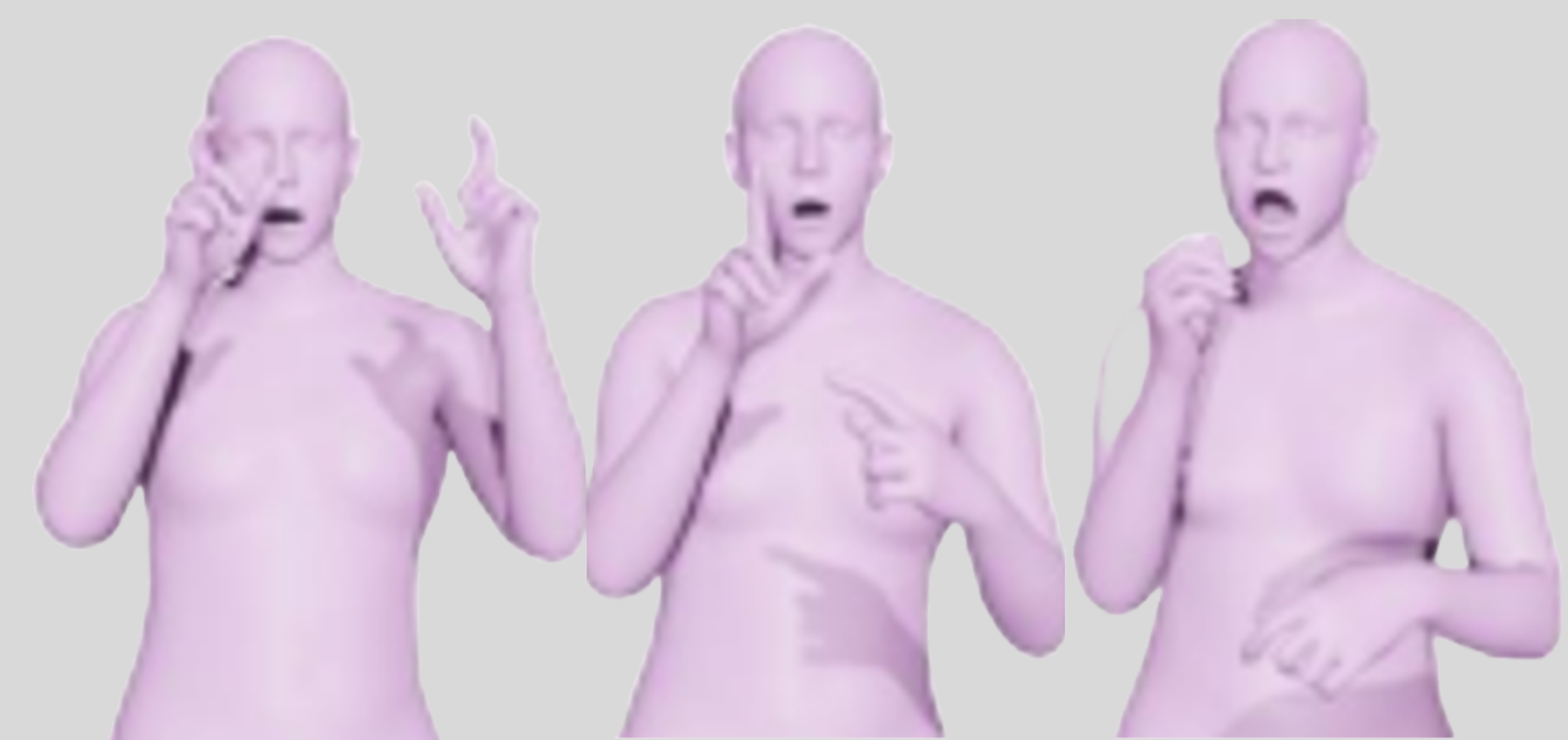


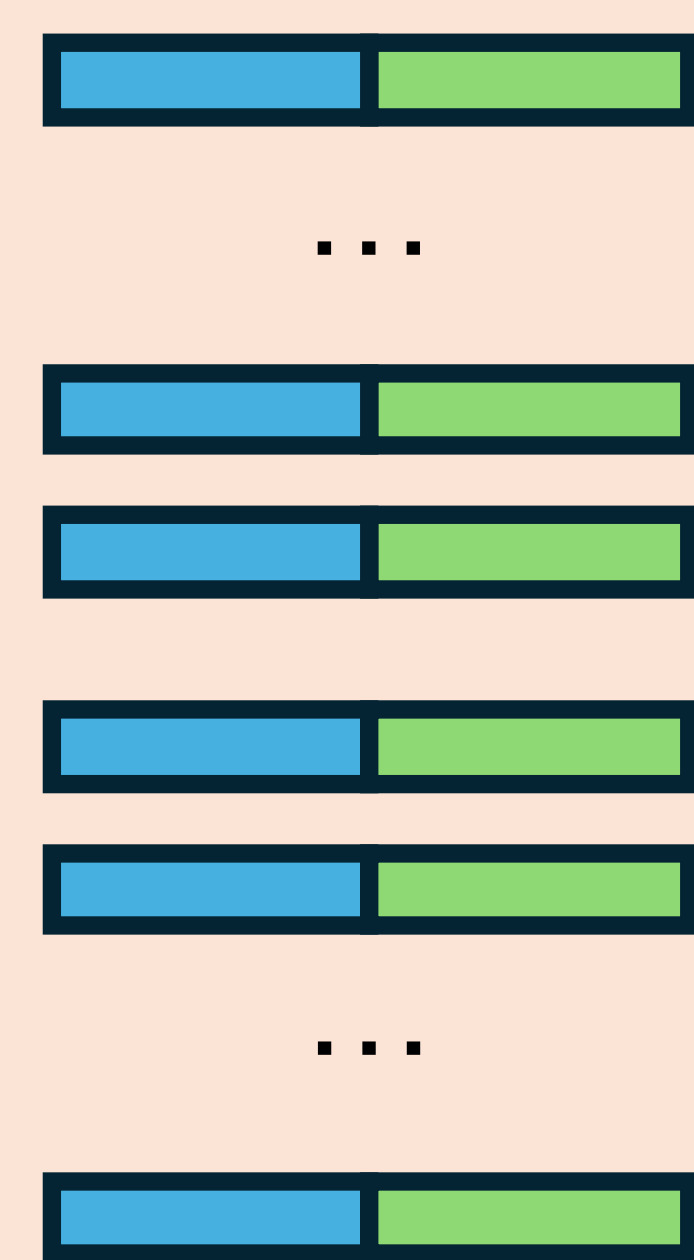
Stage 1: Gesture Feature Alignment



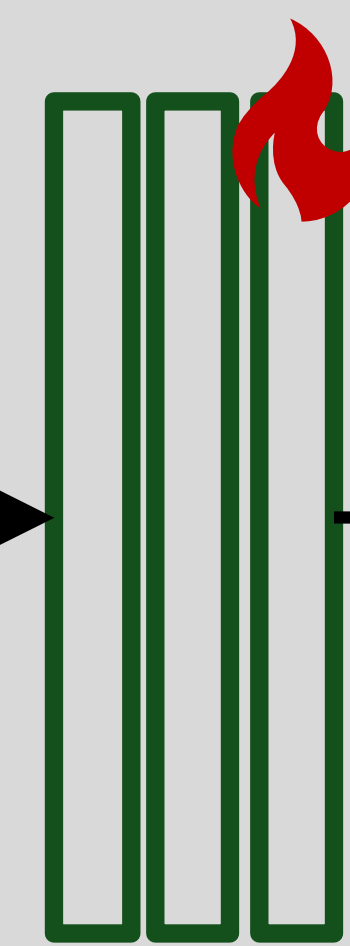
Sign Language Gesture

Gesture
Encoder

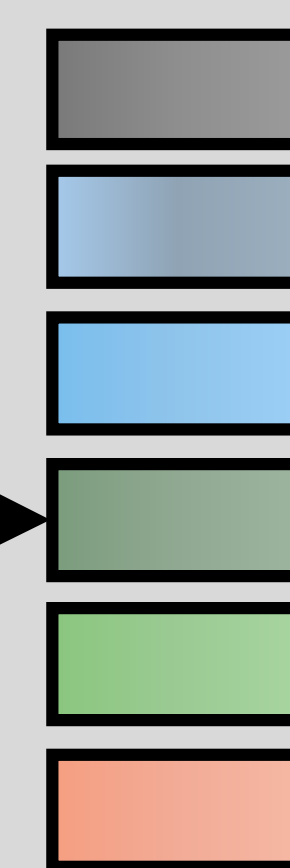
Gesture Codebook



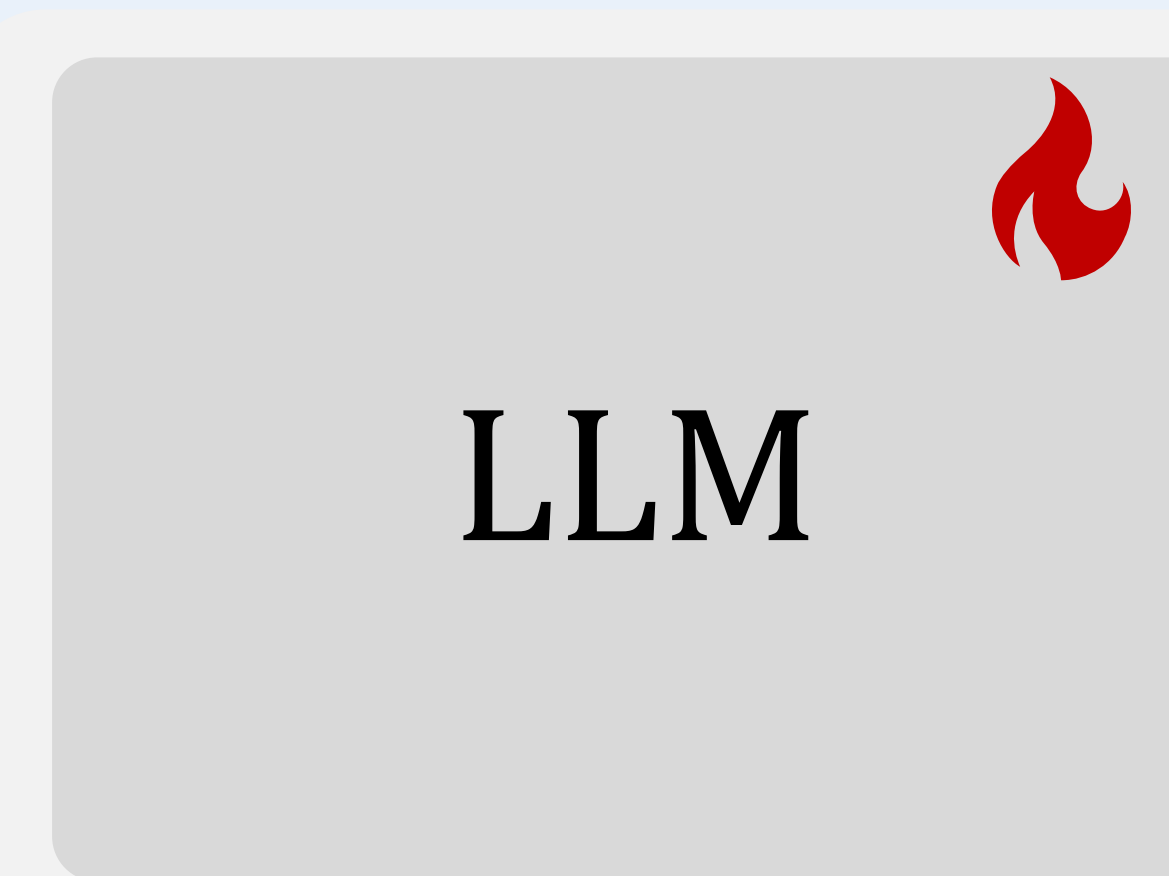
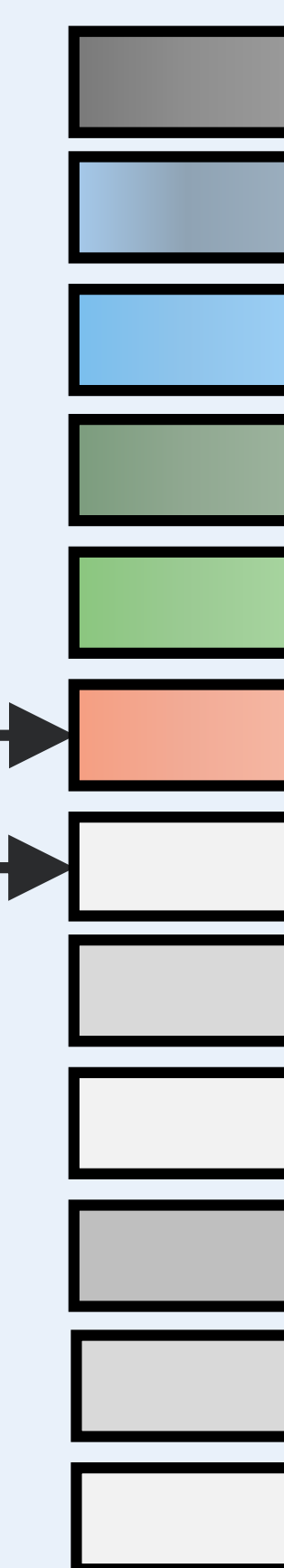
Alignment



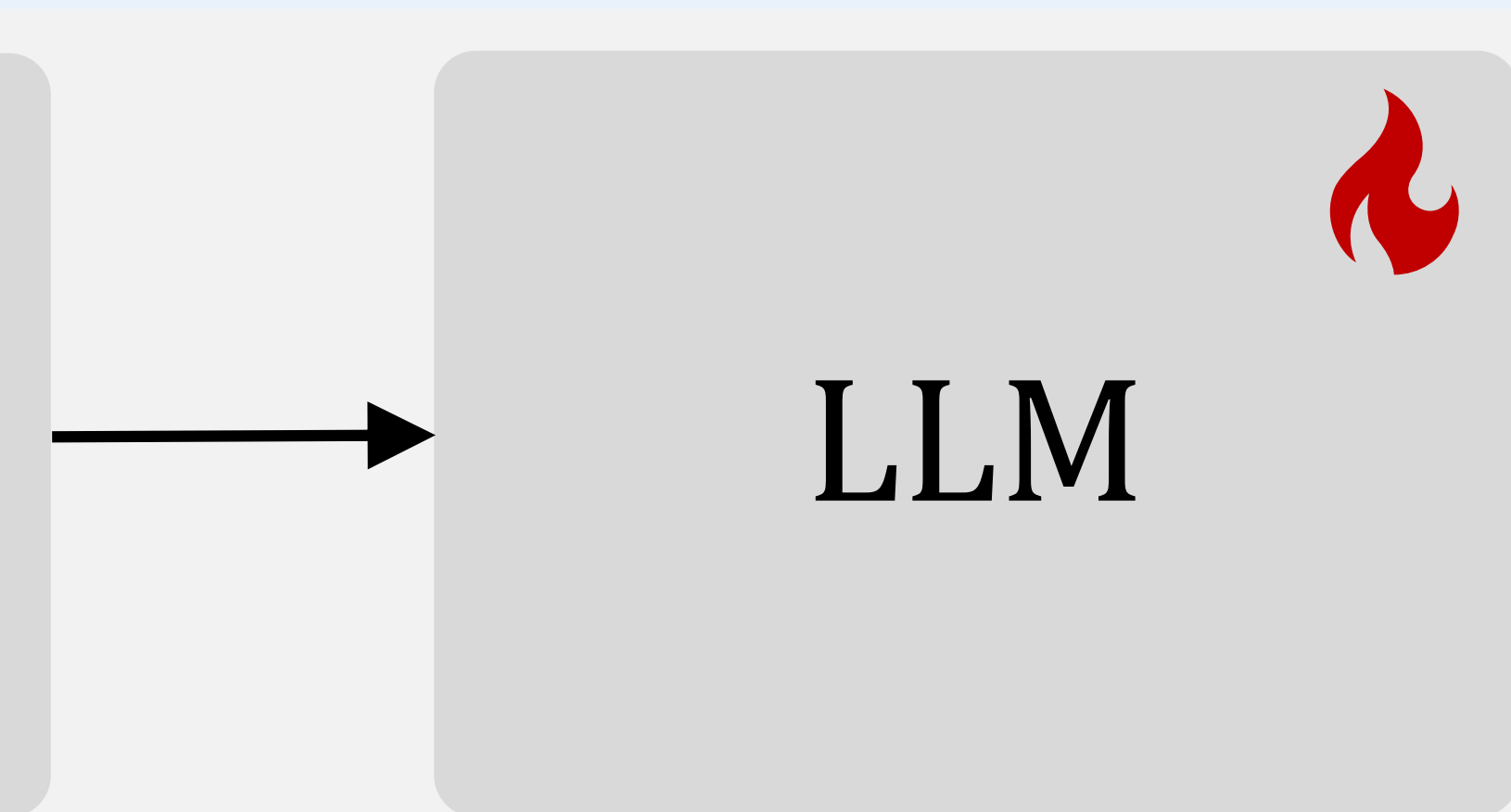
Gesture Embedding



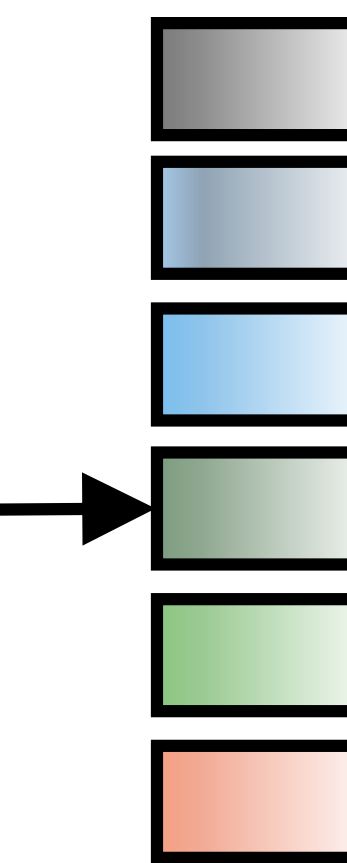
Sign Language Recognition



Stage2:
Pretraining



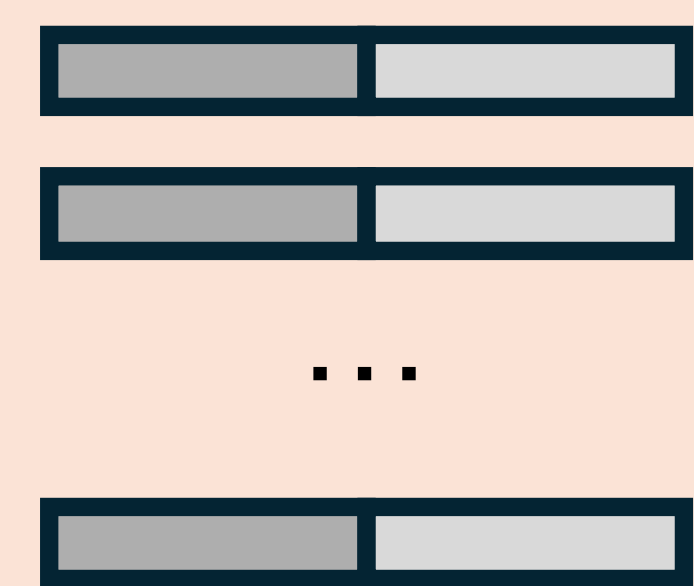
Stage3:
Instruction-tuning



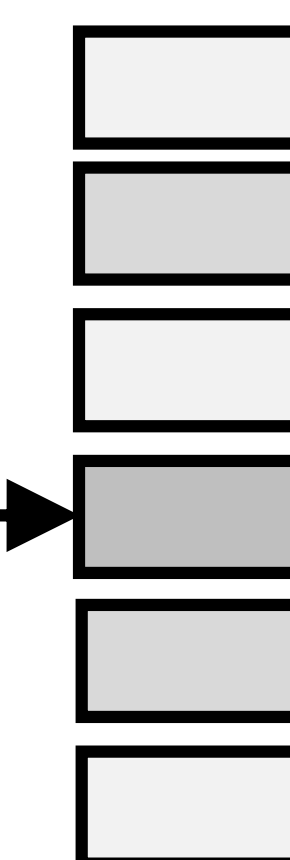
“Next I'm going to show you
how to make the chocolate
really come to life”

“I'm going to show you how to
make a simple chocolate mousse.”

Text Codebook



Language Embedding



“Translate the American Sign Language
represented by <Motion_Placeholder> to English.”

Sign Language Translation Instruction

Sign Language Caption

