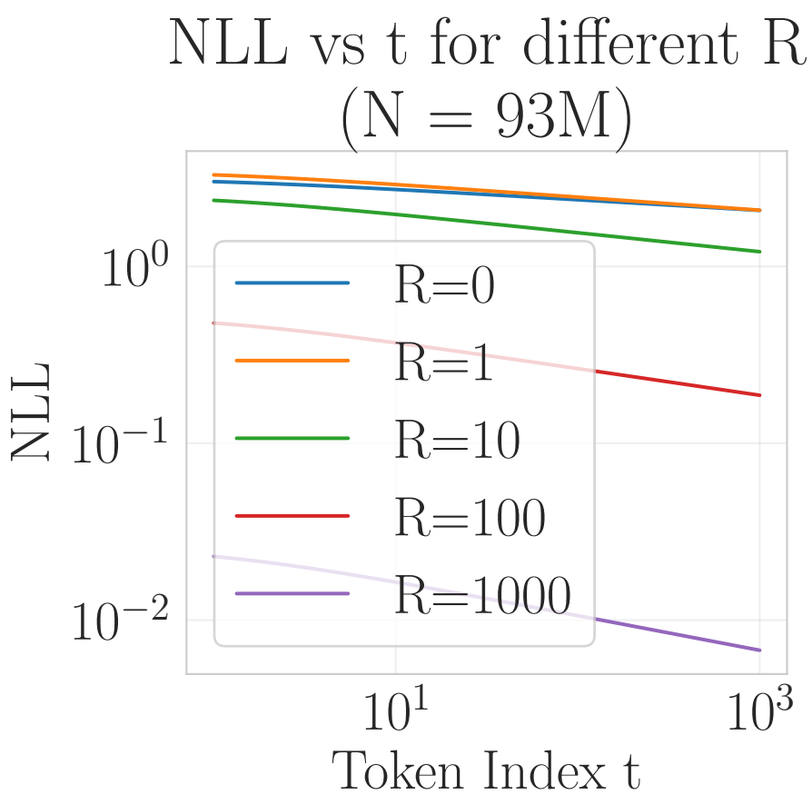
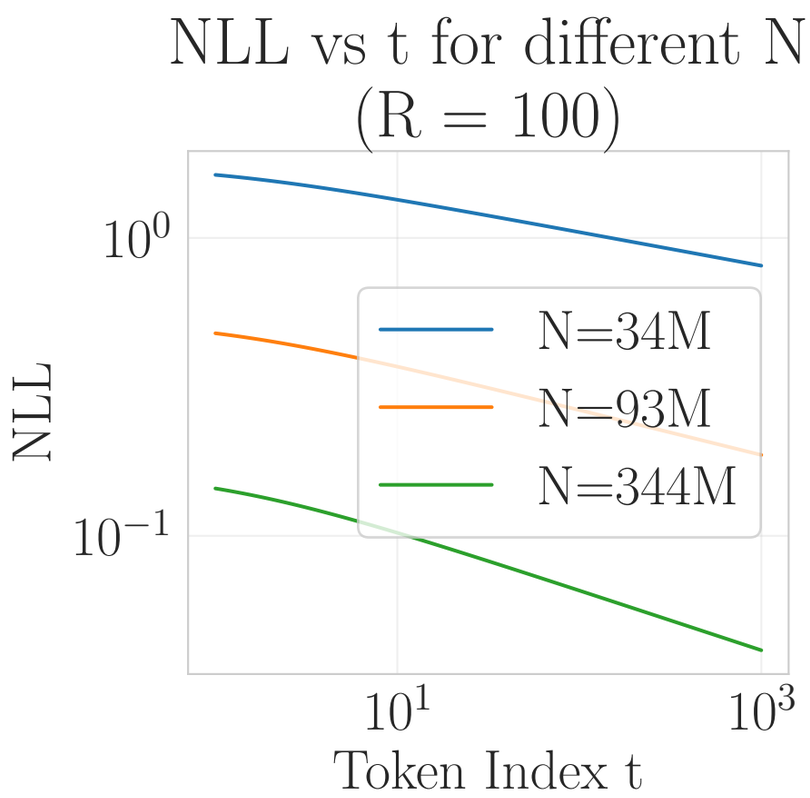
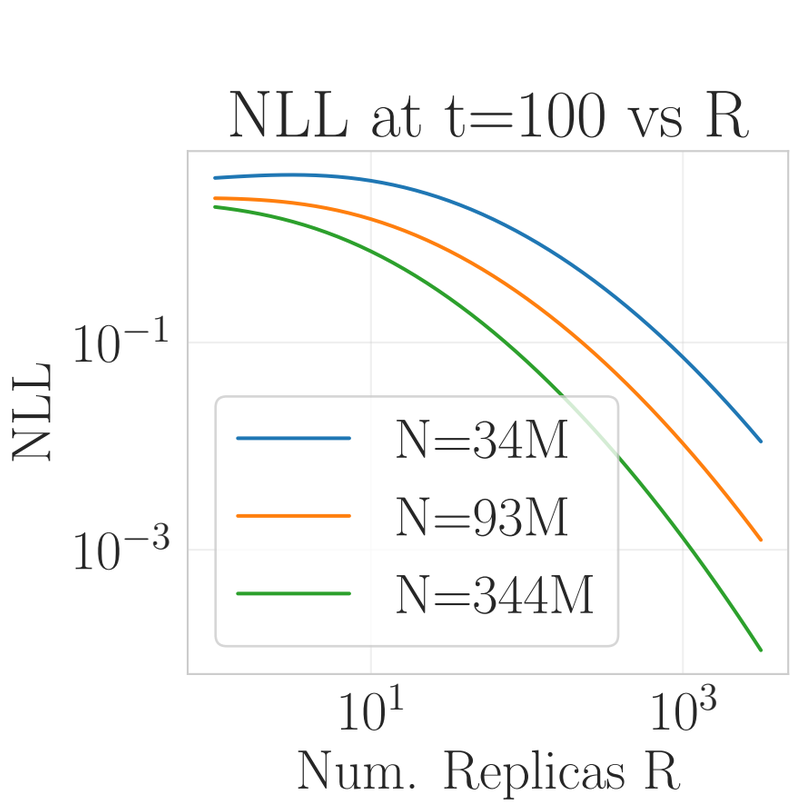
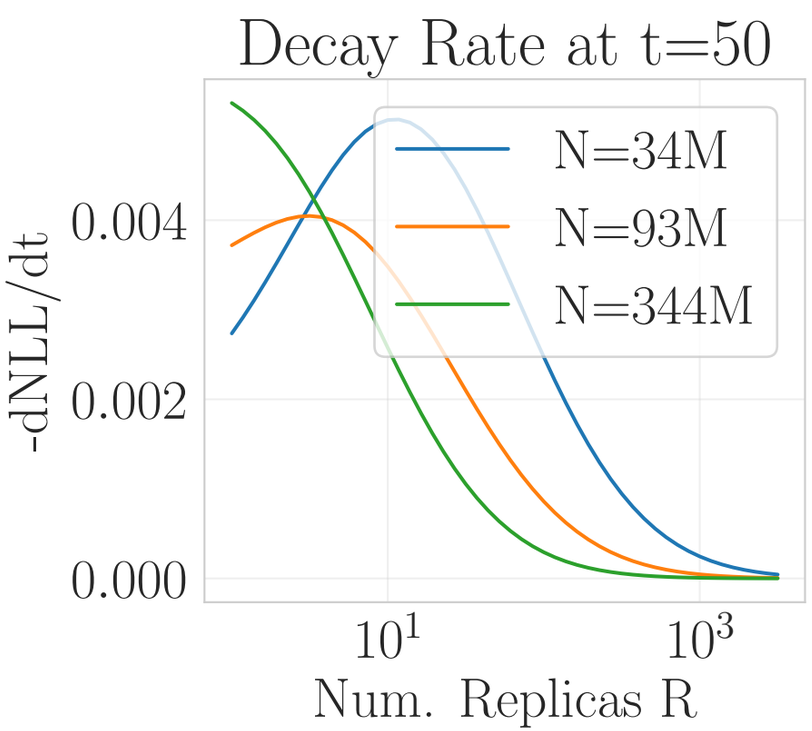
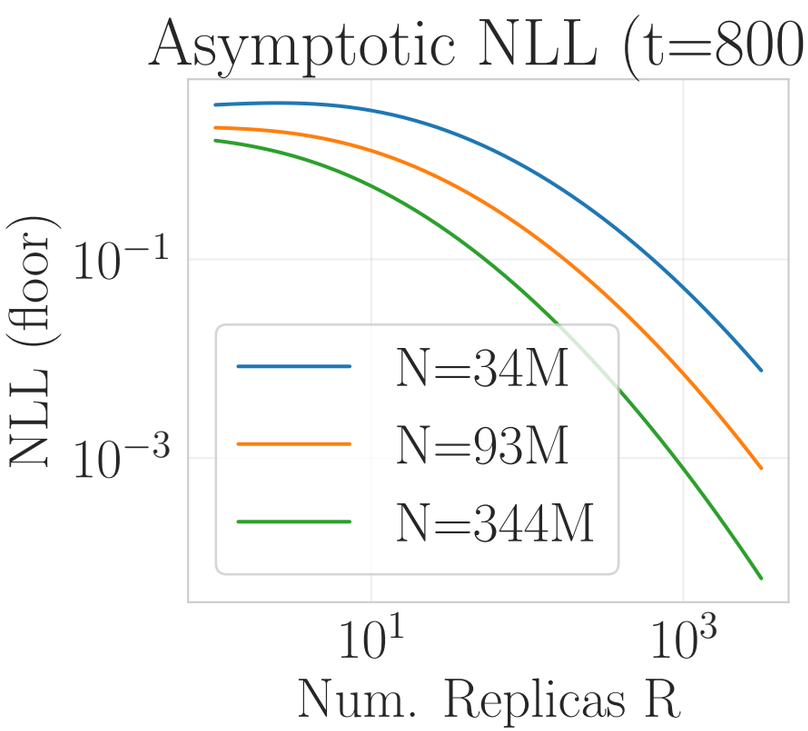
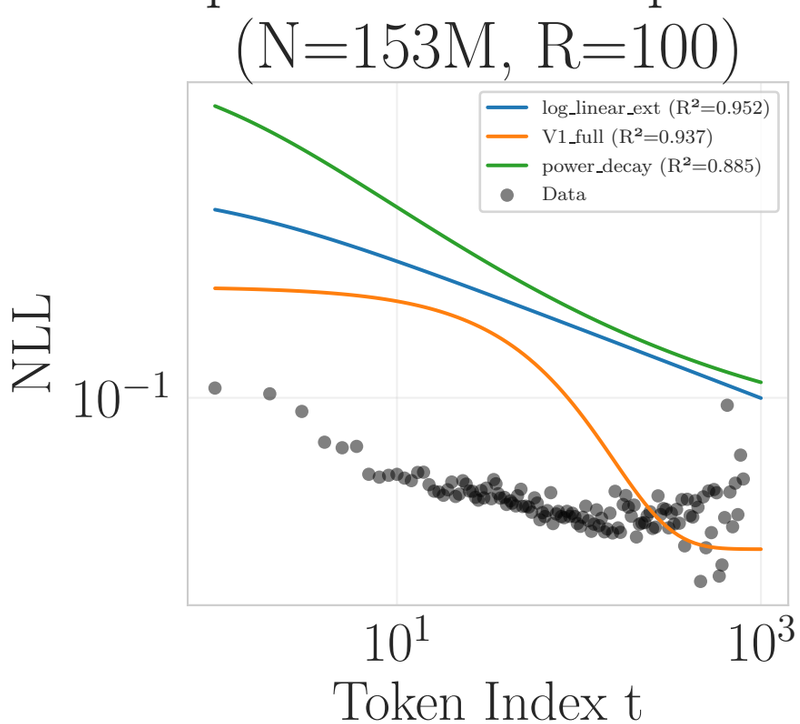

NLL vs t for different R
(N = 93M)
NLL
10^0
10^{-1}
10^{-2}
R=0
R=1
R=10
R=100
R=1000
10^1
10^3
Token Index t


NLL vs t for different N
(R = 100)
NLL
10^0
10^{-1}
N=34M
N=93M
N=344M
10^1
10^3
Token Index t


NLL at t=100 vs R
NLL
10^{-1}
10^{-3}
N=34M
N=93M
N=344M
10^1
10^3
Num. Replicas R


Decay Rate at t=50
-dNLL/dt
0.004
0.002
0.000
N=34M
N=93M
N=344M
10^1
10^3
Num. Replicas R


Asymptotic NLL (t=800)
NLL (floor)
10^{-1}
10^{-3}
N=34M
N=93M
N=344M
10^1
10^3
Num. Replicas R


Top 3 Models Comparison
(N=153M, R=100)
NLL
10^{-1}
log_linear_ext (R^2=0.952)
V1_full (R^2=0.937)
power_decay (R^2=0.885)
Data
10^1
10^3
Token Index t