Name: _____ **Date:** _____

Answer Key: Silicon Synapse Expedition: 9th Grade Neural Architecture Quiz

Deconstruct backpropagation, ethical alignment, and gradient descent through high-level synthesis of modern deep learning frameworks.

---

1. In the context of training a Generative Adversarial Network (GAN), what occurs during the 'minimax' game between the Generator and the Discriminator?

Answer: B) The Generator tries to maximize the probability of the Discriminator making a mistake.

In a GAN, the Generator aims to fool the Discriminator into classifying synthetic data as real, creating a competitive training environment.

2. The process of _____ involves an AI model adjusting its internal weights based on the error rate of its previous output to improve accuracy.

Answer: B) Backpropagation

Backpropagation is the fundamental algorithm used in neural networks to calculate gradients and update weights through the chain rule.

3. Underfitting occurs when a machine learning model is so complex that it captures the 'noise' in the training data rather than the underlying pattern.

Answer: B) False

The description provided actually refers to 'Overfitting.' Underfitting happens when a model is too simple to capture the underlying structure of the data.

4. How does the 'Attention Mechanism' in a Transformer architecture function differently than a standard Recurrent Neural Network (RNN)?

Answer: C) It allows the model to weigh the importance of different parts of the input data regardless of distance.

Self-attention enables Transformers to relate every word in a sentence to every other word, overcoming the 'vanishing gradient' problem in long RNN sequences.

5. A concept known as _____ refers to the lack of transparency in how deep learning models make decisions, making them difficult for humans to interpret.

Name: _____ **Date:** _____

Answer: C) Black Box

The 'Black Box' problem is a major ethical concern in AI, as it is often impossible to trace the exact reasoning behind a specific output.

6. Which of these scenarios best illustrates 'Reinforcement Learning from Human Feedback' (RLHF)?

Answer: B) A human ranking multiple AI responses to help the model align with human preferences.

RLHF uses human ranking to fine-tune Large Language Models, ensuring they produce helpful and safe content.

7. Edge AI refers to processing artificial intelligence algorithms locally on a device rather than in a centralized cloud server.

Answer: A) True

Edge AI reduces latency and increases privacy by keeping data processing on the user's hardware (like a smartphone or IoT device).

8. In AI ethics, the 'Alignment Problem' specifically focuses on which of the following challenges?

Answer: C) Ensuring the AI's goals and behaviors match human values and intentions.

Alignment is the challenge of ensuring autonomous systems act in ways that are beneficial to humans without unintended consequences.

9. A _____ is a specific type of neural network layer that uses mathematical filters to identify spatial hierachies in visual data, like edges and shapes.

Answer: A) Convolutional Layer

Convolutional Neural Networks (CNNs) use these layers to 'scan' images, making them the standard for computer vision tasks.

10. Transfer Learning allows an AI to apply knowledge gained from solving one problem to a different but related problem.

Answer: A) True

Transfer learning is efficient because it prevents models from having to 're-learn' fundamental concepts (like shapes) when starting a new task.