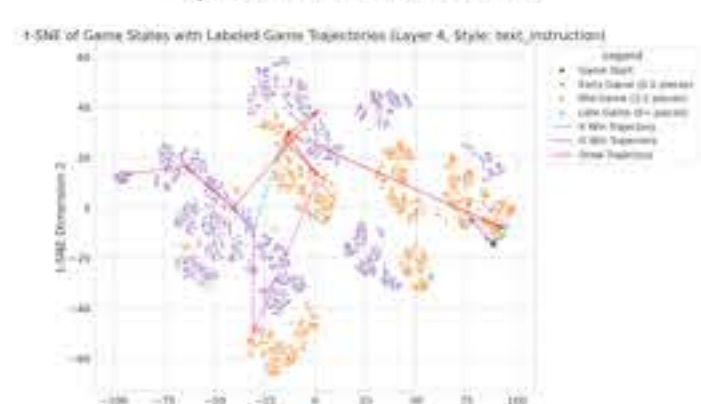


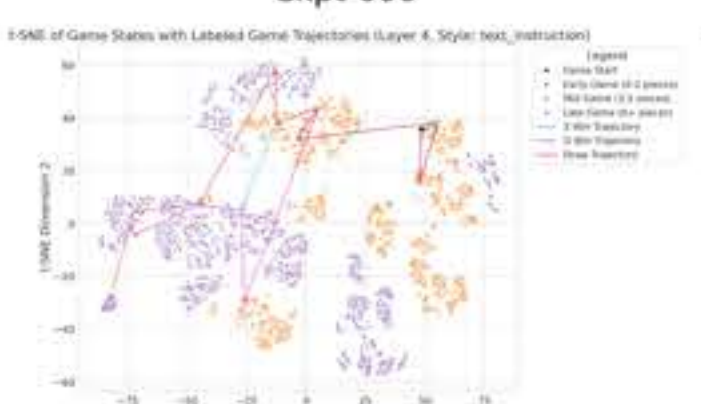
viz hypothesis game trajectories text instruction

Layer 4

Qwen2.5 1.5B Instruct



GRPO Best Move (Base) Ckpt 600



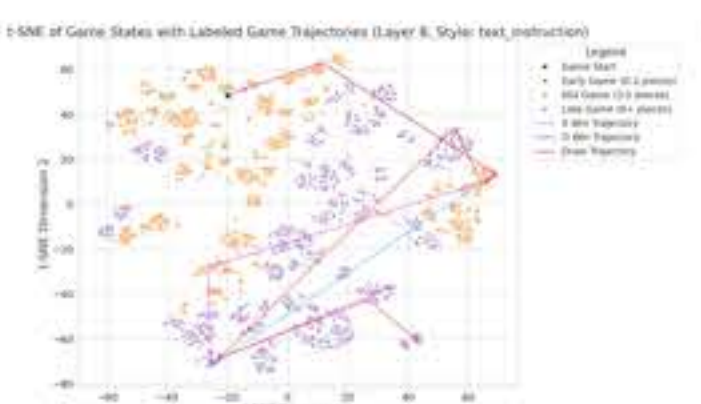
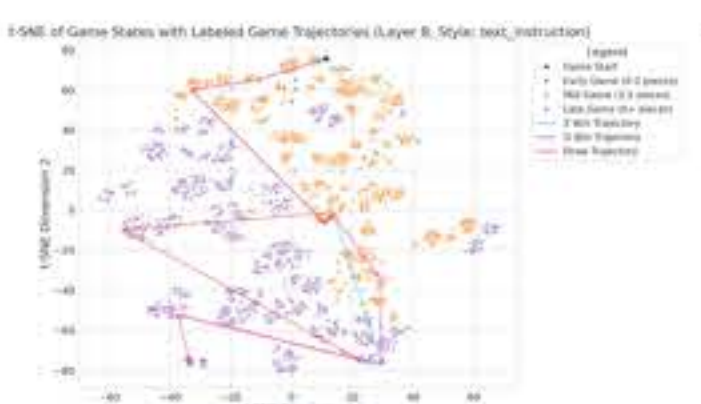
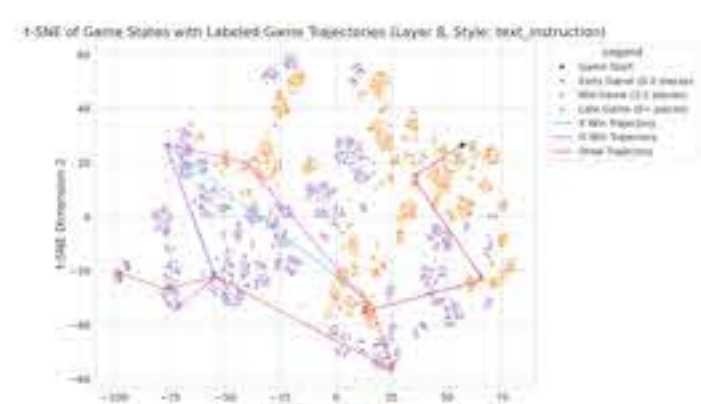
GRPO Best Move (Base) Ckpt 1200



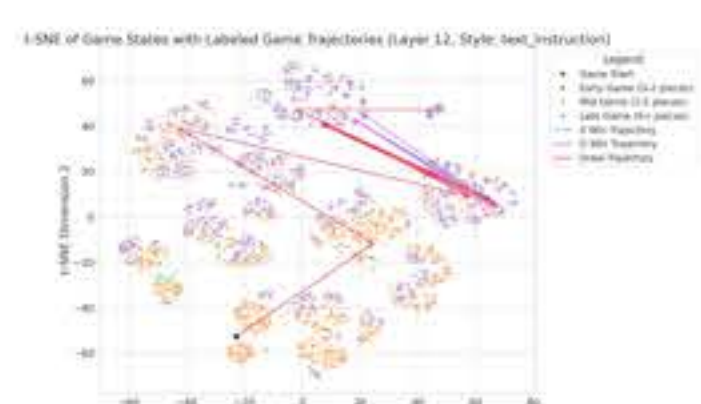
GRPO Best Move (Base)



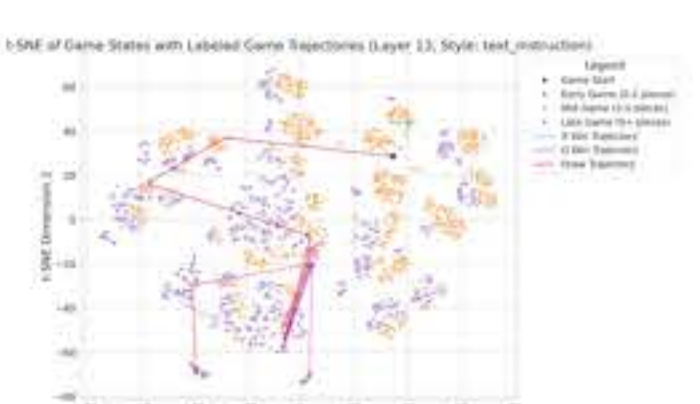
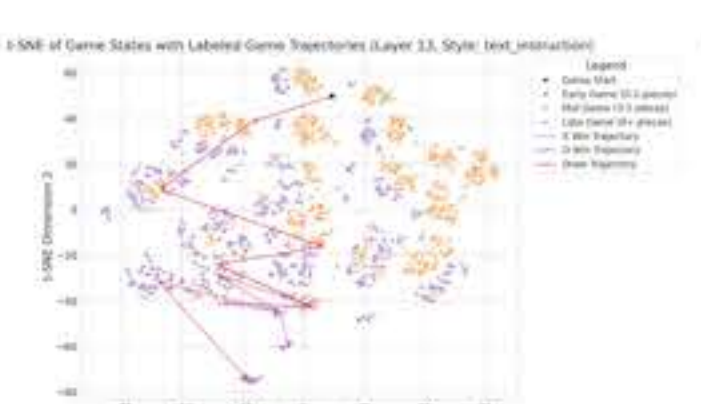
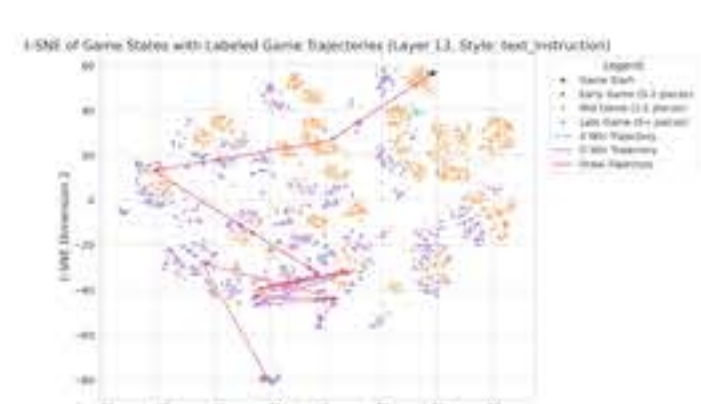
Layer 8



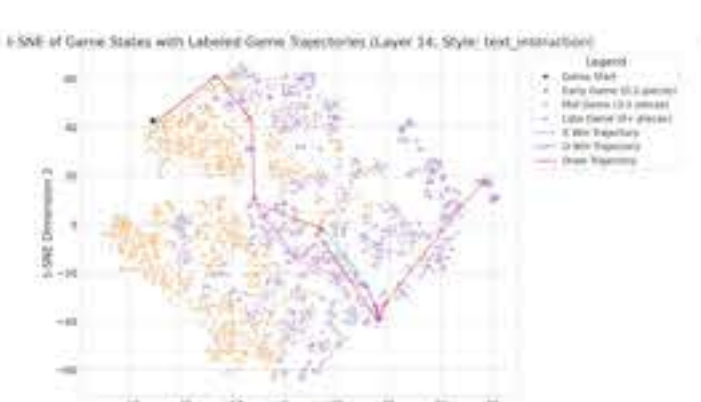
Layer 12



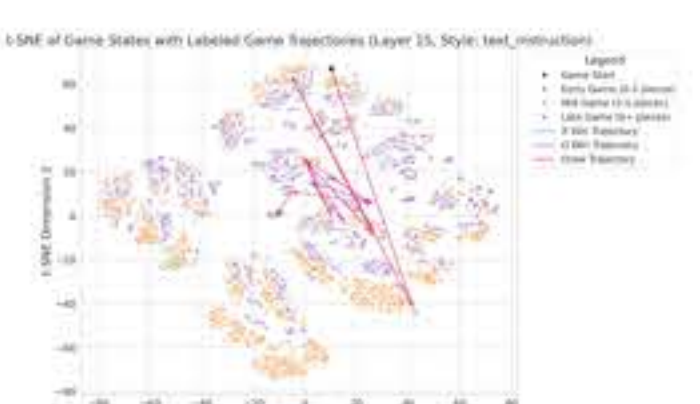
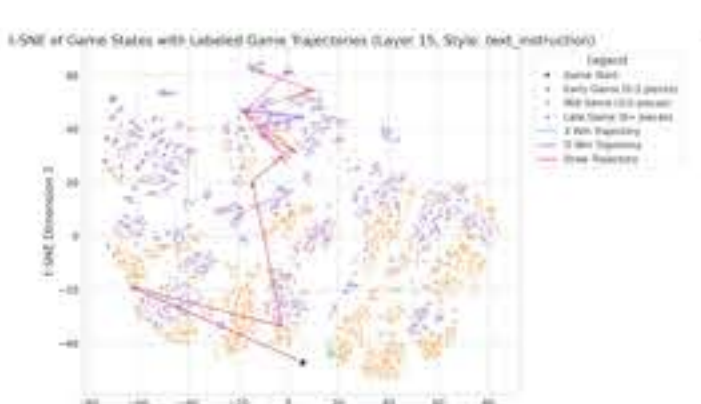
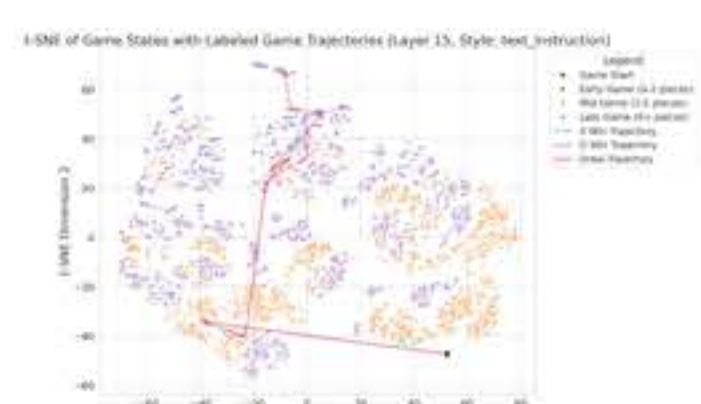
Layer 13



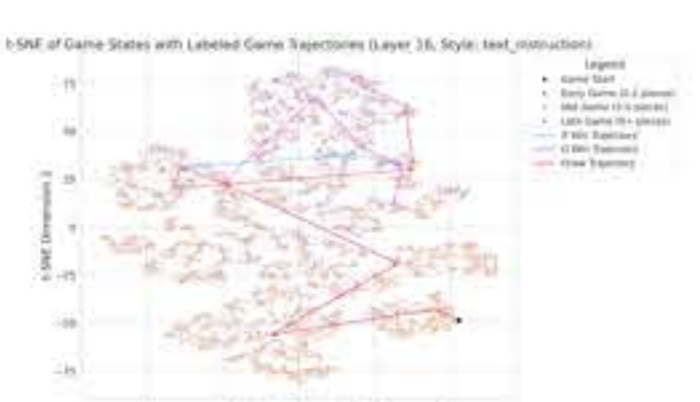
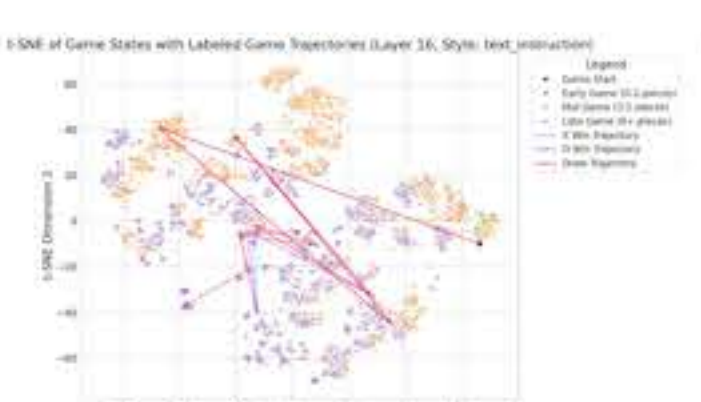
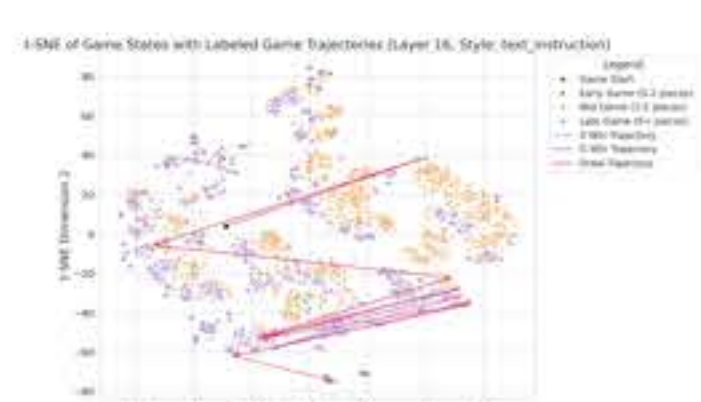
Layer 14



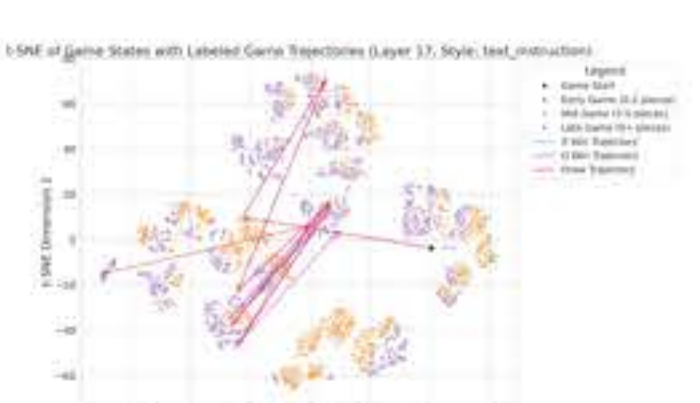
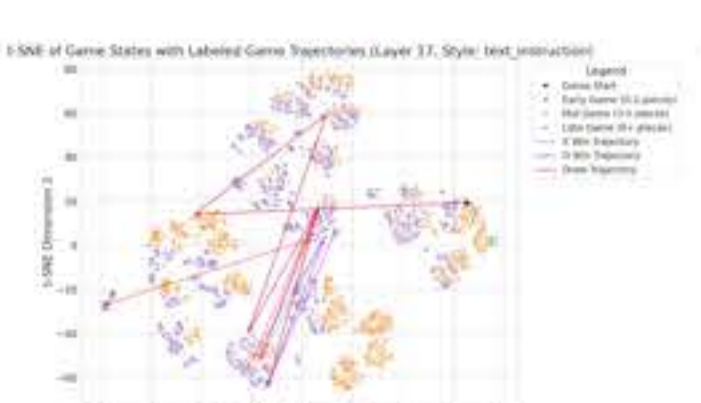
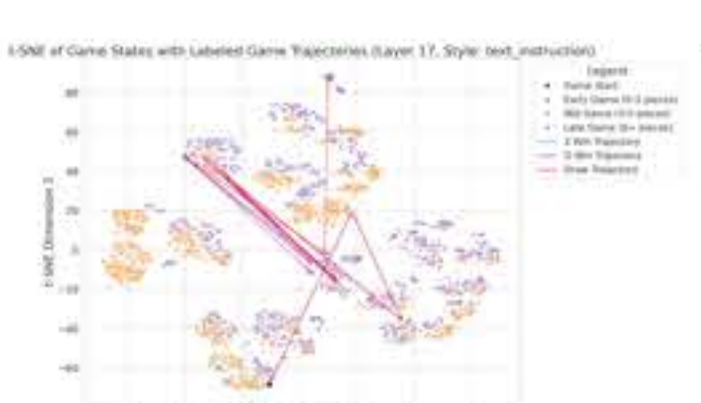
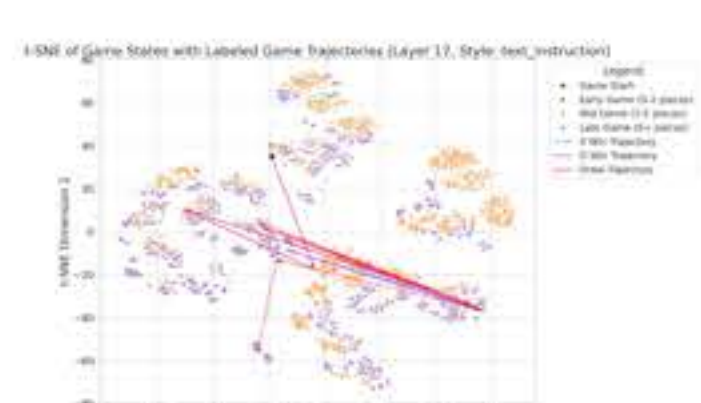
Layer 15



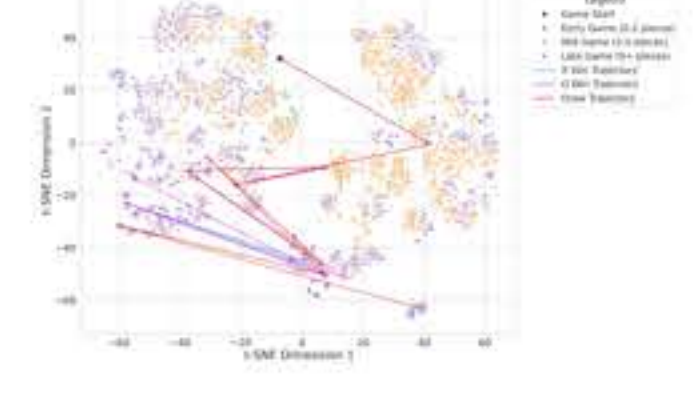
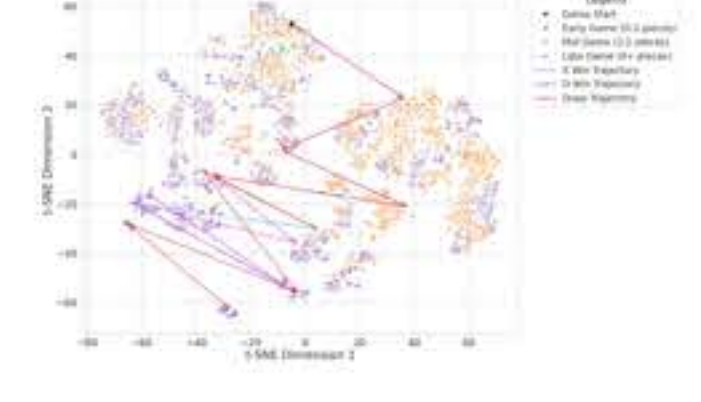
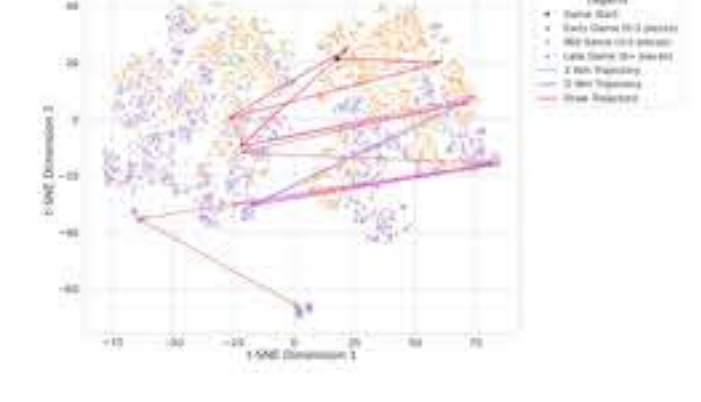
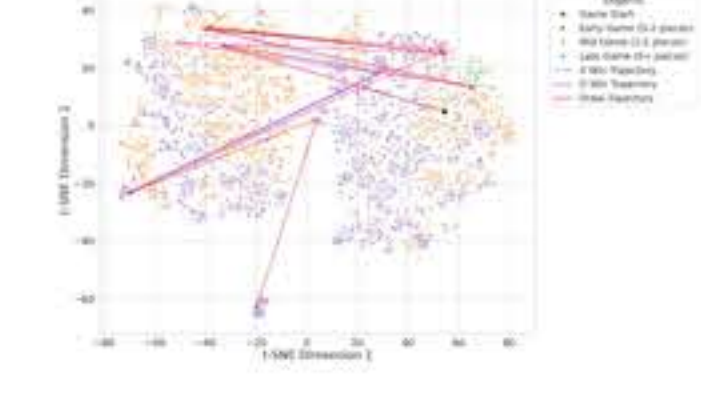
Layer 16



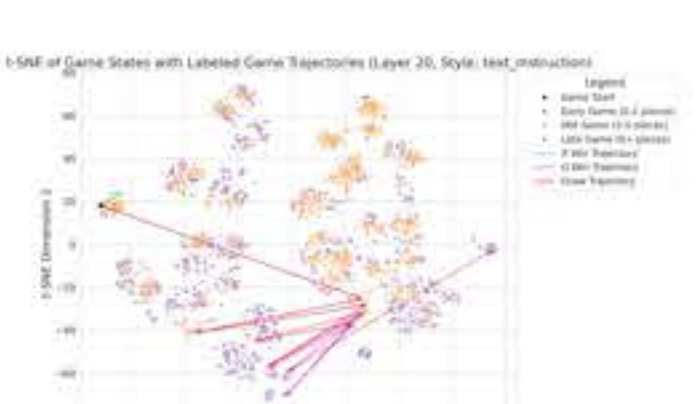
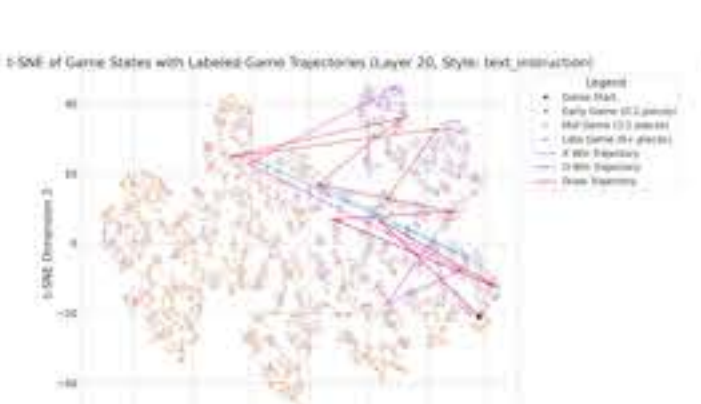
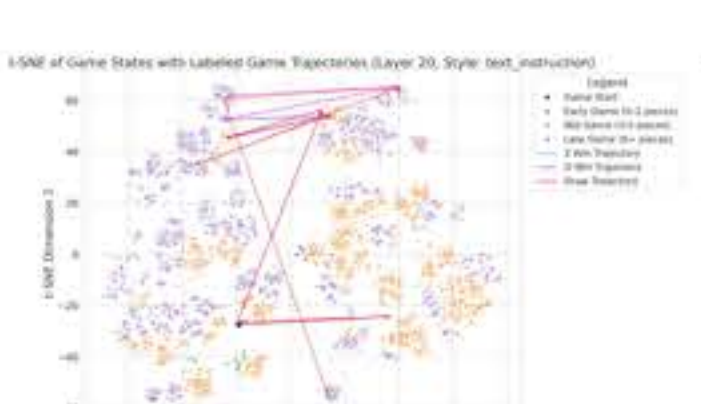
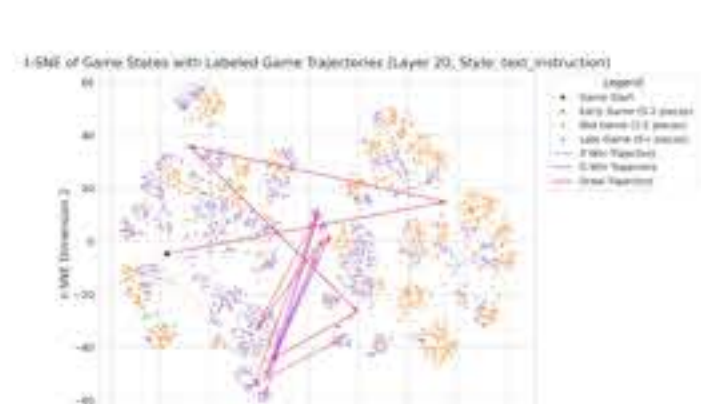
Layer 17



Layer 18



Layer 20



Layer 24

