Codes	Semantics
1	R-arm straight
2	R-arm barely bent
3	R-arm quite bent
8	L-knee straight
9	L-knee barely bent
10	L-knee quite bent
53	L-hand-foot wide
54	L-hand-foot spread
55	L-hand-foot close

Pose Codebook

Encoder

Decoder

Motion Encoder-Decoder (Section 3.1)

Text Description
A person is crouching.

LLM

A person is crouching.
Left arm downward.
Bend at the knee.
Torso leaning forward.

Fine-grained Keywords

CLIP

text
kw_1
kw_2
kw_n

Transformer Layers

Decoder

Motion Generator (Section 3.2)