Use of AI — RNA → TE Prediction Project

[1] **Notebook — Cell #1**
Tool: ChatGPT
Prompt: "Write Python code to load the RNA and TE CSV files into pandas and validate dimensions."
Output Used: Data loading logic, basic shape checks.

[2] **Notebook — Cell #2**
Tool: ChatGPT
Prompt: "Generate a function to align the RNA and TE matrices and ensure matching gene order."
Output Used: Alignment verification code.

[3] **Notebook — Cell #3**
Tool: ChatGPT
Prompt: "Create a train/validation/test split across genes (70/15/15)."
Output Used: Gene-level split function.

[4] **Notebook — Cell #4**
Tool: ChatGPT
Prompt: "Write a preprocessing function that scales RNA features using StandardScaler."
Output Used: Scaling pipeline.

[5] **Notebook — Cell #5**
Tool: ChatGPT
Prompt: "Implement OLS, Ridge, and Lasso regression training + evaluation functions."
Output Used: Linear model training loop and metrics.

[6] **Notebook — Cell #6**
Tool: ChatGPT
Prompt: "Add Random Forest regression code with fit() and R^2/RMSE calculations."
Output Used: Random Forest model implementation.

[7] **Notebook — Cell #7**
Tool: ChatGPT
Prompt: "Write XGBoost regression code compatible with 2D gene × sample matrices."
Output Used: XGBoost model setup, training, and evaluation.

[8] **Notebook — Cell #8**
Tool: ChatGPT

Prompt: “Implement a multi-output PyTorch neural network for predicting 96 TE values.”
Output Used: ANN architecture, training loop, early-stopping logic.

[9] **Notebook — Cell #9**
Tool: ChatGPT
Prompt: “Write code to compute RMSE and R^2 for multi-output predictions.”
Output Used: Evaluation functions.

[10] **Notebook — Cell #10**
Tool: ChatGPT
Prompt: “Generate PCA analysis plots for RNA and TE matrices.”
Output Used: PCA computation and visualization code.

[11] **Notebook — Cell #11**
Tool: ChatGPT
Prompt: “Write a summary table comparing all model performances (RMSE, R^2).”
Output Used: Model comparison table generator.

[12] **Notebook — Cell #12**
Tool: ChatGPT
Prompt: “Write code to save predictions and model outputs for the final report.”
Output Used: Exporting predictions/results.