This is an AI system that evaluates grammar in 45–60s spoken samples via hybrid acoustic/text analysis, scoring 0–5. Initially achieving a **0.45 Pearson correlation** due to limited feature engineering and overfitting, iterative improvements, including text-grammar error metrics, audio-text feature fusion (MFCC + Chroma + Language Tool), dropout regularization, and modality-specific scaling which boosted performance to **0.63 Pearson**, closely aligning with expert ratings. The final model balances speed (2s/sample on GPU) with robust, human-like assessment capabilities.