### Task3:

### ****Training & Transfer Learning Strategy****

#### **1. Freezing the Entire Network**

This turns the model into a fixed feature extractor—great for small datasets where fine-tuning risks overfitting. But it limits adaptability, so it's not ideal if your tasks need domain-specific tweaks.

#### **2. Freezing Only the Transformer Backbone**

Keeping the backbone frozen while training task-specific heads lets you retain general language understanding while adapting to your tasks. It speeds up training and saves compute but may not work well if your domain differs a lot from the model’s original training data.

#### **3. Freezing One Task-Specific Head**

If one task is well-optimized but the other needs fine-tuning, freezing one head prevents performance loss (catastrophic forgetting) while refining the other. This is useful when a task is already performing well, but it assumes no further improvements are needed.

### ****Transfer Learning Done Right****

Picking the right pre-trained model depends on your domain—BERT is great for general NLP, but BioBERT (medical) or FinBERT (finance) might be better for specialized tasks.

A solid approach is to start by freezing the transformer backbone and fine-tuning only the task-specific heads. If performance plateaus, gradually unfreeze the upper layers to refine domain adaptation while keeping core language understanding intact. This keeps training efficient while ensuring the model adapts just enough to perform well.