PART 1 Theoretical Understanding

Q1: Explain the primary differences between TensorFlow and PyTorch. When would you choose one over the other?

1. TensorFlow is developed by Google and emphasizes deployment at scale (e.g., TensorFlow Lite, TensorFlow Serving), while PyTorch, developed by Facebook, is more popular for research due to its dynamic computation graph (eager execution).
2. Choose PyTorch for quick prototyping and research (more Pythonic and intuitive).
3. Choose TensorFlow for production-ready models and deployment on mobile, web, or large-scale cloud systems.

Q2: Describe two use cases for Jupyter Notebooks in AI development.

1. Experimentation & Prototyping: Jupyter Notebooks allow step-by-step testing of AI models, which is ideal for trying out preprocessing steps, tuning hyperparameters, or visualizing data.
2. Educational Demos & Tutorials: Instructors and students use Jupyter to explain concepts interactively, combining code, visualizations, and explanations in one place.

Q3: How does spaCy enhance NLP tasks compared to basic Python string operations?

1. spaCy provides advanced linguistic features such as tokenization, part-of-speech tagging, named entity recognition, and dependency parsing, which are not possible with basic string operations like .split() or .replace().

1. It uses optimized pipelines and pre-trained models for faster, more accurate NLP analysis compared to manual methods.

🔻 2. Comparative Analysis: Scikit-learn vs TensorFlow

Feature	Scikit-learn	TensorFlow
Target Applications	Classical ML (e.g., regression, SVM, KNN)	Deep Learning (e.g., CNNs, RNNs, transformers)
Ease of Use	Beginner-friendly, simple API	Steeper learning curve, especially for custom models

Community Support	Large and mature community; great for ML	Very active and growing community; great for deep learning and deployment

1. Ethical Considerations – Bias in Models

▪ MNIST Bias:

- MNIST is grayscale and well-balanced but lacks diversity in handwriting styles (e.g., age, culture, disability).
- Risk: May underperform on digits written by left-handed or disabled users.

Mitigation:

- Augment dataset with rotated/skewed writing.
- Use TensorFlow Fairness Indicators to measure disparity in predictions across different groups (e.g., age-simulated styles).

▪ Amazon Reviews Bias:

- Sentiment analysis may reflect language or gender bias, especially if reviews include slang or cultural terms.
- spaCy's rule-based approach reduces some risks by keeping logic transparent.

Mitigation:

- Expand keyword lists to include diverse expressions.
- Use spaCy's SpanCategorizer with balanced, labeled training data for fairness-aware sentiment.

3. Troubleshooting Challenge – Debugging Buggy TensorFlow Code

Bug Problem

```
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

model.fit(x_train, y_train, epochs=5)
```

The MNIST labels (`y_train`) are integers (0–9), not one-hot encoded. So:

```
#How to fix
#Use sparse loss for integer labels
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
```

Another Bug

```
x_train = x_train.reshape(28, 28, 1)  # wrong: removes batch
dimension
```

how to fix

```
x_train = x_train.reshape(-1, 28, 28, 1)  # correct: adds batch dimension
```

Screenshots

```
Epoch 1/5
844/844 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 42s 48ms/step - accuracy: 0.8741 -
loss: 0.4320 - val_accuracy: 0.9835 - val_loss: 0.0551
Epoch 2/5
844/844 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 39s 46ms/step - accuracy: 0.9820 -
loss: 0.0592 - val_accuracy: 0.9858 - val_loss: 0.0511
Epoch 3/5
844/844 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 41s 46ms/step - accuracy: 0.9888 -
loss: 0.0365 - val_accuracy: 0.9882 - val_loss: 0.0415
Epoch 4/5
844/844 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 39s 46ms/step - accuracy: 0.9911 -
loss: 0.0303 - val_accuracy: 0.9900 - val_loss: 0.0378
Epoch 5/5
844/844 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 38s 45ms/step - accuracy: 0.9928 -
loss: 0.0217 - val_accuracy: 0.9907 - val_loss: 0.0387
<keras.src.callbacks.history.History at 0x7a077ad2bd50>
```