Part 3: Ethics & Optimization (10%)

1. Ethical Considerations

Potential Biases in Models:

MNIST Model Biases:

- **Demographic bias**: Handwriting styles vary across cultures and age groups
- **Data collection bias**: MNIST primarily contains Western-style digit writing
- **Representation bias**: Limited diversity in writing styles

Amazon Reviews Model Biases:

- **Selection bias**: Reviews may not represent all customer demographics
- **Language bias**: Models trained on English may not work well for other languages
- **Temporal bias**: Product sentiment may change over time

Mitigation Strategies:

Using TensorFlow Fairness Indicators:

```
# Example fairness evaluation
from tensorflow_model_analysis import fairness_indicators

# Evaluate model fairness across different groups
fairness_eval = fairness_indicators.FairnessIndicators(
    eval_shared_model=model,
    slicing_specs=[...],  # Define demographic slices
    example_weight_key='example_weight'
)
```

Using spaCy's Rule-Based Systems:

- Implement custom rules for different cultural contexts
- Use multiple pre-trained models for different languages
- Regular auditing of entity recognition accuracy across groups

2. Troubleshooting Challenge

Common TensorFlow errors and fixes:

- **Dimension mismatches**: Check input shapes and reshape data appropriately
- **Incorrect loss functions**: Match loss function to problem type (categorical vs sparse)
- **Learning rate issues**: Adjust optimizer parameters
- **Overfitting**: Add dropout layers and regularization

Evaluation Criteria

Component	Weight	Key Evaluation Points
Theoretical	40%	Accuracy, depth of understanding, clear explanations
Practical	50%	Code quality, model performance, documentation
Ethics	10%	Critical thinking, bias identification, solutions
Bonus	10%	Deployment success, user interface, functionality

Success Metrics:

- **Iris Classification**: >90% accuracy
- **MNIST CNN**: >95% test accuracy
- **NLP Analysis**: Proper entity extraction and sentiment analysis
- **Code Quality**: Well-commented, reproducible, error-free