

## Artificial Intelligence Answer Sheet - Tanaya (Student 36)

Question 1: Explain the differences between supervised, unsupervised, and reinforcement learning in machine learning. Provide examples of applications for each approach.

Answer:

Supervised learning is a paradigm where algorithms are trained on datasets that contain both inputs and corresponding outputs. The model learns to associate inputs with outputs, making it well-suited for tasks where historical data with known results is available. This approach is foundational to applications like email spam detection, medical diagnosis, credit scoring, and sentiment analysis. Supervised learning is commonly divided into classification (predicting categories, e.g., tumor detection in scans) and regression (predicting continuous values, e.g., house price estimation).

Unsupervised learning, in contrast, works with datasets that lack explicit output labels. The algorithm must autonomously uncover hidden patterns, groupings, or structures within the data. Techniques such as clustering (e.g., customer segmentation for marketing) and dimensionality reduction (e.g., visualizing complex data) are central to this approach. Unsupervised learning is valuable in anomaly detection for cybersecurity, recommendation systems, and market basket analysis in retail.

Reinforcement learning is a distinct paradigm where an agent interacts with an environment, learning by receiving feedback in the form of rewards or penalties. The agent's objective is to maximize cumulative rewards over time, often through trial and error. This approach is highly effective in sequential decision-making scenarios such as game-playing (e.g., AlphaGo), autonomous vehicle navigation, robotic manipulation, and algorithmic trading.

Question 2: Describe the architecture and functioning of Convolutional Neural Networks (CNNs) and explain why they are particularly effective for image recognition tasks.

Answer:

Convolutional Neural Networks (CNNs) are deep learning models specifically designed for processing grid-like data, such as images. Their architecture consists of convolutional layers that apply learnable filters to the input, extracting features like edges and textures. Pooling layers follow, reducing the spatial dimensions of the data while preserving essential information, which improves computational efficiency and provides translation invariance. Activation functions, such as ReLU, introduce non-linearity, and fully connected layers at the end perform classification or regression.

CNNs excel at image recognition because their local connectivity and parameter sharing mirror the spatial structure of images. Early layers detect simple features, while deeper layers combine these to recognize complex patterns and objects. The hierarchical feature extraction in CNNs is similar to how the human visual cortex processes visual information. Pooling operations and the progressive abstraction of features make CNNs robust to variations in position, scale, and orientation, enabling

them to outperform traditional neural networks in tasks like image classification, object detection, and facial recognition.

Question 3: Discuss the ethical considerations and potential societal impacts of implementing artificial intelligence systems in critical decision-making processes.

Answer:

The integration of AI into critical decision-making brings about significant ethical and societal challenges. Algorithmic bias is perhaps the most pressing concern, as AI systems trained on historical data can perpetuate or even amplify existing prejudices, leading to unfair outcomes in areas such as hiring, lending, healthcare, and criminal justice. For example, facial recognition systems have shown higher error rates for women and people of color, raising serious questions about fairness and discrimination.

Transparency and explainability are also crucial. Many advanced AI models, especially deep neural networks, operate as "black boxes," making it difficult to understand or explain their decisions. This lack of interpretability can undermine trust, particularly in high-stakes domains like healthcare, where understanding the rationale behind recommendations is essential for safety and accountability.

Privacy concerns are significant as well, since AI systems often require access to large volumes of personal data, raising questions about consent, data security, and surveillance. Accountability is also complex, as responsibility for AI-driven decisions may be distributed among data collectors, developers, system integrators, and end-users, making it difficult to assign liability when harm occurs.

Additionally, the widespread adoption of AI can disrupt labor markets, potentially leading to job displacement and increased economic inequality if the benefits of automation are not broadly shared. Addressing these challenges requires a combination of technical solutions (such as fairness-aware algorithms and explainable AI), regulatory oversight, and ongoing stakeholder engagement to ensure AI systems are ethical, transparent, and aligned with societal values.

Question 4: Explain the concept of transfer learning in deep neural networks and discuss its advantages and limitations.

Answer:

Transfer learning is a machine learning technique where a model trained on one task is adapted for a different but related task. In deep learning, this often involves taking a neural network pre-trained on a large dataset (such as ImageNet for images) and fine-tuning it for a new task with less data. Typically, the final layers of the pre-trained model are replaced with new layers suited to the target task, and the network is retrained—sometimes only the new layers, sometimes the entire model with a lower learning rate.

The main advantage of transfer learning is that it leverages knowledge from large, diverse datasets, making it possible to train effective models even when the target

dataset is small. This reduces the need for extensive labeled data, speeds up training, and often leads to better performance by providing a strong starting point for learning. Transfer learning also helps prevent overfitting by transferring general feature representations learned from the source task.

However, transfer learning works best when the source and target tasks are similar; if the domains are too different, the transferred knowledge may not be useful or could even hinder performance (a phenomenon known as negative transfer). Additionally, pre-trained models may carry over biases from their original training data, and their architectures may not always be optimal for the new task.

Question 5: Describe the principles of natural language processing (NLP) and how transformer-based models like BERT have revolutionized language understanding tasks.

Answer:

Natural Language Processing (NLP) is the field focused on enabling computers to analyze, interpret, and generate human language. Traditional NLP approaches relied on rule-based systems and statistical models, which often struggled to capture the complexity and context of language.

Transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have revolutionized NLP. Transformers use self-attention mechanisms, allowing the model to consider relationships between all words in a sentence simultaneously, rather than processing them sequentially. BERT introduced bidirectional context, enabling the model to understand the meaning of a word based on both its left and right surroundings.

These models are pre-trained on massive text corpora using tasks like masked language modeling and next sentence prediction, then fine-tuned for specific applications such as sentiment analysis, question answering, and text classification. This approach allows for the development of rich, contextual word representations that capture both semantic and syntactic nuances. As a result, transformer-based models have set new standards for NLP performance, enabling more accurate and natural human-computer interactions. Despite challenges such as high computational requirements and potential biases, transformers have become the backbone of modern NLP systems.