



CVPR 2026



Multi-level Causal LLM-based Text-to-Motion Generation with Human Alignment

Xiaodong Chen^{1,2} Qian Bao^{2†} Xudong Liu² Jianping Fang² Jintao Fang²
Yongdong Zhang¹ Tao Mei³ Wu Liu^{1†}

¹University of Science and Technology of China, Hefei, China

² Unaffiliated, Beijing, China ³HiDream.ai, Beijing, China

Text-driven 3D Human Motion Generation

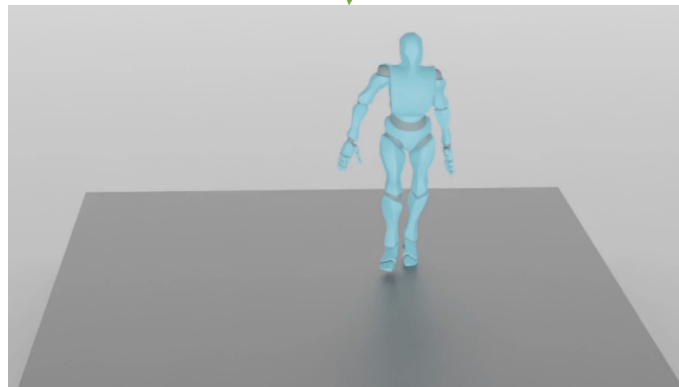
Text-driven 3D human motion generation aims to synthesize realistic 3D human motion sequences from natural language descriptions

“A person does a couple of powerful kicks with their right leg”

“A person does a lunge with right foot leading”

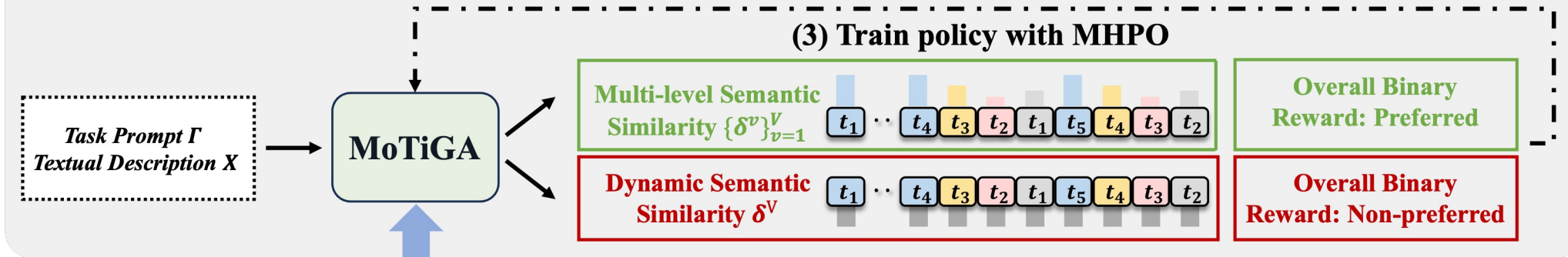
“A person dodges something to the right, then the left and then the right again”

MoTiGA: Motion Generator

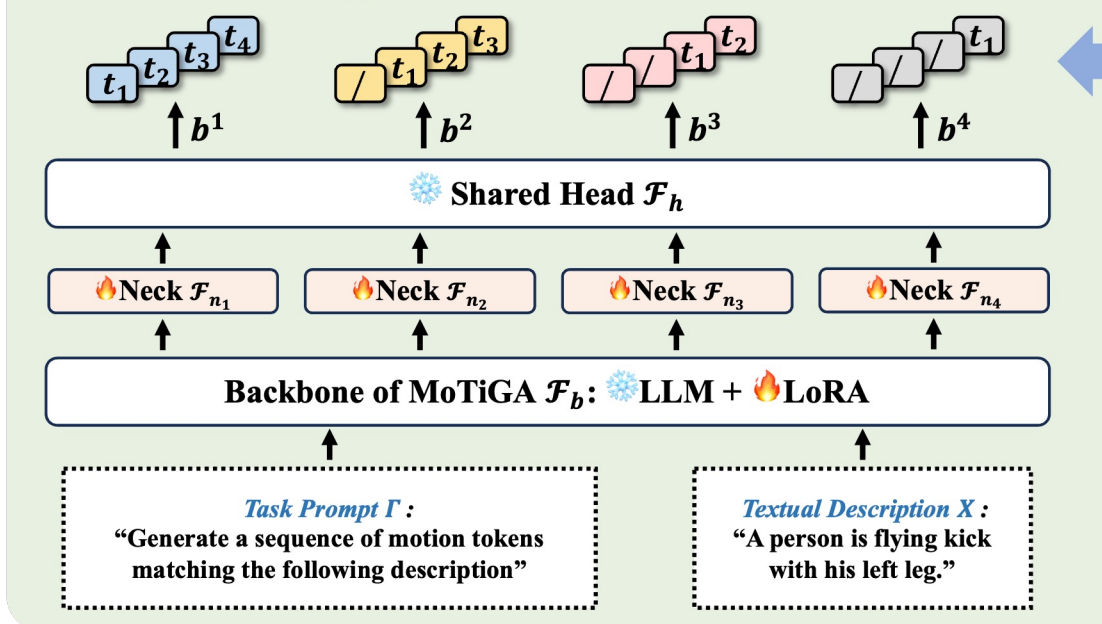


Approach Overview of MoTiGA

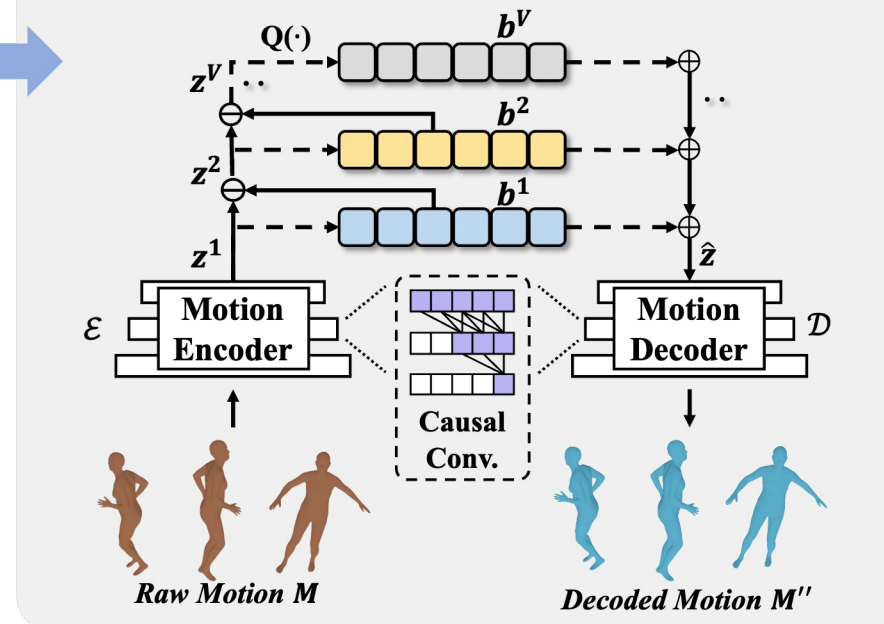
MoTiGA: Multi-level Causal LLM-based Text-to-Motion Generation with Human Alignment



(1) Time-lagged Causal Prediction of MoTiGA



(2) Causal RVQ-VAE Tokenizer of MoTiGA

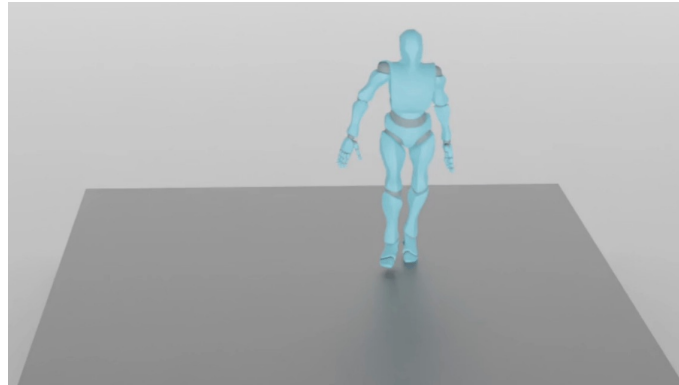


Gallery of Motion Generation

“A person does a couple of powerful kicks with their right leg”



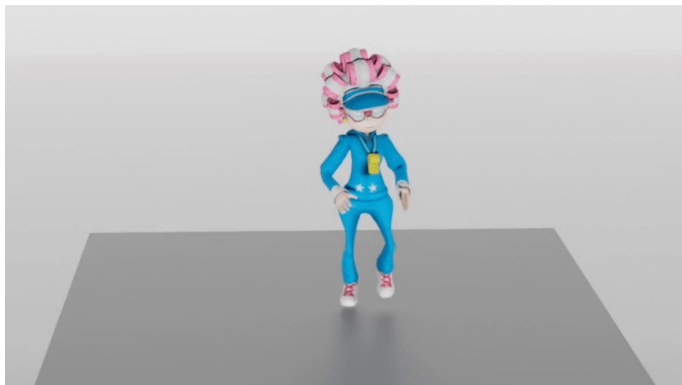
“A person does a lunge with right foot leading”



“A person dodges something to the right, then the left and then the right again”



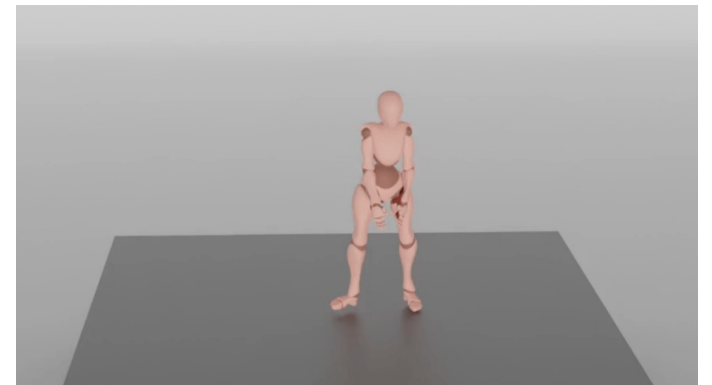
“A person jogs on the spot”



“A person is walking while kicking out legs”



“A golfer takes a golf swing”



Impact of Human Alignment

“A person **does a couple of powerful kicks** with their right leg”



W/o Human Alignment



With Human Alignment

Impact of Human Alignment

“A person dodges something **to the right**, then **the left** and then **the right again**”



W/o Human Alignment



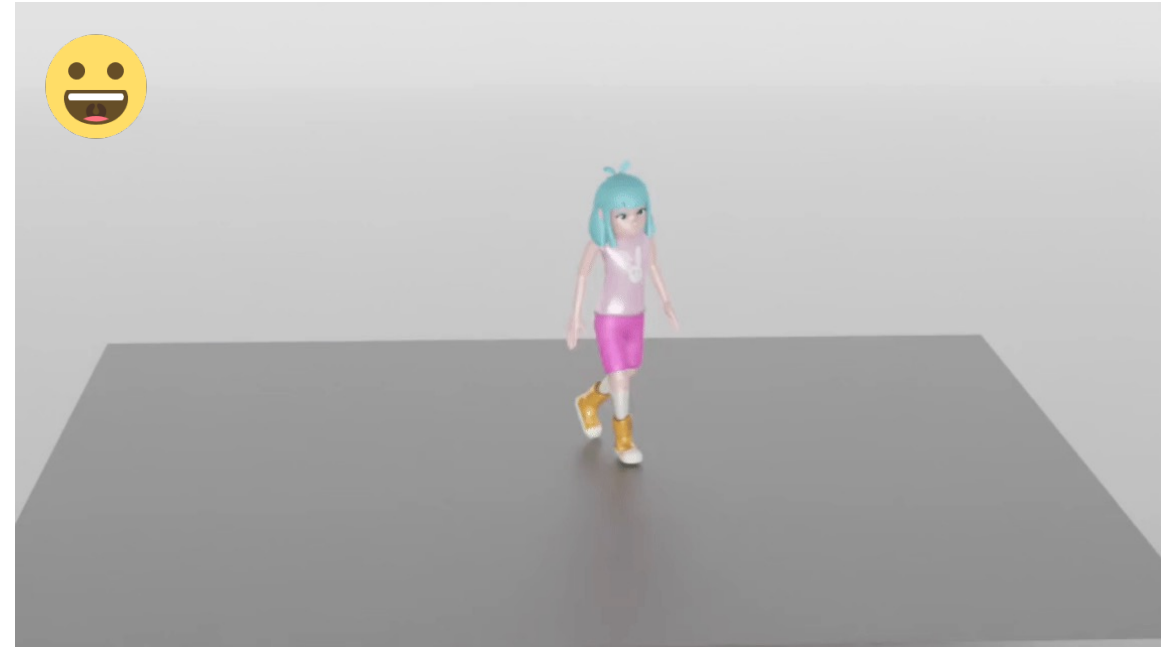
With Human Alignment

Impact of Human Alignment

“A person is **walking** while **kicking out legs**”



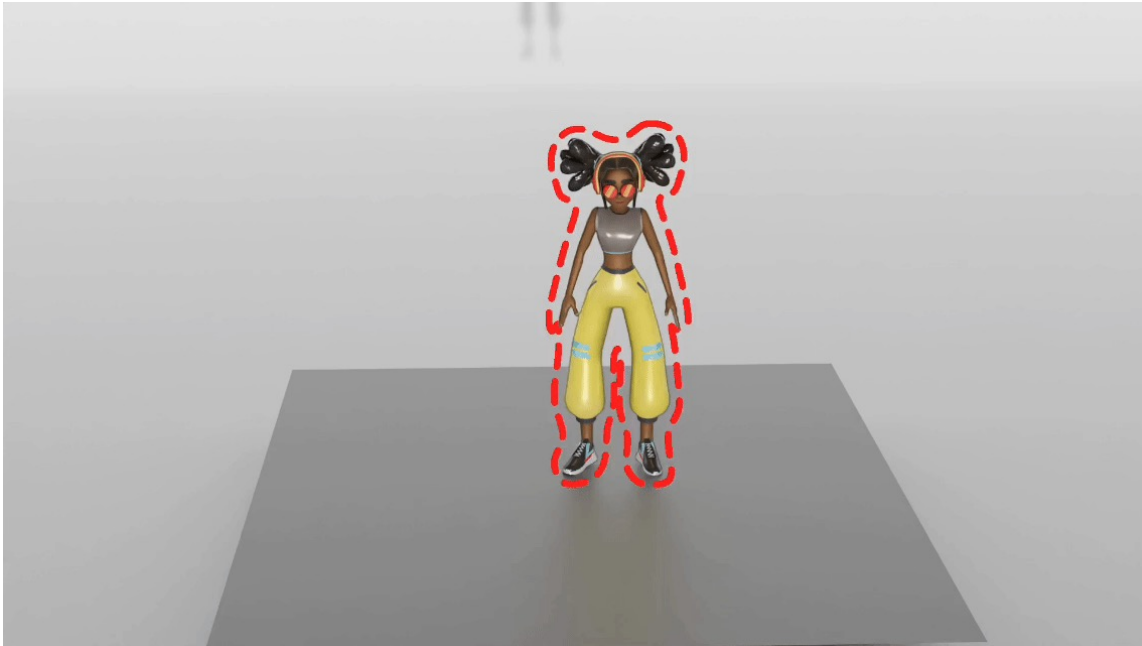
W/o Human Alignment



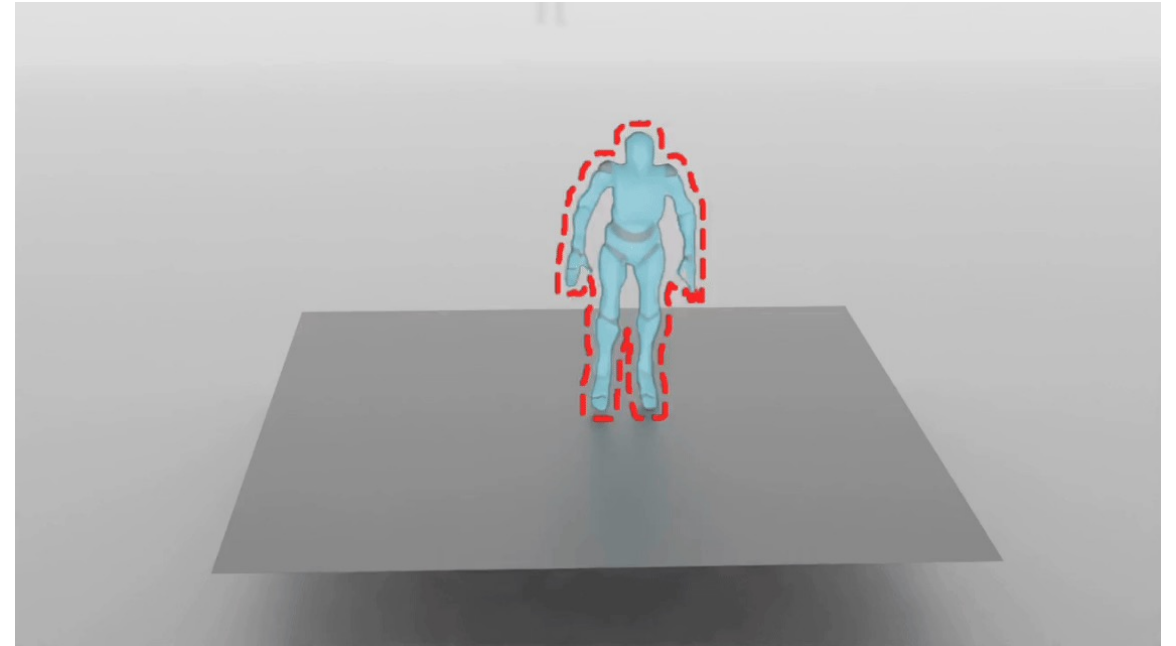
With Human Alignment

Generation with Given Poses

“a person raises both their arms and then places them back down”



“a person kicks out with the left leg”



The Given Poses are Highlighted in Red Dashed Line