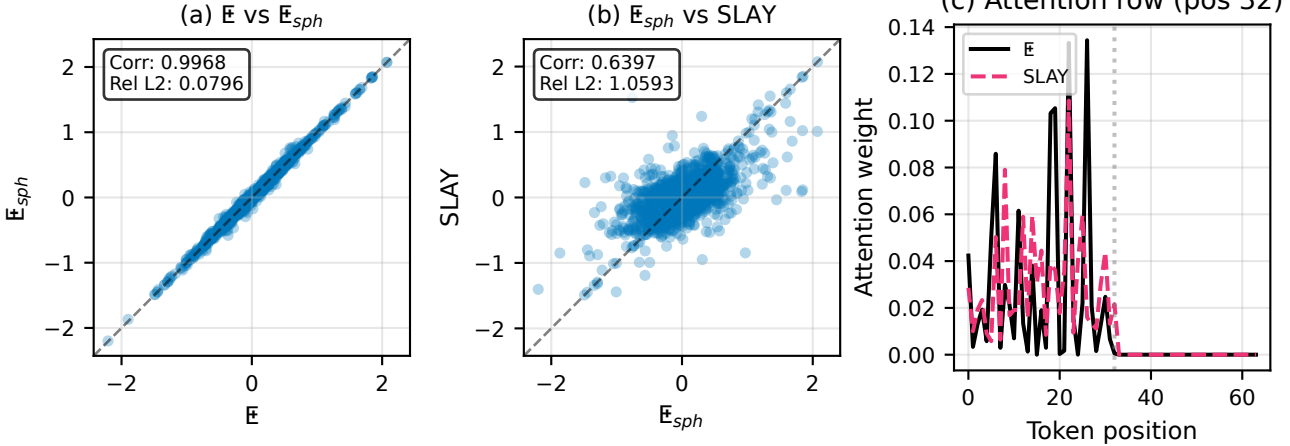
(a) Ɛ vs Ɛ$_{sph}$
Corr: 0.9968
Rel L2: 0.0796
Ɛ$_{sph}$
Ɛ
(b) Ɛ$_{sph}$ vs SLAY
Corr: 0.6397
Rel L2: 1.0593
SLAY
Ɛ$_{sph}$
(c) Attention row (pos 32)
Ɛ
SLAY
Attention weight
Token position