

Part 3: Ethics & Optimization

1. Ethical Considerations

Potential Biases:

MNIST Model: May include data representation bias — primarily Western handwriting styles, reducing accuracy for diverse handwriting forms. **Amazon Reviews Model:** Rule-based sentiment may misclassify sarcasm or cultural expressions, introducing linguistic bias. **Iris Model:** Sampling bias from limited and balanced data may reduce generalizability to real-world flower variations.

Mitigating Bias with Fairness Tools:

TensorFlow Fairness Indicators: Used to measure subgroup disparities such as per-class accuracy or recall differences in MNIST digits. **spaCy Rule-Based Systems:** Custom lexical and pattern-based rules reduce unfair sentiment or NER misclassification. **Ethical Best Practices:** Ensure dataset transparency and proper documentation. Test fairness across demographic and contextual groups. Include human validation for sensitive or impactful outputs. Prefer interpretable models when fairness is critical.

2. Troubleshooting Challenge

Buggy Code:

```
model = tf.keras.Sequential([
    tf.keras.layers.Conv2D(32, (3,3), activation='relu', input_shape=(28,28)), # Missing channel
    dimension
    tf.keras.layers.MaxPooling2D(2,2),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(10, activation='relu'), # Wrong output activation
    tf.keras.layers.Dense(1, activation='softmax') # Wrong output dimension
])
model.compile(optimizer='sgd', loss='mse', metrics=['accuracy'])
```

Fixed Code:

```
model = tf.keras.Sequential([
    tf.keras.layers.Conv2D(32, (3,3), activation='relu', input_shape=(28,28,1)),
    tf.keras.layers.MaxPooling2D(2,2),
    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),
    tf.keras.layers.MaxPooling2D(2,2),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(10, activation='softmax')
])
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
```

Problem	Fix
Missing input channel dimension	Added (28,28,1)
Wrong activation (relu for output)	Changed to softmax
Wrong output layer size (Dense(1))	Changed to Dense(10)
Wrong loss (mse)	Changed to sparse_categorical_crossentropy
Inefficient optimizer (sgd)	Upgraded to adam