Comparison: In-Context Learning vs. Traditional Supervised Learning

Aspect	Traditional Supervised Learning	In-Context Learning (ICL)
Prediction function	$\hat{y}_{\text{test}} = f_{\hat{\theta}}(x_{\text{test}})$	$\hat{y}_{\text{test}} = f_{\phi}(X_{\text{train}}, y_{\text{train}}, X_{\text{test}})$
Generalization target	New samples from same distribution	New datasets from distribution $p(D)$
Training objective	Optimize $\hat{\theta}$ on single dataset	Learn predictor f_{ϕ} across many tasks
Loss function	MLE: $-\log p_{\theta}(y \mid x)$	Conditional NLL of test labels given training data
Learning approach	Per-task weight adaptation via SGD	Single model inference via forward pass
Bayesian view	Point estimate: $p(y \mid x, \hat{\theta})$	Posterior predictive: $p(y_{\text{test}} \mid X_{\text{train}}, y_{\text{train}}, X_{\text{test}})$
Advantages	Efficient inference; strong on large data	No retraining; excels on small data; task generalization