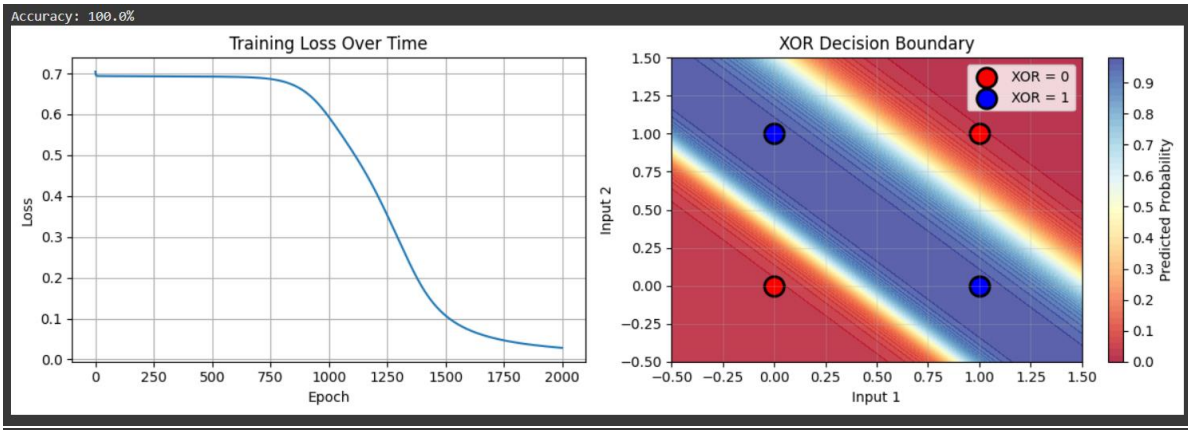
```
=== XOR PROBLEM ===
XOR Truth Table:
Input1 | Input2 | Output
   0   |   0    |   0
   0   |   1    |   1
   1   |   0    |   1
   1   |   1    |   0

Training Data:
Input: [0 0] -> Expected Output: 0
Input: [0 1] -> Expected Output: 1
Input: [1 0] -> Expected Output: 1
Input: [1 1] -> Expected Output: 0

=== TRAINING NEURAL NETWORK ===
Initial Predictions (before training):
Input: [0 0] -> Prediction: 0.4467
Input: [0 1] -> Prediction: 0.4303
Input: [1 0] -> Prediction: 0.4270
Input: [1 1] -> Prediction: 0.4126
```

```
Epoch 0, Loss: 0.704555
Epoch 100, Loss: 0.693436
Epoch 200, Loss: 0.693192
Epoch 300, Loss: 0.693012
Epoch 400, Loss: 0.692809
Epoch 500, Loss: 0.692471
Epoch 600, Loss: 0.691700
Epoch 700, Loss: 0.689375
Epoch 800, Loss: 0.681078
Epoch 900, Loss: 0.654608
Epoch 1000, Loss: 0.593856
Epoch 1100, Loss: 0.510578
Epoch 1200, Loss: 0.411963
Epoch 1300, Loss: 0.292354
Epoch 1400, Loss: 0.177371
Epoch 1500, Loss: 0.107591
Epoch 1600, Loss: 0.072489
Epoch 1700, Loss: 0.053241
Epoch 1800, Loss: 0.041513
Epoch 1900, Loss: 0.033758

=== RESULTS AFTER TRAINING ===
Final Predictions:
Input: [0 0] -> Probability: 0.0335 -> Binary: 0 -> Expected: 0 ✓
Input: [0 1] -> Probability: 0.9743 -> Binary: 1 -> Expected: 1 ✓
Input: [1 0] -> Probability: 0.9742 -> Binary: 1 -> Expected: 1 ✓
Input: [1 1] -> Probability: 0.0266 -> Binary: 0 -> Expected: 0 ✓
```

```
Accuracy: 100.0%
```




```
=== KEY CONCEPTS DEMONSTRATED ===
1. PERCEPTRON LIMITATION: A single layer cannot solve XOR (not linearly separable)
2. HIDDEN LAYERS: Added non-linearity allows the network to solve XOR
3. FORWARD PROPAGATION: Data flows from input -> hidden -> output
4. ACTIVATION FUNCTIONS: Sigmoid adds non-linearity (essential for XOR)
5. BACKPROPAGATION: Errors propagate backward to update weights
6. GRADIENT DESCENT: Weights updated to minimize loss
7. LEARNING RATE: Controls how big steps the network takes while learning
8. BIAS TERMS: Additional parameters that help the network fit better

=== NETWORK ARCHITECTURE ===
Input Layer: 2 neurons (for 2 inputs)
Hidden Layer: 4 neurons with sigmoid activation
Output Layer: 1 neuron with sigmoid activation
Total Parameters: 17
  - W1: (2, 4) = 8 weights
  - b1: (1, 4) = 4 biases
  - W2: (4, 1) = 4 weights
  - b2: (1, 1) = 1 biases
```