Comparison of Model Merging Methods

Method	What It Is	Pros	Cons
Task Vector Algorithms	Merging by adjusting weight differences for task-specific improvements.	• Enhances task-specific capabilities. • Supports merging several models • Offers flexibility with various merging strategies.	• Dependent on a suitable base model. • Increased complexity with more models. • Risk of overfitting to specific tasks.
SLERP	An interpolation technique for smoothly transitioning between two models.	• Preserves distinct features of each model. • Single hyperparameter to tune.	• Restricted to two models. • May not fully capture the advantages of each.
Frankenmerging	Stacking different layers from multiple models sequentially to create a new model.	• Enables integration of models with varied architectures. • Highly customizable.	• Demands significant experimentation. • May face layer compatibility challenges.
Mixture of Experts (MoE)	Dynamically combines expert models through per-layer, per-token routing for precise, task-adaptive activation.	• Precision in activating models for specific tasks. • Adapts dynamically to the complexity of input. • Efficiently utilizes model capacity by activating relevant experts.	• Complexity in configuring and managing routing. • Potential overhead from dynamic routing decisions. • Requires careful balancing to avoid overloading specific experts.