Bioinformatics + Deep Learning Technical Test

Welcome! This test evaluates your practical and theoretical knowledge in **bioinformatics**, **deep learning**, and **scientific communication**, specifically in the context of **single-cell data analysis** and **generative AI**.

Instructions

- Total time: **24 hours**
- Submit: A Jupyter notebook and a brief report (markdown or PDF)
- Include: All code, visualizations, and explanations inline
- Recommended: Use `scanpy.datasets.pbmc3k()` or any publicly available dataset if none is provided.

PART 1: Single-cell RNA-seq Analysis (Scanpy)

Goal: Preprocess and analyze scRNA-seq data.

Tasks:

1. Load the dataset (`scanpy.datasets.pbmc3k()`).
2. Preprocess:
   - Filter cells and genes
   - Normalize and log-transform
   - Identify highly variable genes
   - PCA + neighbors + UMAP
3. Cluster using Leiden or Louvain.
4. Identify marker genes.
5. Plot:
   - UMAP colored by cluster

- Heatmap of marker genes

6. Summary: Brief biological interpretation (12 paragraphs)

PART 2: Deep Learning Model (Autoencoder or scVI)

Goal: Apply deep learning to dimensionality reduction and reconstruction.

Tasks:

1. Use PyTorch, TensorFlow, or scVI.
2. Build a simple autoencoder:
   - Input: gene expression matrix
   - Output: reconstruction and latent space
3. Visualize latent space (UMAP or PCA).
4. Optionally, fine-tune a pre-trained `scVI` model.

PART 3: Multi-Omics Integration (Mini Task)

Goal: Show theoretical and practical knowledge of omics integration.

Tasks:

1. Briefly describe **two strategies** for integrating multi-omics data (e.g., RNA + protein).
2. Provide pseudocode or actual Python code to preprocess and concatenate RNA and protein matrices for downstream analysis.

PART 4: Scientific Reasoning + Communication

Tasks:

1. You observed poor clustering between two biologically similar cell types:

   - List 3 possible causes

   - Suggest 2 resolution strategies

2. Bonus: Draft a short abstract (max 200 words) summarizing your analysis from Parts 1 & 2.

Submission Checklist

- [ ] Jupyter Notebook with code and outputs

- [ ] Short report/abstract with biological interpretations

- [ ] Include necessary plots: UMAP, heatmap, etc.