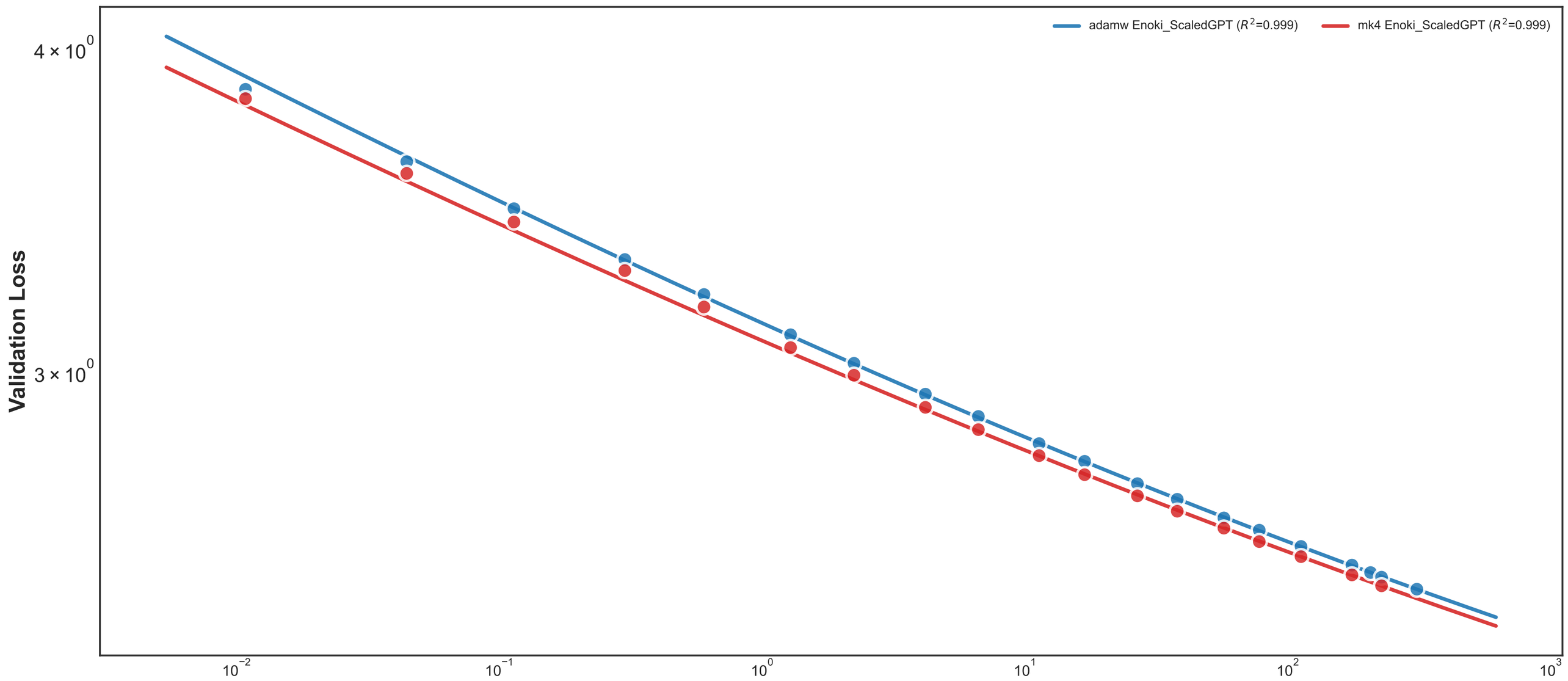
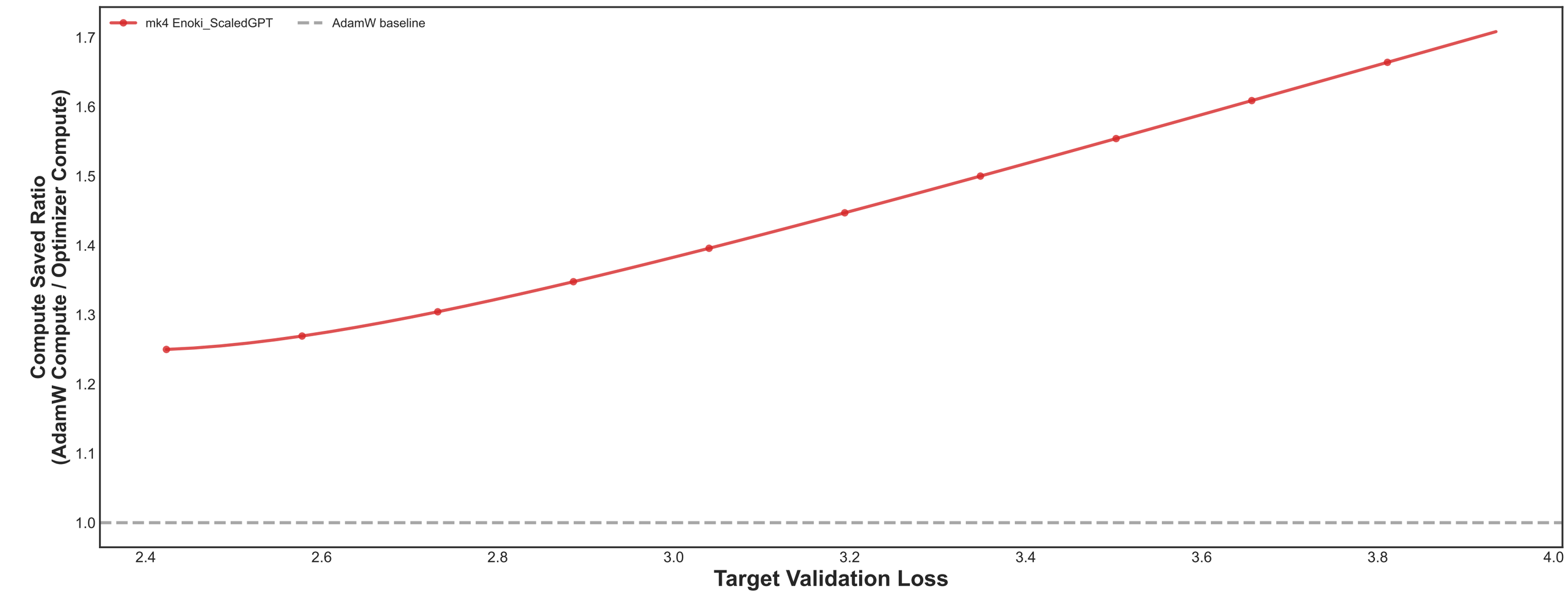
Scaling Laws (Broken Power Law): Enoki_ScaledGPT
Optimizers: adamw, mk4 (Shared a = 0.1058)



Compute Efficiency Relative to AdamW (Higher = More Efficient)