

F.3 Choice of Optimizer

As detailed in the Section F.1, we use Adafactor as our optimizer for all downstream fine-tuning experiments. A reasonable concern would be if the choice of optimizer could influence the reported results in Section 4.3. To address these we conducted a targeted ablation and offer additional theoretical justification below.

We replicated the downstream experiment on the CTI MCQ dataset, scaling the data size from 4k to 84k using the Muon optimizer (Jordan et al., 2024). We opted for the adaptive Muon optimizer because of its structural distinctness from Adafactor (i.e., momentum-orthogonal vs. factored second-moment). As shown in Figure 12, we observe that the relative performance trends hold: The Full Simula system consistently outperforms the Baseline and different variants across data sizes. These results suggest that the benefits provided by our reasoning-driven approach are robust to the choice of optimizer.

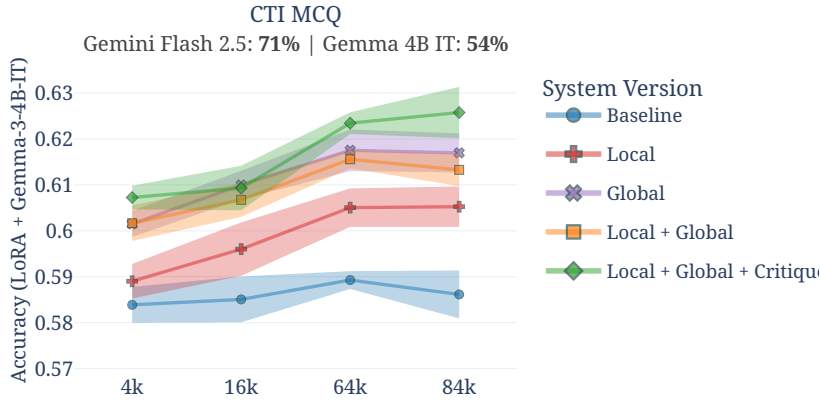


Figure 12: **Downstream Performance using Muon Optimizer.** We report mean accuracy with 95% CI for the CTI MCQ dataset for student models fine-tuned using the Muon optimizer. We continue to see that the full Simula system version with the critic, provides the best performance compared to the other versions.

Informed Hyperparameter Selection. The reported results in Section 4.3 using Adafactor are not based on default settings, but the outcome of rigorous hyperparameter sweeps over data sizes and configurations (see Appendix F.1). As noted in (Schmidt et al., 2021), while optimizer performance can vary greatly across tasks, the authors show that evaluating multiple optimizers with default parameters works approximately as well as tuning the hyperparameters of a single, fixed optimizer.

Scaling Laws for Relative Trends. The goal of our ablations is to measure the *relative* performance differences of the different system components. We can therefore distinguish between the *exponent* of the scaling law (rate at which error decreases with data) and the *intercept* (training efficiency). The former is shown to be strongly correlated with data composition (Sorscher et al., 2022; Chen et al., 2025), while the latter is influenced by optimization and architecture (Schmidt et al., 2021; Everett et al., 2024). Since our claims focus on optimal scaling trajectories (exponent) of Simula data, the choice of optimizer (primarily intercept) should not invalidate the relative advantages observed.