PRAGYAMD: PARAMETER-EFFICIENT ADAPTATION OF HUMANOID MOTION GENERATORS TO INDIC LANGUAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

English-centric text encoders can obscure distinctions between Indic instructions before those instructions reach a humanoid motion generator. We address this bottleneck with **PragyaMD**, a **4.7M-parameter adapter** that extends a pretrained, frozen OMG motion diffusion model to Hindi, Bengali, Tamil, and Telugu, *without inference-time translation or generator retraining*.

PragyaMD maps representations from a frozen multilingual encoder into the generator's conditioning space. A single shared adapter is trained on Unitree G1-format trajectories using a motion-pivot objective comprising **(i)** *motion-denoising supervision*, which couples each instruction to its corresponding observed motion continuation; **(ii)** *cross-language agreement*, which encourages compatible conditioning across language views of the same motion; and **(iii)** *English-pathway regularization*, which weakly anchors the learned interface to the model's original English conditioning pathway. Motion thus provides a shared training target across languages; every component except the adapter remains frozen.

On 1,824 leakage-controlled motion groups, PragyaMD improves English-mediated top-1 retrieval from 13.3–14.8% with direct Indic input to **27.5–28.3%** across the four evaluated Indic languages, compared with 26.2–27.2% for an IndicTrans2 translate-then-generate baseline. The translation baseline has 2.4–$2.6\times$ PragyaMD's Fréchet distance in the evaluator's motion space. A matched English-adapter control shows that the Fréchet improvement arises from replacing the conditioning pathway, not from Indic-language adaptation itself. Text-encoder-free interventions reveal consistent forward/backward and speed responses but weak left/right separation; cross-language tests show consistent responses across languages but expose sensitivity to instruction wording. These findings support parameter-efficient language-interface adaptation for extending pretrained humanoid motion models, while showing that retrieval improvements alone do not establish reliable attribute-level instruction grounding.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.