to simple questions.

Smart Assistants

- Personal AI assistants, like Alexa and Siri, use natural language processing to receive instructions from users to perform a variety of ‘smart tasks.’ They can carry out commands like setting reminders, searching for online information or turning off your kitchen lights.

Self-Driving Cars

- Self-driving cars are a recognizable example of deep learning, since they use deep neural networks to detect objects around them, determine their distance from other cars, identify traffic signals and much more.

Wearables

- Many wearable sensors and devices used in the healthcare industry apply deep learning to assess the health condition of patients, including their blood sugar levels, blood pressure and heart rate. They can also derive patterns from a patient’s prior medical data and use that to anticipate any future health conditions.

Visual Filters

- Filters used on social media platforms like TikTok and Snapchat rely on algorithms to distinguish between an image’s subject and the background, track facial movements and adjust the image on the screen based on what the user is doing.

1.5 How AI Works

How AI Systems Work (Step-by-Step)

Step 1: Data Collection

- AI systems begin by collecting large volumes of data.
- Data can be structured (databases) or unstructured (images, text, audio).

Step 2: Data Preprocessing

- Clean and transform data into a usable format.

- May involve handling missing values, normalization, or tokenization (for text).

Step 3: Model Selection and Training

- Choose an appropriate algorithm (e.g., decision tree, neural network).
- Feed training data to the model to help it learn patterns and relationships.

Step 4: Evaluation

- Use a separate test dataset to assess model accuracy and performance using metrics like precision, recall, and F1-score.

Step 5: Deployment

- The trained AI model is integrated into applications or systems to make predictions or automate tasks in real-time.

Step 6: Feedback and Learning

- AI systems can improve over time using new data (continuous learning or online learning).

1.6 Key AI Algorithms

1. Linear Regression

- **Description:** A simple algorithm used for predicting a continuous target variable based on one or more input features.
- **Use Case:** Predicting house prices based on features like square footage, location, etc.
- **Type:** Supervised learning.

2. Logistic Regression

- **Description:** A classification algorithm used to predict the probability of a binary outcome (0 or 1).
- **Use Case:** Predicting whether an email is spam or not.
- **Type:** Supervised learning.

3. Decision Trees

- **Description:** A model that splits data into subsets based on feature values. It creates a tree-like structure for decision-making.
- **Use Case:** Credit scoring or medical diagnosis.
- **Type:** Supervised learning.

4. Random Forest

- **Description:** An ensemble method that builds multiple decision trees and merges them to improve prediction accuracy and control overfitting.
- **Use Case:** Predicting customer churn or fraud detection.
- **Type:** Supervised learning.

5. Support Vector Machines (SVM)

- **Description:** A classification algorithm that finds the optimal hyperplane to separate different classes in a high-dimensional space.
- **Use Case:** Image classification or text categorization.
- **Type:** Supervised learning.

6. K-Nearest Neighbors (KNN)

- **Description:** A non-parametric algorithm that classifies data points based on the majority class of their nearest neighbors.
- **Use Case:** Recommender systems or anomaly detection.
- **Type:** Supervised learning.

7. Naive Bayes

- **Description:** A classification algorithm based on Bayes' Theorem, assuming independence between features.
- **Use Case:** Sentiment analysis, spam detection.
- **Type:** Supervised learning.

8. K-Means Clustering

- **Description:** An unsupervised learning algorithm that groups data into K clusters based on similarity.
- **Use Case:** Customer segmentation or image compression.
- **Type:** Unsupervised learning.

9. Principal Component Analysis (PCA)

- **Description:** A dimensionality reduction technique that transforms data into a set of orthogonal (uncorrelated) features, ordered by variance.
- **Use Case:** Reducing features for better visualization or improving model performance.
- **Type:** Unsupervised learning.

10. Deep Neural Networks (DNN)

- **Description:** A type of artificial neural network with multiple layers that can model complex patterns in data.

- **Use Case:** Image recognition, natural language processing, and speech recognition.
- **Type:** Supervised learning (typically).

11. Convolutional Neural Networks (CNN)

- **Description:** A deep learning algorithm specialized in processing structured grid-like data, such as images or video.
- **Use Case:** Object detection, facial recognition, autonomous vehicles.
- **Type:** Supervised learning.

12. Recurrent Neural Networks (RNN)

- **Description:** A class of neural networks designed to handle sequential data by retaining memory of previous time steps.
- **Use Case:** Time series forecasting, language modeling, speech recognition.
- **Type:** Supervised learning.

13. Generative Adversarial Networks (GANs)

- **Description:** A framework where two neural networks, a generator and a discriminator, compete against each other to improve data generation.
- **Use Case:** Image generation, video generation, and data augmentation.
- **Type:** Unsupervised learning.

14. Q-Learning (Reinforcement Learning)

- **Description:** A reinforcement learning algorithm where an agent learns to take actions by receiving rewards or penalties from the environment.
- **Use Case:** Game AI, robotics, and autonomous navigation.
- **Type:** Reinforcement learning.

15. Deep Q-Networks (DQN)

- **Description:** An extension of Q-learning that integrates deep learning to approximate the Q-values for large, complex state spaces.
- **Use Case:** Video game AI, robotics.
- **Type:** Reinforcement learning.

1.7 AI REGULATION AND THE FUTURE OF AI

AI Regulation

- As artificial intelligence algorithms grow more complex and powerful, AI technologies — and the companies that create them — have increasingly drawn scrutiny from regulators across the world.

- In 2021, the European Union Parliament proposed a regulatory framework that aims to ensure AI systems deployed within the European Union are "safe, transparent, traceable, non-discriminatory and environmentally friendly." Under this framework, AI systems that can be used to perform real-time surveillance, or to manipulate people, categorize populations or discriminate against vulnerable groups, would be banned from use within the EU (though some limited exceptions may be made for law enforcement purposes).

- In 2022, the Biden White House introduced an AI Bill of Rights, outlining principles for responsible use of AI. And in 2023, the Biden-Harris administration introduced The Executive Order on Safe, Secure and Trustworthy AI, which aims to regulate the AI industry while maintaining the United States' status as a leader in artificial intelligence innovation.

- The order requires the companies operating large AI systems to perform safety testing and report results to the federal government before making their products publicly available. It also calls for labeling of AI-generated content and increased efforts to answer questions about the impact of AI on intellectual property rights. Additionally, the executive order calls for several worker protections including against unsafe AI implementation and harmful disruptions of the labor force. The order also calls for the United States

government to work alongside other countries to establish global standards for mitigating the risks of AI and promoting AI safety more generally.

Future of Artificial Intelligence

- In the near future, AI is poised to advance in machine learning capabilities and related frameworks like generative adversarial networks (GANs), which can help further develop generative AI and autonomous systems. Inevitably, AI will continue to make an impact across multiple industries, potentially causing job displacement, but also new job opportunities.

- Looking ahead, one of the next big steps for artificial intelligence is to progress beyond weak or narrow AI and achieve artificial general intelligence (AGI). With AGI, machines will be able to think, learn and act the same way as humans do, blurring the line between organic and machine intelligence. This could pave the way for increased automation and problem-solving capabilities in medicine, transportation and more — as well as sentient AI down the line.

- While likely groundbreaking, future advancements in AI have raised concerns such as heightened job loss, widespread disinformation, unpredictable AI behavior and possible moral dilemmas associated with reaching technological singularity.

- For now, society is largely looking toward federal and business-level AI regulations to help guide the technology’s future