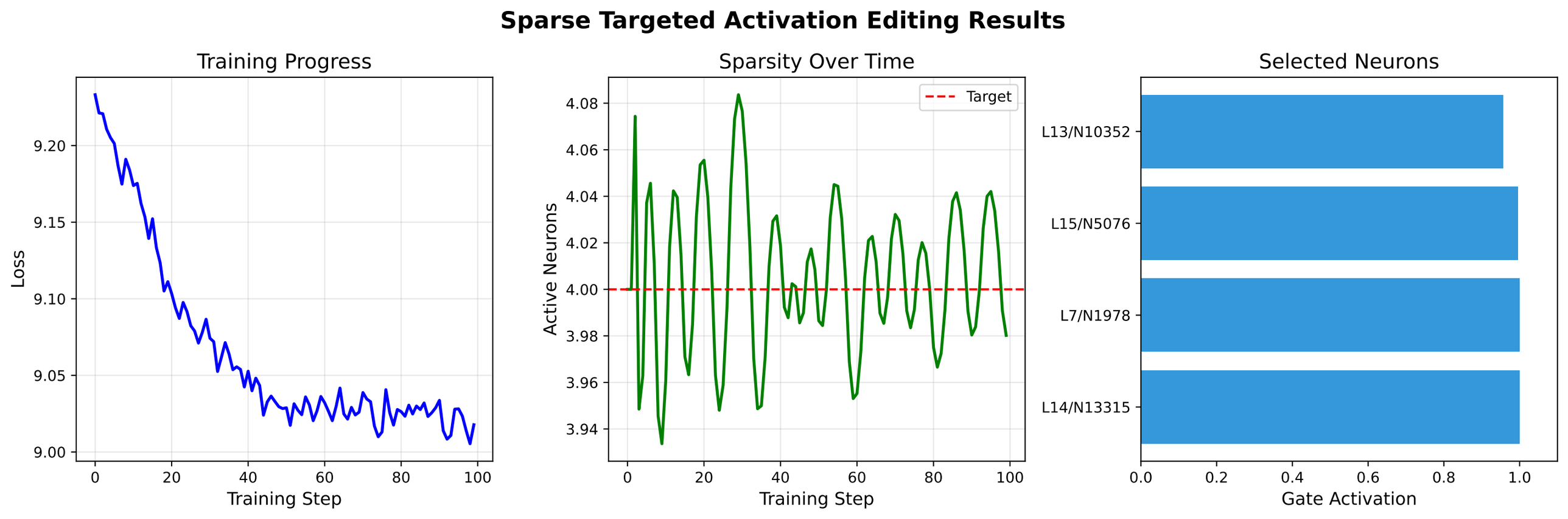
Sparse Targeted Activation Editing Results
Training Progress
Loss
Training Step
Sparsity Over Time
Active Neurons
Target
Training Step
Selected Neurons
L13/N10352
L15/N5076
L7/N1978
L14/N13315
Gate Activation