Nepal College of Information Technology

Affiliated to Pokhara University

POKHARA UNIVESITY

NEPAL COLLEGE OF INFORMATION TECHNOLOGY

BALKUMARI,LALITPUR, NEPAL

ARTFICIAL INTELLIGENCE AND NEURAL NETWORK

By

Er. Rudra Nepal

NOVEMBER, 2024

CHAPTER ONE: INTRODUCTION

1.1   Introduction To Artificial Intelligence

Artificial Intelligence (AI) refers to the development of computer systems that are
capable of performing tasks that would typically require human intelligence. It
involves the creation of intelligent machines that can analyze, reason, learn, and
make decisions similar to humans. AI encompasses a wide range of techniques
and approaches, including machine learning, natural language processing, computer
vision, and robotics.

Machine learning plays a crucial role in AI, enabling systems to learn and improve
from experience without explicit programming. Natural language processing fo-
cuses on enabling computers to understand and communicate in human language,
while computer vision enables machines to interpret and analyze visual information.
Robotics, on the other hand, involves the design and development of physical AI
systems capable of interacting with the physical world.

With advancements in AI, machines can now perform complex tasks such as speech
recognition, image classification, recommendation systems, autonomous driving, and
even beat human experts in specific domains. The ultimate goal of AI is to develop
machines that possess general intelligence, meaning they can rival human intelligence
across various cognitive tasks. AI has the potential to revolutionize various industries
and improve our lives through automation, enhanced decision-making capabilities,
and the ability to tackle complex problems at scale.

But in 2024, most AI researchers and practitioners—and most AI-related head-
lines—are focused on breakthroughs in generative AI (gen AI), a technology that
can create original text, images, video and other content. To fully understand gen-
erative AI, it's important to first understand the technologies on which generative
AI tools are built: machine learning (ML) and deep learning.

A simple way to think about AI is as a series of nested or derivative concepts
that have emerged over more than 70 years: A diagram showing how artificial
intelligence, machine learning, deep learning and generative AI are related Directly
underneath AI, we have machine learning, which involves creating models by training
an algorithm to make predictions or decisions based on data. It encompasses a broad
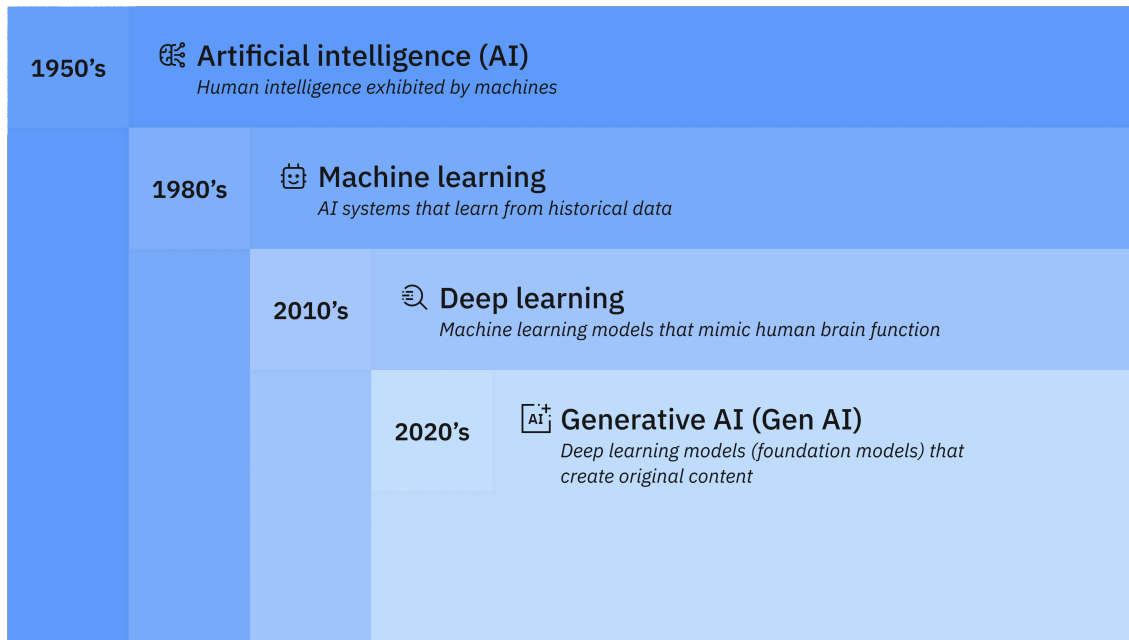
Figure 1.1: diagram-comparing-ai-ml-deep-learning-genai

range of techniques that enable computers to learn from and make inferences based on data without being explicitly programmed for specific tasks.

There are many types of machine learning techniques or algorithms, including linear regression, logistic regression, decision trees, random forest, support vector machines (SVMs), k-nearest neighbor (KNN), clustering and more. Each of these approaches is suited to different kinds of problems and data.

But one of the most popular types of machine learning algorithm is called a neural network (or artificial neural network). Neural networks are modeled after the human brain's structure and function. A neural network consists of interconnected layers of nodes (analogous to neurons) that work together to process and analyze complex data. Neural networks are well suited to tasks that involve identifying complex patterns and relationships in large amounts of data.

The simplest form of machine learning is called supervised learning, which involves the use of labeled data sets to train algorithms to classify data or predict outcomes accurately. In supervised learning, humans pair each training example with an output label. The goal is for the model to learn the mapping between inputs and outputs in the training data, so it can predict the labels of new, unseen data.

1.1.1 Intelligence

Intelligence is a complex concept that can be categorized in various ways. Here are some common types and definitions:

1. **Linguistic Intelligence:** This involves a strong ability in language, including reading, writing, and verbal communication. People with high linguistic intelligence are often good storytellers, poets, or effective speakers.

2. **Logical-Mathematical Intelligence:** This type focuses on logical reasoning, problem-solving, and mathematical skills. Individuals with high logical-mathematical intelligence excel in recognizing patterns, analyzing data, and solving mathematical problems.

3. **Spatial Intelligence:** This involves the ability to visualize and manipulate objects in space. People with strong spatial intelligence often excel in fields like architecture, engineering, and art, as they can think in three dimensions and create visual representations.

4. **Bodily-Kinesthetic Intelligence:** This type pertains to physical coordination and the ability to use one's body to express feelings or create products. Athletes, dancers, and surgeons typically exhibit high bodily-kinesthetic intelligence, as they can control their movements skillfully.

5. **Musical Intelligence:** This involves sensitivity to rhythm, tone, pitch, and sound. Individuals with high musical intelligence often excel in singing, playing instruments, or composing music. They can recognize musical patterns and appreciate different musical styles.

6. **Interpersonal Intelligence:** This type focuses on the ability to understand and interact effectively with others. People with high interpersonal intelligence are adept at recognizing others' feelings, motivations, and moods, making them great leaders, teachers, and collaborators.

7. **Intrapersonal Intelligence:** This involves self-awareness and the ability to understand one's own emotions, thoughts, and motivations. Individuals with high intrapersonal intelligence are reflective and often excel in self-directed learning, personal development, and introspection.

1.1.2 Components of Intelligence

1. **Cognition:** This refers to the mental processes involved in acquiring knowledge and understanding, including thinking, problem-solving, and decision-making.

2. **Memory:** Memory is the ability to store, retain, and retrieve information. It plays a crucial role in learning and applying knowledge.

3. **Perception:** Perception is how we interpret sensory information from our environment. It shapes our understanding and responses to the world around us.

4. **Language:** Language is essential for communication and is closely tied to thought. It allows us to express ideas, share knowledge, and connect with others.

5. **Reasoning:** This involves the ability to think logically, draw conclusions, and solve problems. It can be both inductive (generalizing from specific instances) and deductive (applying general principles to specific cases).

6. **Creativity:** Creativity is the ability to generate new ideas, approaches, or solutions. It involves thinking outside the box and connecting seemingly unrelated concepts.

7. **Emotional Intelligence:** This refers to the ability to recognize, understand, and manage our own emotions, as well as the emotions of others. It plays a critical role in interpersonal relationships and social interactions.

8. **Social Intelligence:** Social intelligence is the capacity to navigate social situations and understand social dynamics. It involves empathy, communication skills, and the ability to build relationships.

9. **Adaptability:** Adaptability is the ability to adjust to new conditions and respond effectively to change. It reflects flexibility in thinking and behavior.

10. **Practical Knowledge:** This encompasses the skills and know-how to deal with everyday tasks and challenges. It involves applying knowledge to real-world situations effectively.

1.2 Approaches to AI

Thinking Humanly	Thinking Rationally
"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense." (Haugeland, 1985)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)
"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)	"The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998)
"The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

1.2.1 Acting Humanly: Turing Test Approach

The "acting humanly" approach to AI, often referred to as the Turing Test approach, is centered around creating machines that can simulate human-like behavior and thought processes. This approach was proposed by Alan Turing in 19501.

Key Points of the Turing Test Approach:

Objective: To determine if a machine can exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.
Method: Involves an interrogator communicating with both a human and a machine without knowing which is which. If the interrogator cannot reliably distinguish the machine from the human, the machine is said to have passed the test.
Significance: This test provides an operational definition of intelligence, focusing on the machine's ability to perform human-like cognitive tasks.

This approach emphasizes the importance of natural language processing(Ability to communicate successfully), knowledge representation(Store what it knows or hears), automated reasoning(Use the stored information to answer questions and draw new conclusions), and machine learning(Adapt to new circumstances and to detect and

extraploate patterns) to create systems that can mimic human interactions convincingly.

1.2.2 Thinking humanly: The cognitive modeling approach

The exciting new effort to make computers think...machines with minds, in the full and literal sense.(Haugeland, 1985)

- "The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning..."(Bellman, 1978)

Once we gather enough data, we can create a model to simulate the human process. This model can be used to create software that can think like humans. Of course this is easier said than done! All we care about is the output of the program given a particular input. If the program behaves in a way that matches human behavior, then we can say that humans have a similar thinking mechanism.

Within computer science, there is a field of study called Cognitive Modeling that deals with simulating the human thinking process. It tries to understand how humans solve problems. It takes the mental processes that go into this problem solving process and turns it into a software model. This model can then be used to simulate human behavior. Cognitive modeling is used in a variety of AI applications such as deep learning, expert systems, Natural Language Processing, robotics, and so on.

The interdisciplinary field of cognitive science brings together computer models from AI and experimental techniques from psychology to construct precise and testable theories of the human mind.

1.2.3 Thinking Rationally: The "Laws of Thought" Approach

Rationality refers to doing the right thing in a given circumstance.

The Greek philosopher Aristotle was one of the first to attempt to codify "right thinking," that is, irrefutable reasoning processes. His syllogisms provided patterns for argument structures that always yielded correct conclusions when given correct premises — for example, "Socrates is a man; all men are mortal; therefore, Socrates is mortal." These laws of thought were supposed to govern the operation of the mind; their study initiated the field called logic.

There are two main obstacles to this approach:

1. It is not easy to take informal knowledge and state it in the formal terms required by logical notation, particularly when the knowledge is less than 100% certain.

2. There is a big difference between solving a problem "in principle" and solving it in practice. Even problems with just a few hundred facts can exhaust the computational resources of any computer unless it has some guidance as to which reasoning steps to try first.

1.2.4   Acting Rationally: The Rational Agent Approach

An agent is just something that acts (agent comes from the Latin *agere*, to do). Of course, all computer programs do something, but computer agents are expected to do more: operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals.

Rational agents need to perform in such a way that there is maximum benefit to the entity performing the action. An agent is said to act rationally if, given a set of rules, it takes actions to achieve its goals. It perceives and acts according to the information that's available. This system is widely used in AI to design robots when they are sent to navigate unknown terrains.

All the skills needed for the Turing Test also allow an agent to act rationally. Knowledge representation and reasoning enable agents to reach good decisions. We need to be able to generate comprehensible sentences in natural language to get by in a complex society.

The rational-agent approach has two advantages over the other approaches:

1. It is more general than the "laws of thought" approach because correct inference is just one of several possible mechanisms for achieving rationality.

2. It is more amenable to scientific development than are approaches based on human behavior or human thought.

1.3 Foundations of AI
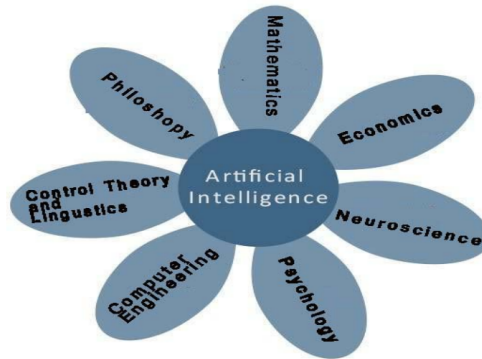
The foundation of AI is multidisciplinary, drawing from a variety of fields, each contributing unique insights and methodologies. Here's an overview of how each of the specified fields plays a role:

1. **Philosophy**: Explores fundamental questions about intelligence, consciousness, and ethics. It helps frame discussions about the implications of AI, including moral responsibility and the nature of decision-making.

2. **Mathematics**: Provides the theoretical underpinnings for algorithms and models. Key areas include linear algebra for data representation, calculus for optimization, and graph theory for network analysis.

3. **Statistics**: Essential for making inferences from data and understanding uncertainty. It forms the backbone of many machine learning techniques, enabling the analysis of patterns and predictions.

4. **Economics**: Offers insights into decision-making processes, game theory, and market dynamics, which can inform the development of AI systems for optimization and resource allocation.

5. **Neuroscience**: Informs the design of neural networks by mimicking the structure and function of the human brain. Understanding brain processes can lead to better algorithms for learning and adaptation.

6. **Psychology**: Provides insights into human behavior and cognition, which can inform user-centered design and improve AI interactions. Understanding cognitive biases can also help in developing fair AI systems.

7. **Computer Engineering**: Involves the design and implementation of hardware and software systems. It plays a crucial role in creating efficient AI architectures and optimizing performance.

8. **Control Theory**: Focuses on the behavior of dynamical systems. It's important for developing AI in robotics and automation, helping to design systems that can operate and adapt in real-time environments.

9. **Linguistics**: Essential for natural language processing (NLP), linguistics helps AI systems understand, interpret, and generate human language, bridging communication between humans and machines.

Together, these fields contribute to the robust and evolving landscape of AI, shaping its development, application, and impact on society.

1.4 History of AI

The idea of "a machine that thinks" dates back to ancient Greece. However, since the advent of electronic computing, several important events and milestones have marked the evolution of artificial intelligence (AI).

Key Milestones

- **1950**: Alan Turing publishes *Computing Machinery and Intelligence*. In this paper, Turing, famous for breaking the German ENIGMA code during WWII, asks, "Can machines think?" He introduces the now-famous *Turing Test*, where a human interrogator attempts to distinguish between computer and human text responses. Despite scrutiny, this test remains a significant concept in AI and philosophy, particularly in linguistics.

- **1956**: John McCarthy coins the term "artificial intelligence" at the first-ever AI conference at Dartmouth College. McCarthy later invents the Lisp programming language. That same year, Allen Newell, J.C. Shaw, and Herbert Simon create the Logic Theorist, the first running AI computer program.

- **1967**: Frank Rosenblatt builds the Mark 1 Perceptron, the first computer based on a neural network that learns through trial and error. The following year, Marvin Minsky and Seymour Papert publish *Perceptrons*, which becomes a landmark work on neural networks, but also argues against future neural network research initiatives.

- **1980**: Neural networks, which utilize a backpropagation algorithm for self-training, become widely adopted in AI applications.

- **1995**: Stuart Russell and Peter Norvig publish *Artificial Intelligence: A Modern Approach*, a leading textbook in AI. The book explores four potential goals or definitions of AI, differentiating computer systems based on rationality and thinking versus acting.

- **1997**: IBM's Deep Blue defeats world chess champion Garry Kasparov in a chess match and a subsequent rematch.

- **2004**: John McCarthy writes *What Is Artificial Intelligence?*, proposing a widely cited definition of AI. By this time, the era of big data and cloud computing begins, enabling organizations to manage larger data estates that will be used to train AI models.

- **2011**: IBM Watson® defeats champions Ken Jennings and Brad Rutter on *Jeopardy!*. Around this time, data science emerges as a popular discipline.

- **2015**: Baidu's Minwa supercomputer employs a convolutional neural network to identify and categorize images with higher accuracy than the average human.

- **2016**: DeepMind's AlphaGo, powered by a deep neural network, defeats world champion Go player Lee Sedol in a five-game match, a notable achievement given the game's complexity. Google later purchases DeepMind for approximately USD 400 million.

- **2022**: A rise in large language models (LLMs), such as OpenAI's ChatGPT, leads to significant advancements in AI performance and its potential to create enterprise value. These generative AI models are pretrained on vast amounts of data.

- **2024**: The latest AI trends indicate a continuing renaissance, with multimodal models that can process multiple types of input data, providing richer, more robust experiences. These models integrate computer vision and natural language processing capabilities. Smaller models are also making progress, countering the diminishing returns observed with massive models.

1.5 Benefits of AI

AI offers numerous benefits across various industries and applications. Some of the most commonly cited benefits include:

- Automation of repetitive tasks

- More and faster insight from data

- Enhanced decision-making

- Fewer human errors

- 24/7 availability

- Reduced physical risks

Automation of Repetitive Tasks AI can automate routine, repetitive, and often tedious tasks—including digital tasks such as data collection, entry, preprocessing, and physical tasks such as warehouse stock-picking and manufacturing processes. This automation frees humans to work on higher-value, more creative tasks.

Enhanced Decision-Making Whether used for decision support or for fully automated decision-making, AI enables faster, more accurate predictions and reliable, data-driven decisions. Combined with automation, AI allows businesses to act on opportunities and respond to crises in real-time and without human intervention.

Fewer Human Errors AI can reduce human errors in various ways, from guiding people through the proper steps of a process, to flagging potential errors before they

occur, and fully automating processes without human intervention. This is especially important in industries such as healthcare, where AI-guided surgical robotics enable consistent precision.

Machine learning algorithms can continually improve their accuracy and further reduce errors as they're exposed to more data and "learn" from experience.

Round-the-Clock Availability and Consistency AI is always on, available around the clock, and delivers consistent performance every time. Tools such as AI chatbots or virtual assistants can alleviate staffing demands for customer service or support. In other applications—such as materials processing or production lines—AI helps maintain consistent work quality and output levels when completing repetitive or tedious tasks.

Reduced Physical Risk By automating dangerous work—such as animal control, handling explosives, performing tasks in deep ocean water, high altitudes, or outer space—AI can eliminate the need to put human workers at risk of injury or worse. While still being perfected, self-driving cars and other vehicles offer the potential to reduce the risk of injury to passengers.

1.5.1 AI challenges and risks

Organizations are scrambling to take advantage of the latest AI technologies and capitalize on AI's many benefits. This rapid adoption is necessary, but implementing and maintaining AI workflows comes with challenges and risks.

Data Risks AI systems rely on datasets that might be vulnerable to data poisoning, data tampering, data bias, or cyberattacks that can lead to data breaches. Organizations can mitigate these risks by protecting data integrity and implementing security and availability throughout the entire AI lifecycle, from development to training, deployment, and post-deployment.

Model Risks Threat actors can target AI models for theft, reverse engineering, or unauthorized manipulation. Attackers might compromise a model's integrity by tampering with its architecture, weights, or parameters—the core components that determine a model's behavior, accuracy, and performance.

Operational Risks Like all technologies, AI models are susceptible to operational risks such as model drift, bias, and breakdowns in the governance structure. Left

unaddressed, these risks can lead to system failures and cybersecurity vulnerabilities that threat actors can exploit.

Ethics and Legal Risks If organizations don't prioritize safety and ethics when developing and deploying AI systems, they risk committing privacy violations and producing biased outcomes. For example, biased training data used for hiring decisions might reinforce gender or racial stereotypes and create AI models that favor certain demographic groups over others. AI use cases

1.5.2 AI Usecases

The real-world applications of AI are many. Here is just a small sampling of use cases across various industries to illustrate its potential:

Customer experience, service and support

Companies can implement AI-powered chatbots and virtual assistants to handle customer inquiries, support tickets and more. These tools use natural language processing (NLP) and generative AI capabilities to understand and respond to customer questions about order status, product details and return policies.

Chatbots and virtual assistants enable always-on support, provide faster answers to frequently asked questions (FAQs), free human agents to focus on higher-level tasks, and give customers faster, more consistent service.

Fraud detection

Machine learning and deep learning algorithms can analyze transaction patterns and flag anomalies, such as unusual spending or login locations, that indicate fraudulent transactions. This enables organizations to respond more quickly to potential fraud and limit its impact, giving themselves and customers greater peace of mind.

Personalized marketing

Retailers, banks and other customer-facing companies can use AI to create personalized customer experiences and marketing campaigns that delight customers, improve sales and prevent churn. Based on data from customer purchase history and behaviors, deep learning algorithms can recommend products and services customers are likely to want, and even generate personalized copy and special offers for individual customers in real time.

Human resources and recruitment

AI-driven recruitment platforms can streamline hiring by screening resumes, matching candidates with job descriptions, and even conducting preliminary interviews using video analysis. These and other tools can dramatically reduce the mountain of administrative paperwork associated with fielding a large volume of candidates. It can also reduce response times and time-to-hire, improving the experience for candidates whether they get the job or not.

Application development and modernization

Generative AI code generation tools and automation tools can streamline repetitive coding tasks associated with application development, and accelerate the migration and modernization (reformatting and replatorming) of legacy applications at scale. These tools can speed up tasks, help ensure code consistency and reduce errors.

Predictive maintenance

Machine learning models can analyze data from sensors, Internet of Things (IoT) devices and operational technology (OT) to forecast when maintenance will be required and predict equipment failures before they occur. AI-powered preventive maintenance helps prevent downtime and enables you to stay ahead of supply chain issues before they affect the bottom line.

1.6   AI ethics and governance

AI ethics is a multidisciplinary field that studies how to optimize AI's beneficial impact while reducing risks and adverse outcomes. Principles of AI ethics are applied through a system of AI governance, consisting of guardrails that help ensure that AI tools and systems remain safe and ethical.

AI governance encompasses oversight mechanisms that address risks. An ethical approach to AI governance requires the involvement of a wide range of stakeholders, including developers, users, policymakers, and ethicists, helping to ensure that AI-related systems are developed and used to align with society's values.

Here are common values associated with AI ethics and responsible AI:

Explainability and Interpretability

As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm arrived at a result. Explainable AI is a set of processes and methods that enables human users to interpret, comprehend, and trust the results and outputs created by algorithms.

Fairness and Inclusion

Although machine learning, by its very nature, is a form of statistical discrimination, the discrimination becomes objectionable when it places privileged groups at a systematic advantage and certain unprivileged groups at a systematic disadvantage, potentially causing varied harms. To encourage fairness, practitioners can try to minimize algorithmic bias across data collection and model design, and build more diverse and inclusive teams.

Robustness and Security

Robust AI effectively handles exceptional conditions, such as abnormalities in input or malicious attacks, without causing unintentional harm. It is also built to withstand intentional and unintentional interference by protecting against exposed vulnerabilities.

Accountability and Transparency

Organizations should implement clear responsibilities and governance structures for the development, deployment, and outcomes of AI systems. In addition, users should be able to see how an AI service works, evaluate its functionality, and comprehend its strengths and limitations. Increased transparency provides information for AI consumers to better understand how the AI model or service was created.

Privacy and Compliance

Many regulatory frameworks, including GDPR, mandate that organizations abide by certain privacy principles when processing personal information. It is crucial to protect AI models that might contain personal information, control what data goes into the model, and build adaptable systems that can adjust to changes in regulation and attitudes around AI ethics.

1.7 AI and Society: Work, Automation, Employment, Privacy, and Security

The intersection of AI and society touches on several key themes, including work, automation, employment, privacy, and security. This document provides an overview of each aspect.

Work and Automation

- **Job Displacement**: Automation through AI can lead to significant job losses in sectors like manufacturing, retail, and even some professional services. Routine and repetitive tasks are the most vulnerable.

- **Job Creation**: While some jobs are lost, new roles may emerge in AI development, maintenance, and oversight. There will be a growing need for skills in data analysis, machine learning, and AI ethics.

- **Changing Skill Requirements**: Workers may need to adapt to new technologies and acquire skills that complement AI, such as critical thinking, creativity, and emotional intelligence.

Employment

- **Gig Economy**: AI can facilitate gig work, providing flexible employment opportunities but often without the benefits of traditional employment (like health insurance and job security).

- **Remote Work**: AI tools support remote collaboration, enabling more flexible work arrangements but also blurring the lines between work and personal life.

- **Inequality**: There's a risk that the benefits of AI will disproportionately favor those with access to technology and education, potentially widening socioeconomic gaps.

Privacy

- **Data Collection**: AI systems often rely on large datasets, which can include personal information. Concerns arise about how this data is collected, stored, and used.

- **Surveillance**: AI technologies are increasingly used for surveillance, raising ethical questions about consent and the potential for abuse by governments or corporations.

- **Regulation**: There's a growing call for regulations to protect individuals' privacy, such as GDPR in Europe, which aims to give people control over their personal data.

Security

- **Cybersecurity Risks**: AI can enhance cybersecurity by detecting threats, but it can also be used by malicious actors to conduct sophisticated cyber attacks.

- **Bias and Fairness**: AI systems can perpetuate or even exacerbate existing biases, leading to unfair treatment in areas like hiring, law enforcement, and lending.

- **Ethical AI**: There's a push for developing ethical guidelines to ensure that AI technologies are designed and implemented responsibly, prioritizing safety and fairness.

Conclusion Navigating the challenges posed by AI requires a collaborative approach among governments, businesses, and civil society. Fostering public discourse, implementing effective regulations, and investing in education and training are crucial steps to harness the benefits of AI while mitigating its risks.