Table 1: Comparison of LLM Post-training Techniques

Feature	Instruction Tuning	Reinforcement Learning from Human Feedback (RLHF)
Core Idea	Fine-tuning the LLM on a dataset of human-written instructions and their corresponding correct responses [2305.09246v1].	Using reinforcement learning to train an LLM, where human feedback on the quality of responses is used to guide the model towards desired behaviors [2204.05862v1].
Main Goal	To improve the model's ability to understand and execute human instructions across a wide range of tasks, from natural language understanding to code generation [2203.11171v4; 2108.07732v1].	To align the model with human values and preferences, thereby improving its safety and reliability by reducing the generation of harmful or inappropriate content [2309.16609v1; 2404.15974v1].
Advantages	- Significantly improves performance on diverse tasks [2203.11171v4; 2108.07732v1]. - Recent research shows competitive performance is achievable with significantly less training data [2305.09246v1].	- Effectively reduces the output of unsafe or undesirable content [2309.16609v1; 2404.15974v1]. - Enables the model to continuously learn from its mistakes and improve over time.
Challenges	Requires large amounts of high-quality instruction-response data, which can be time-consuming and expensive to create and collect [2305.09246v1].	- The process is computationally expensive. - The model's performance can be skewed by biases present in the human feedback [2204.05862v1].