

How Much Do Large Language Models Know about Human Motion?

A Case Study in 3D Avatar Control

Kunhang Li¹ Jason Naradowsky¹ Yansong Feng² Yusuke Miyao^{1,3}

¹The University of Tokyo ²Peking University ³NII LLMC

{kunhangli, narad, yusuke}@is.s.u-tokyo.ac.jp

fengyansong@pku.edu.cn

Abstract

We explore the human motion knowledge of Large Language Models (LLMs) through 3D avatar control. Given a motion instruction, we prompt LLMs to first generate a high-level movement plan with consecutive steps (**High-level Planning**), then specify body part positions in each step (**Low-level Planning**), which we linearly interpolate into avatar animations. Using 20 representative motion instructions that cover fundamental movements and balance body part usage, we conduct comprehensive evaluations, including human and automatic scoring of both high-level movement plans and generated animations, as well as automatic comparison with oracle positions in low-level planning. Our findings show that LLMs are strong at interpreting high-level body movements but struggle with precise body part positioning. While decomposing motion queries into atomic components improves planning, LLMs face challenges in multi-step movements involving high-degree-of-freedom body parts. Furthermore, LLMs provide reasonable approximations for general spatial descriptions, but fall short in handling precise spatial specifications. Notably, LLMs demonstrate promise in conceptualizing creative motions and distinguishing culturally specific motion patterns.¹

1 Introduction

Recent approaches in text-conditioned human motion generation attempt to improve the generalization to unseen instructions by leveraging Large Language Models (LLMs) to extract motion-relevant information, such as active body parts (Athanasios et al., 2023), detailed body part descriptions (Huang et al., 2024), and keyframe coordinates (Huang et al., 2023). However, these methods only utilize LLMs as auxiliary components, leaving the extent of their human motion knowledge largely unexplored.

¹<https://github.com/KunhangL/MotionDecomposition>

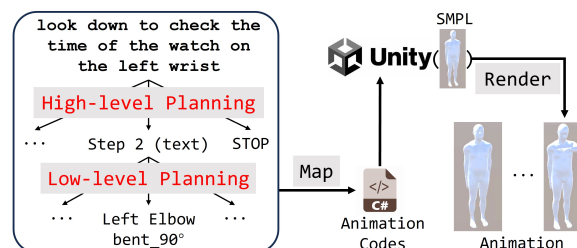


Figure 1: Our motion knowledge grounding pipeline converts the instruction into a high-level movement plan, followed by body part position predictions, which are mapped to animation codes and rendered in Unity.

In this paper, we explore LLMs’ knowledge of human motion through their capabilities to drive a 3D human avatar. Following the natural hierarchy from action sequences to body part movements (Flash and Hochner, 2005), our approach (Figure 1) consists of two stages: (1) **High-level Planning**, where an LLM generates step-by-step body part movements in natural language from the input motion instruction, and (2) **Low-level Planning**, where the LLM selects a position for every body part from a predefined set of poses within each step (e.g., neutral, bent_in_90_degrees for left elbow). These predictions are then converted into animation codes for the Skinned Multi-Person Linear (SMPL; Loper et al., 2015) 3D human model in Unity,² using predefined rules that map body part positions to SMPL joint rotations (e.g., bending left elbow to 90 degrees maps to m_avg_L_Elbow rotation from (0, 0, 0) to (0, 90, 0)). The rendered animations from linearly interpolated LLM-selected poses provide a clear verification lens for human evaluators.

We carefully design 20 representative motion instructions with full coverage of basic movement primitives and balanced body part usage, and evaluate both commercial (e.g., Claude 3.5 Sonnet) and

²<https://unity.com/>

open-source (e.g., Llama-3.1-70B) LLMs through three complementary approaches: (1) human and GPT-4.1-based evaluation of high-level planning feasibility and key movement completeness, assessing the conceptual understanding of motion; (2) quantitative comparison of low-level body part positioning against oracle annotations for scalability and reproducibility; and (3) human and Gemini 2.5 Pro-based judgement of animation quality for both complete motion and individual body parts, capturing multiple valid motion variations with holistic feedback on naturalness.

Firstly, we find that LLMs demonstrate high competence in generating high-level plans with physically proper key movements. However, they struggle with precise body part positioning, especially for multi-step motions involving high-degree-of-freedom body parts (e.g., upper arm). These positioning errors often accumulate across multiple body parts, resulting in low-quality animations.

Secondly, breaking down motion queries into atomic components enhances performance compared to single-round generation. For high-level planning, iterative querying of individual motion aspects (e.g., body movements, states) proves more effective. For low-level planning, hierarchical decomposition and position-by-position selection consistently yield superior results.

Finally, while LLMs provide reasonable approximations for general spatial descriptions (e.g., the bending motion for wiping a one-meter high table), they fail to handle precise spatial specifications in text (e.g., pick up the object by foot), and fall short in generating accurate spatial and temporal parameters for avatar control. However, LLMs show promise in conceptualizing creative motions (e.g., strut like a peacock showing off feathers) and distinguishing culturally-specific motion patterns (e.g., differentiating normal kneel to bow and kneel to perform a Japanese bow), suggesting their potential to provide enhanced semantic understanding when combined with high-quality low-level motion generators from traditional supervised approaches.

2 Related Work

Contemporary generative models show remarkable progress in synthesizing realistic human body movements from natural language instructions (Guo et al., 2022; Tevet et al., 2023; Zhang et al., 2024a; Guo et al., 2024). However, these models often fail on novel motion instructions out of the

limited training datasets, such as compositional instructions, rare activities, or nuanced movement descriptions. To address this generalization problem, recent work uses LLMs to extract specific motion-relevant information, indicating that LLMs might contain rich human motion knowledge.

Athanasiou et al. (2023) use LLMs to identify relevant body parts for action labels like “stroll”, showing LLMs’ understanding of the anatomical requirements for different movements. However, they only focus on simple action-to-body-part mapping without exploring complex motion reasoning capabilities. Later research further prompts LLMs to decompose abstract motion descriptions into sequential, step-by-step movement specifications (Li and Feng, 2024). More advanced approaches leverage LLMs for hierarchical motion planning and control. For instance, CoMo (Huang et al., 2024) and Fg-T2M++ (Wang et al., 2025) employ LLMs to parse ambiguous instructions into structured descriptions targeting specific body parts, enabling fine-grained control over motion generation. More comprehensively, Fan et al. (2024) propose a framework that converts instructions into atomic motion plans organized by predefined body segments, such as spine, left upper limb, etc. Instead of using LLMs for motion-related text generation, recent work also shows that LLMs can directly generate keyframe coordinates to be interpolated as motions (Zhang et al., 2024b; Huang et al., 2023).

The aforementioned approaches focus on leveraging LLMs as auxiliary tools to optimize text-to-motion systems. However, they overlook the fundamental question of what motion knowledge LLMs actually possess and how accurately they understand human movement principles. We address this research gap by grounding LLM responses into 3D avatar animations, and probing their motion knowledge across multiple levels of abstraction.

3 Methodology

This section presents our comprehensive methodology for motion knowledge grounding and evaluation. We introduce a hierarchical pipeline that converts natural language instructions into Unity animations (§3.1). Within the pipeline, we develop systematic querying strategies that guide LLMs through high-level motion decomposition and low-level body part specification (§3.2). Finally, we design a novel evaluation framework to assess LLM capabilities at multiple levels of the pipeline (§3.3).

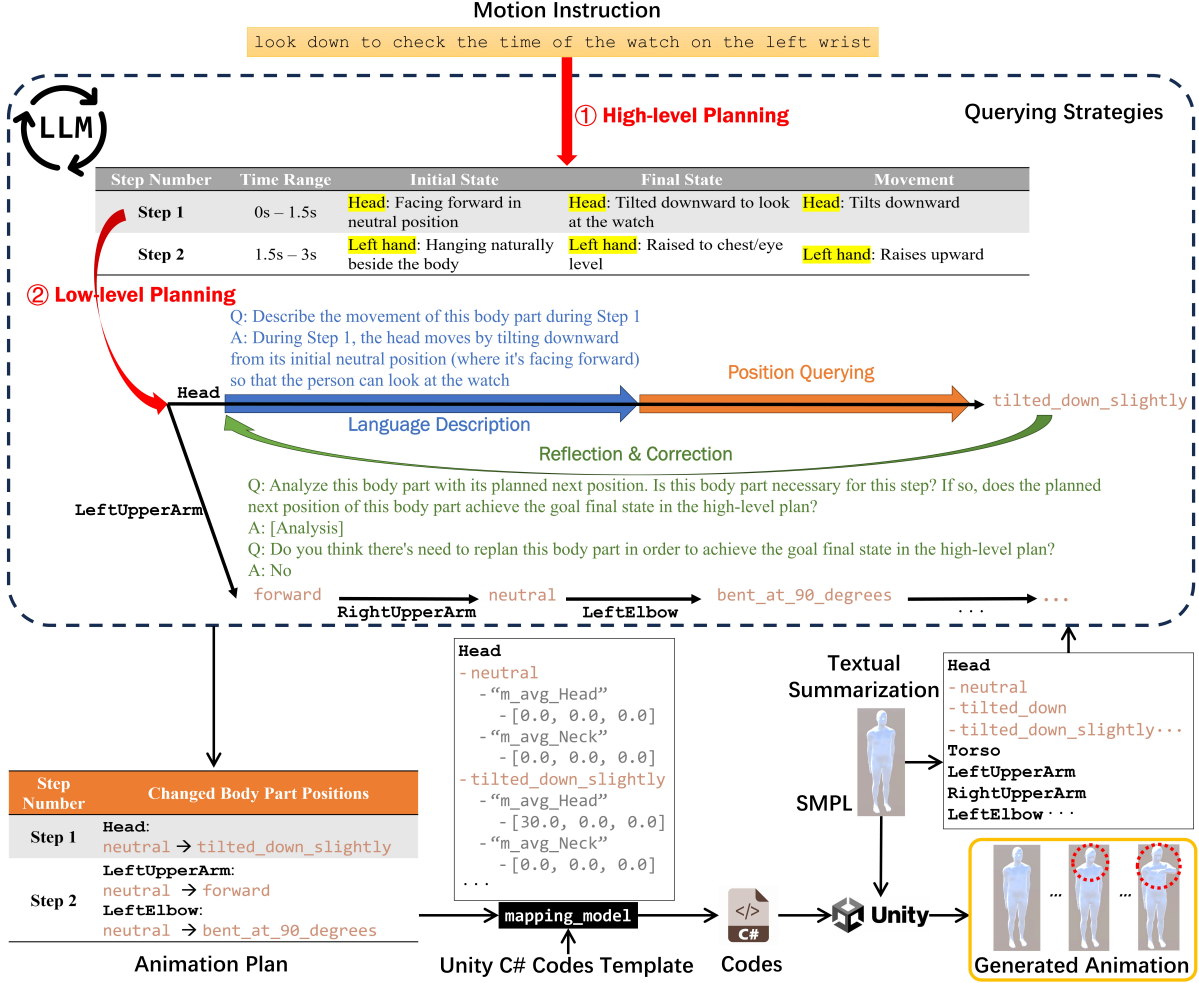


Figure 2: Our motion knowledge grounding pipeline: (1) An LLM processes natural language motion instructions using hierarchical querying strategies to generate an animation plan with specific body part positions, (2) A mapping model converts the animation plan into Unity-compatible codes by transforming body part positions into SMPL joint rotations, (3) Unity renders the final animation by executing the codes.

3.1 Motion Knowledge Grounding Pipeline

Figure 2 illustrates our motion knowledge grounding pipeline using schematic prompts. Firstly, based on the hierarchical joint structure of the SMPL model (e.g., `m_avg_Head`'s child joint is `m_avg_Neck`), we define a finite set of textually summarized positions for preset body parts. For example, we define the body part Head to incorporate `m_avg_Head` and `m_avg_Neck` with available positions like `neutral`, `tilted_down`, `tilted_down_slightly`, etc. We design hierarchical querying strategies that first decompose the input motion instruction into sequential high-level steps, then for each step, iteratively specify body part positions by selecting from the textually summarized positions. These strategies guide the LLM to generate the animation plan.

Secondly, we employ a mapping model to con-

vert the animation plan into Unity codes by transforming the specified body part positions into joint rotations on SMPL, and incorporating them into a code template.

Finally, we render the animation by executing Unity codes on SMPL, where joint rotations are linearly interpolated between consecutive steps.³

3.2 Querying Strategies

High-level Planning. We employ two strategies to translate the instruction into step-by-step descriptions: (1) *piece_by_piece* — querying sequentially about body movements, states, timing, and completion status per step, and (2) *in_one_go* — generating the complete motion plan in one query. This

³We provide the full prompting details in Appendix A.1, and the details of preset body parts and their mapping rules to predefined positions in Appendix A.2.

comparison evaluates the effectiveness of atomic versus holistic planning approaches.

Low-level Planning. To convert high-level plans into low-level body part positions for SMPL mapping, we first prompt the LLM to generate a language description of the queried body part, then implement three approaches: (1) *hierarchical* — querying from large to atomic components (e.g., first determining if elbow is straight or bent, then if bent, specifying slightly, 90 degrees, or fully), (2) *one_by_one* — offering positions sequentially (e.g., first whether straight, then whether slightly bent, etc.), and (3) *all* — presenting all positions simultaneously. These approaches balance structural guidance against decision complexity. We query related body parts sequentially (e.g., LeftUpperArm, RightUpperArm, LeftElbow) for symmetry and employ self-reflection for improved accuracy.

3.3 Evaluation Framework

We propose an evaluation framework aligned with our motion knowledge grounding pipeline, covering high-level planning, low-level planning, and complete animation generation. In the absence of existing metrics for this purpose, we introduce new ones tailored to each level: for high-level plans and animations, human judgement captures nuanced aspects of instruction adherence and naturalness, while multimodal LLMs enable scalable and consistent assessment; for low-level plans, automated metrics provide reproducible evaluation of body part positioning against annotated ground truth. Based on this framework, we design representative instructions to assess LLMs’ motion understanding in diverse scenarios.

High-level Planning. To evaluate LLMs’ abilities to conceptually decompose motion instructions, we introduce **High-level Plan Score (HPS)**, a human-scored 5-point Likert scale metric assessing the physical feasibility and completeness of generated movement plans. Each high-level plan is scored by three independent annotators following rubrics.⁴ Additionally, GPT-4.1 is utilized as the automated evaluator, employing the identical evaluation guidelines provided to the human evaluators, with three evaluations per high-level plan.

Low-level Planning. To consistently evaluate LLMs’ capabilities to specify body part positions, we use fixed high-level plans from GPT-4o with the

in_one_go strategy followed by manual correction. LLMs predict body part positions based on these plans. For scalable and reproducible evaluation, we annotate oracle positions of all body parts across steps, and calculate **Body Part Position Accuracy (BPPA)** by comparing LLM-generated positions against the annotated ones.

Complete Animation Generation. We conduct human evaluation of the complete pipeline’s animations to better accommodate valid motion variations and assess overall naturalness. Five independent annotators rate each animation using: (1) **Whole Body Score (WBS)** — a 5-point Likert scale measuring adherence to the motion instruction, and (2) **Body Part Quality (BPQ)** — classification of six key body parts (Head, Torso, Left Arm, Right Arm, Left Leg, Right Leg) into “Good”, “Partially Good”, “Bad”, or “Not Relevant” categories. We introduce “Not Relevant” to distinguish between motion-critical body parts (e.g., arms during throwing) and those with little involvement in the action (e.g., legs during a standing wave), while still marking any unnatural movement as “Bad”, helping evaluators provide targeted feedback on the quality of key motion components. Oracle animations are evaluated separately to establish performance bounds without biasing annotators.⁵ Additionally, we employ Gemini 2.5 Pro as an automated evaluator using the same criteria provided to humans. Each animation is evaluated five times, with frames sampled at 1-second intervals as image inputs.⁶

Motion Instructions. Given the extensive human evaluation required, we design a focused set of instructions that maximize coverage while remaining feasible for thorough assessment, following HCI practices that emphasize human-in-the-loop assessment through focused, representative examples (Heuer and Buschek, 2021). We create 20 motion instructions fully covering basic primitives from which complex motions can be composed, with balanced coverage across body parts (head: 15, torso: 16, arms: 16 each, legs: 13 each).⁷

⁵Full animation evaluation details are in Appendix A.3.2.

⁶We observe that using full video inputs results in notably low agreement with humans, likely due to the unnatural motion artifacts introduced through linear interpolation.

⁷The details of these motion instructions are provided in Appendix A.4.

⁴Evaluation details are provided in Appendix A.3.1.

4 Results and Analysis

Using the designed motion instructions, we run the motion knowledge grounding pipeline on both commercial and open-source LLMs including Claude 3.5 Sonnet, GPT-4o, GPT-4o-mini, GPT-3.5-turbo and Llama-3.1-70B.⁸ Nine evaluators with artificial intelligence research backgrounds participate in the human evaluation, where we calculate the inter-annotator agreement. For HPS and WBS using Likert scales, we calculate the pairwise weighted kappa (Fleiss and Cohen, 1973). For category-based BPQ, we apply average pairwise agreement, calculated as the mean percentage of matching categories between evaluator pairs.

This section presents our evaluation results and analysis across three key components of the motion knowledge grounding process: high-level planning that decomposes motion descriptions into sequential steps (§4.1), low-level planning that predicts precise body part positions (§4.2), and complete animation generation that synthesizes animations from the instructions (§4.3).

4.1 High-level Planning

As shown in Table 1, while GPT-4.1 tends to give lower scores, the *piece_by_piece* approach consistently outperforms *in_one_go* across LLMs in both human and GPT-4.1 evaluations. Under *piece_by_piece*, Claude 3.5 Sonnet and GPT-4o variants achieve similarly high HPS, while Llama-3.1-70B exceeds GPT-3.5-turbo by a large margin of 0.57. These results suggest that while most LLMs possess sophisticated understanding of high-level body movements, this knowledge is more effectively accessed through incremental guidance rather than one-round generation. To assess the consistency between humans and GPT-4.1 judgments, we compute the average HPS per high-level plan and report a Pearson correlation coefficient of 0.665 ($p = 2.47 \times 10^{-24}$), a Spearman correlation coefficient of 0.549 ($p = 1.49 \times 10^{-15}$), and a Krippendorff’s alpha of 0.653. For the inter-annotator agreement, we report an average kappa of 0.74, indicating substantial agreement based on the interpretation of Landis and Koch (1977).

4.2 Low-level Planning

As shown in Table 2, Claude 3.5 Sonnet and GPT-4o maintain top across three low-level planning

⁸Llama-3.1-8B struggles to follow the output schema, and is easily distracted by body part positions in the chat history. LLM hyperparameters and costs are in Appendix A.5.

LLM	HPS	HPS
	(<i>piece_by_piece</i>)	(<i>in_one_go</i>)
Claude 3.5 Sonnet	4.57 / 4.55	4.42 / 4.53
GPT-4o	4.68 / 4.53	4.55 / 4.28
GPT-4o-mini	4.67 / 4.28	3.93 / 3.73
GPT-3.5-turbo	3.50 / 3.35	3.33 / 3.13
Llama-3.1-70B	4.07 / 3.92	-

Table 1: HPS for each tested LLM across two high-level planning strategies. Each score pair represents the mean HPS rated by human annotators (left) and by GPT-4.1 (right). Llama-3.1-70B is excluded from *in_one_go* due to output schema compliance issues. In addition, we report motion-wise mean scores with standard deviation and variance in Appendix A.8.1.

LLM	BPPA (%)	BPPA (%)	BPPA (%)
	(<i>hierarchical</i>)	(<i>one_by_one</i>)	(<i>all</i>)
Claude 3.5 Sonnet	73.52	71.23	70.75
GPT-4o	70.87	71.70	67.49
GPT-4o-mini	68.10	67.80	65.32
GPT-3.5-turbo	67.19	62.76	21.70
Llama-3.1-70B	52.60	53.34	45.87

Table 2: BPPA across three low-level planning strategies for each LLM. Each value is averaged from two runs. Bold and blue values respectively indicate the highest and second-highest accuracy for the same strategy. In addition, we report motion-wise mean scores with standard deviation and variance in Appendix A.8.2.

strategies. Llama-3.1-70B falls far behind all closed-source LLMs, with large gaps of 9%–15% from GPT-3.5-turbo. Exceptionally, GPT-3.5-turbo receives a low BPPA of 21.70% when running the *all* strategy, possibly due to its limitation in long context understanding.

To examine what drives BPPA differences across LLMs, we visualize body-part-wise BPPA for LLMs using the *hierarchical* strategy in Figure 3. The relative performance of LLMs remains consistent with the overall BPPA shown in Table 2 for most body parts, except that GPT-3.5-turbo shows notably higher BPPA for Knee and Wrist. Upper Arm shows notably low BPPA, possibly due to its high degree-of-freedom. We discuss the correlation between body part complexity and BPPA in §5.2.⁹

For the same LLM, the performance ranking of strategies from highest to lowest (Table 2) is mostly *hierarchical*, *one_by_one*, and *all*, which indicates that increased structural guidance in prompting enables LLMs to generate more precise body part positions, suggesting enhanced utilization of their in-

⁹Additionally, higher-ranked LLMs show higher accuracy in identifying and correcting inaccurate body part positions during self-reflection (Appendix A.6).

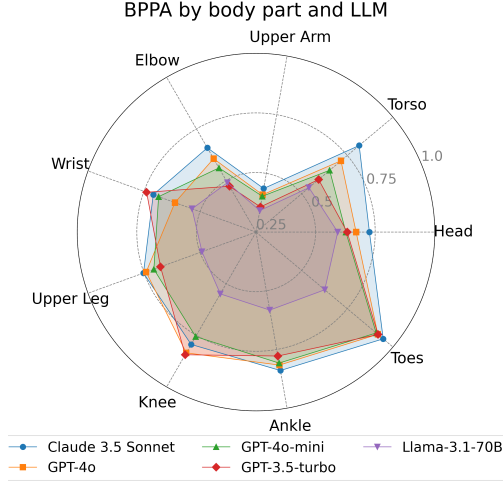


Figure 3: Body-part-wise BPPA for LLMs with the *hierarchical* strategy. BPPA is averaged for paired body parts, e.g., “Elbow” for “LeftElbow” and “RightElbow”.

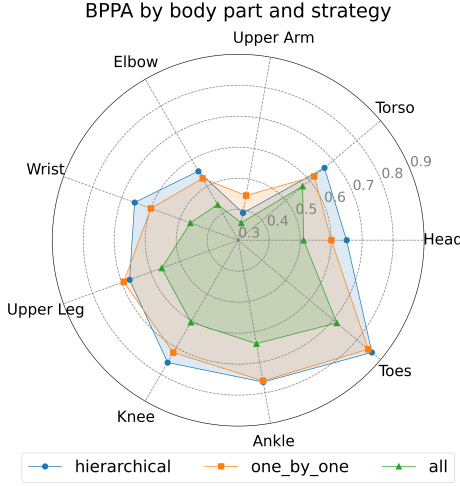


Figure 4: Body-part-wise BPPA for strategies averaged across LLMs. BPPA is averaged for paired body parts.

herent understanding of human motion details. Furthermore, we visualize the body-part-wise BPPA for strategies in Figure 4. The *all* strategy falls far behind across all body parts. The *hierarchical* strategy outperforms *one_by_one* in most body parts, except for high-degree-of-freedom body parts Upper Arm and Upper Leg. This indicates that heuristically defined querying structures might be too rigid compared to more flexible position-by-position selection for complex articulations.

4.3 Complete Animation Generation

We compare LLMs’ complete generation performance under *piece_by_piece* high-level planning and *hierarchical* low-level planning. As shown in Table 3, LLMs mostly maintain their relative rank-

LLM	WBS
Claude 3.5 Sonnet	3.29 / 3.65
GPT-4o	3.13 / 3.22
GPT-4o-mini	2.87 / 2.73
GPT-3.5-turbo	2.14 / 2.20
Llama-3.1-70B	2.13 / 2.29
(Oracle Annotation)	4.57 / 3.97

Table 3: WBS with *piece_by_piece* high-level planning and *hierarchical* low-level planning. Each score pair shows human-rated (left) and Gemini 2.5 Pro-rated (right) WBS. Motion-wise statistics are reported in Appendix A.8.3.

ings from BPPA (Table 2) when evaluated with WBS, except for GPT-3.5-turbo and Llama-3.1-70B’s rankings are reversed in Gemini 2.5 Pro evaluation. While GPT-3.5-turbo largely outperforms Llama-3.1-70B in BPPA, both achieve similarly low WBS, likely because evaluators assign similarly low scores when animation quality falls below a certain threshold. The best performing LLM, Claude 3.5 Sonnet, scores well below the oracle animations, indicating considerable room for motion understanding improvement.

To quantify alignment between human and model evaluations for WBS scoring, we report a Pearson correlation coefficient of 0.585 ($p = 2.24 \times 10^{-12}$), a Spearman correlation coefficient of 0.597 ($p = 6.15 \times 10^{-13}$), and a Krippendorff’s alpha of 0.578. We also observe that Gemini 2.5 Pro tends to produce more conservative and averaged scores compared to human evaluators — it assigns notably lower scores to oracle animations with a difference of 0.6, while assigning comparatively higher scores to LLM-generated motions. For the inter-annotator agreement, WBS achieves a moderate kappa of 0.531.¹⁰ We hypothesize this moderate agreement stems from the inherent uncertainty in human motion, where people can move and express themselves in various valid ways.

A detailed analysis of BPQ reveals distinct performance tendencies across different body parts (Table 4). When comparing the percentages of averaged results among LLMs and oracle animations, head and torso movements demonstrate relatively smaller deficits, while arm and leg motions exhibit notably larger inaccuracies. Among the LLMs, Claude 3.5 Sonnet and GPT-4o consistently achieve higher percentages in the “Good” and “Par-

¹⁰Most human agreement pairs fall in the moderate to substantial agreement range (0.41–0.80), as shown in Figure A7.

LLM	Head			Torso			Left Arm			Right Arm			Left Leg			Right Leg		
	G	PG	B	G	PG	B	G	PG	B	G	PG	B	G	PG	B	G	PG	B
Claude 3.5 Sonnet	74.1	22.2	3.7	72.6	17.7	9.7	25.0	53.9	21.1	29.3	53.3	17.3	38.6	31.8	29.5	31.7	29.3	39.0
GPT-4o	63.8	19.1	17.0	60.7	25.0	14.3	15.2	58.2	26.6	16.9	64.9	18.2	46.8	36.2	17.0	29.5	47.7	22.7
GPT-4o-mini	80.7	8.8	10.5	59.4	28.1	12.5	12.8	47.4	39.7	12.2	52.7	35.1	17.9	33.3	48.7	11.1	33.3	55.6
GPT-3.5-turbo	34.2	13.2	52.6	29.1	16.4	54.5	3.8	41.8	54.4	3.8	46.2	50.0	10.3	30.8	59.0	5.4	18.9	75.7
Llama-3.1-70B	44.0	32.0	24.0	34.8	34.8	30.4	6.9	41.4	51.7	9.4	38.8	51.8	15.5	7.0	77.5	5.9	5.9	88.2
(Average)	59.4	19.0	21.6	51.3	24.4	24.3	12.8	48.6	38.7	14.3	51.2	34.5	25.8	27.8	46.3	16.7	27.0	56.2
(Oracle)	89.6	10.4	0.0	80.3	18.2	1.5	74.0	19.5	6.5	76.3	19.7	4.0	76.6	14.9	8.5	76.1	13.0	10.9

Table 4: Percentage (%) of Body Part Quality (BPQ) from human evaluation after excluding “Not Relevant” across evaluated LLMs. **G**, **PG**, and **B** respectively stand for “Good”, “Partially Good”, and “Bad”. Highest percentages for each category are highlighted in pink (**G**), yellow (**PG**), and gray (**B**).

tially Good” categories, whereas GPT-3.5-turbo and Llama-3.1-70B show higher frequencies in the “Bad” category across all body parts. Furthermore, we find that Gemini 2.5 Pro has limited alignment with nuanced human judgements of body parts.¹¹

5 Discussion

To reveal both the capabilities and limitations of LLMs in each stage of motion knowledge grounding in more details, we analyze the performance patterns in high-level planning, low-level planning, and complete animation generation through quantitative error analysis and representative case studies.

5.1 High-level Planning

To better understand how LLMs perform differently in HPS (Table 1), we count the numbers of high-level plans with wrong or incomplete action descriptions generated using the *piece_by_piece* strategy (Table 5). The ranking of combined error counts aligns with the HPS results. Specifically, Llama-3.1-70B and GPT-3.5-turbo show inferior performance due to higher combined error counts, with GPT-3.5-turbo receiving the lowest HPS primarily due to having the most incomplete plans. Two representative cases demonstrate the two error types: For the motion instruction *lift the right shoe with both hands and put it on in the air*, the wrong plan only specifies lifting the foot while reaching down to grab the shoe, deviating from the intended sequence of lifting the shoe then putting it on (Figure A8). For *look down to check the time of the watch on the left wrist*, the incomplete plan omits the crucial action of positioning the left arm to make the watch visible (Figure A9).

¹¹The agreement scores between Gemini 2.5 Pro and human majority votes for body parts remain below 0.8 (Table A8), far from the expected near-perfect agreement.

LLM	#Wrong	#Incomplete
Claude 3.5 Sonnet	1	5
GPT-4o	1	4
GPT-4o-mini	2	3
GPT-3.5-turbo	4	9
Llama-3.1-70B	5	5

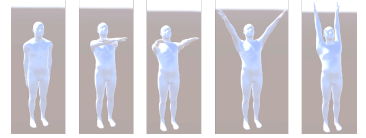
Table 5: Counts of high-level plans with wrong or incomplete action descriptions for each LLM with the *piece_by_piece* strategy.

5.2 Low-level Planning

LLMs typically achieve BPPA between 50% and 75% (Table 2), revealing their limitations in predicting precise body part positions. When observing animations predicted using the *hierarchical* strategy from the low-level planning evaluation, we find that although BPPA is high, there might be critical errors in animations. The example animation (Figure 5a), despite achieving BPPA of 0.7812, wrongly crosses the arms and fails to toss the ball, showing that positioning errors often accumulate across multiple body parts, resulting in low-quality animations. In another example (Figure 5b), despite achieving high BPPA of 0.9688, the animation fails to rotate the shoulder to align the wrist with the face, demonstrating that errors in key articulated joints can severely impact the overall motion quality regardless of high BPPA.

Furthermore, to clearly understand how different body parts and motions affect BPPA, we analyze the correlation between their complexity and BPPA when using the *hierarchical* strategy.

Body Part Complexity. We observe two phenomena in the correlation between the number of possible positions and BPPA for different body parts (Figure A10). First, BPPA tends to inversely correlate with movement flexibility — body parts with more possible positions show lower accuracy



(a) “lean back fully and toss the ball into the air at a 45-degree angle using both hands”



(b) “look down to check the time of the watch on the left wrist”

Figure 5: Key frames at one-second intervals from the example animations predicted by Claude 3.5 Sonnet using the *hierarchical* strategy.

compared to more constrained parts (e.g., upper arm versus upper leg). Second, LLMs demonstrate higher accuracy for lower body parts compared to their upper body counterparts.

Motion Complexity. LLM performance in position prediction declines as motions become more complex with increased steps and body parts (Figure A11). This degradation likely stems from two factors: LLMs’ difficulty in maintaining spatial relationships across extended movement sequences, and training data bias where humans describe only core movements (e.g., “raise arms”) while omitting auxiliary ones (e.g., shoulder and elbow adjustments). Unlike humans who can intuitively infer these auxiliary movements, LLMs appear limited in developing such implicit understanding.

5.3 Complete Animation Generation

We present a comparative case study between the keyframes extracted from animations generated by our pipeline and those produced by MoMask (Guo et al., 2024), a state-of-the-art text-to-motion generative model. This comparison provides insights into how LLMs’ motion understanding differs from specialized motion generative models.¹²

Spatial Precision. LLMs show mixed spatial understanding capabilities. For precise positioning requirements such as picking up objects near feet, LLMs consistently fail where MoMask succeeds (Figure 6a). However, LLMs provide reasonable approximations for less stringent specifications, such as bending to wipe a one-meter high table (Fig-

ure 6b), while MoMask generates generic wiping motions without proper height adaptation, likely because it learns average motion patterns from training data rather than interpreting specific height measurements. LLMs demonstrate solid comprehension of basic directional concepts like left or right, forward or backward, and up or down.

Imagination. We find that LLMs can predict animations from animal imitation instructions, demonstrating sign of creative motion imagination. For instance, GPT-4o successfully predicts a coherent sequence of arm and leg movements that mimic a peacock displaying feathers (Figure 6c). While MoMask also performs adequately on this instruction, its success likely stems from exposure to similar motions in its training data (specifically, humans imitating birds with wing-like movements). Interestingly, Claude 3.5 Sonnet demonstrates more flexible conceptualization by generating plausible woodpecker-like head movements that MoMask completely fails to reproduce (Figure 6d). These observations suggest that LLMs possess a fundamental understanding of motion-related semantics that extends beyond simple pattern matching.

Cultural Awareness. LLMs demonstrate the capability in distinguishing culturally specific motion patterns. For instance, when prompted about Japanese bowing, Claude 3.5 Sonnet correctly generates the formal full-body bow with a kneeling posture and hands positioned on the ground (Figure 6e), differentiating it from a simple knee-bending bow (Figure 6f). In contrast, MoMask produces similar general bowing motions for both instructions, suggesting its limited cultural understanding.

Generating Raw Avatar Control Parameters.

We further examine LLMs’ capabilities to directly generate SMPL control parameters, with prompting strategies illustrated in Appendix A.11. Our analysis reveals three key limitations. First, LLMs generate timing for high-level steps primarily in whole and half seconds (Figure A12, A13), lacking precise temporal control. Second, with GPT-4o as an example, the LLM shows poor comprehension of joint rotations, failing in both directional recognition and movement quantity generation (Figure A14). Third, while the LLM can roughly identify root movement directions (Figure A15), they fail to generate reasonable values, likely because human language rarely specifies body root concepts and precise global movements.

¹²Instructions for spatial precision case study are taken from the main experiments, while the others are newly created.

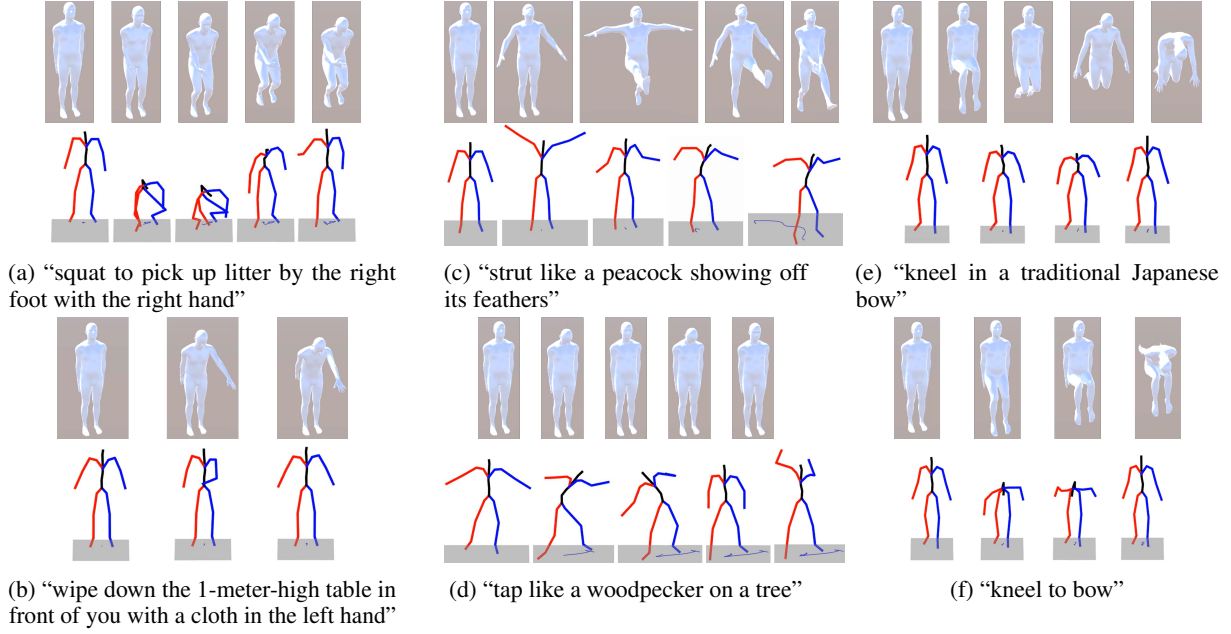


Figure 6: Key frames for example animations reflecting spatial precision (a, b), imagination by animal imitation (c, d), and cultural awareness (e, f). The frames in human figures are generated using our pipeline, while the frames in stick figures are produced by the text-to-motion generative model MoMask (Guo et al., 2024).

6 Conclusion

Through our hierarchical framework for 3D avatar control, we evaluate LLMs’ human motion understanding on carefully designed representative instructions. Our findings reveal that LLMs possess substantial motion-related semantic knowledge despite limitations in precise spatial understanding, particularly for multi-step motions involving high-degree-of-freedom body parts. Breaking down movements into atomic components improves both high-level conceptual planning and low-level body part coordination, yet LLMs fall short in handling precise spatial specifications and generating accurate spatial-temporal parameters for avatar control. Notably, LLMs show promise in conceptualizing creative motions and distinguishing culturally-specific motion patterns.

Practical Value & Future Directions. Our work suggests LLMs could enhance natural language interfaces for avatar control and culturally-aware, creative motion synthesis. To address spatial precision limitations, we propose a hybrid approach: using LLMs for high-level planning and motion decomposition, while delegating low-level execution to specialized motion models (Tevet et al., 2023; Guo et al., 2024) that handle joint dynamics (e.g., velocity, acceleration) and physical constraints (e.g., foot contact, collision). This leverages LLMs’ semantic

strengths with motion models’ spatial precision, enabling interpretable language-based control with enhanced motion realism. While implementing such a system exceeds our current scope of evaluating LLMs’ human motion knowledge, we hope our findings provide insights for future development.

Limitations

Our work, as the first exploration of LLMs’ motion understanding through avatar control, faces two main limitations. First, due to the high workload and cost of human evaluation, we focus on twenty representative motion instructions. While providing initial insights into LLMs’ motion understanding, they are limited in scale. A more comprehensive evaluation would require a larger dataset covering a broader range of motion scenarios and edge cases. Second, while our choice of linear interpolation between keyframes enables clear verification of LLMs’ human motion understanding, it produces mechanical movements not immediately ready for practical applications. Further research could investigate hybrid approaches that combine LLMs’ sophisticated high-level motion understanding with specialized motion synthesis models to generate more natural animations.

Ethics Statement

This work demonstrates both the potential and limitations of using LLMs for human body movement prediction, which has implications for various fields including animation, robotics, and human-computer interaction. While the ability to generate human movements from natural language could democratize animation creation, it also raises potential risks. The technology could be misused to create misleading or deceptive content, particularly in combination with other AI tools for digital human generation. There are also ethical considerations around consent and representation, as such systems could potentially reproduce and amplify biases in human movement patterns. Additionally, as these technologies become more sophisticated, there may be privacy concerns regarding the capture and reproduction of distinctive individual movement styles. Therefore, future development in this area should carefully consider these ethical implications and incorporate appropriate safeguards.

Acknowledgments

This work was partially supported by the “R&D Hub Aimed at Ensuring Transparency and Reliability of Generative AI Models” project of the Ministry of Education, Culture, Sports, Science and Technology. We acknowledge the assistance of Claude 3.5 Sonnet and GPT-4o for manuscript polishing and visualization code development. We would like to thank Bowen Chen for helpful discussion. We would also like to express our appreciation to all annotators for their valuable contributions and dedication to this work.

References

- Nikos Athanasiou, Mathis Petrovich, Michael J Black, and Gül Varol. 2023. Sinc: Spatial composition of 3d human motions for simultaneous action generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9995.
- Ke Fan, Jiangning Zhang, Ran Yi, Jingyu Gong, Yabiao Wang, Yating Wang, Xin Tan, Chengjie Wang, and Lizhuang Ma. 2024. Textual decomposition then sub-motion-space scattering for open-vocabulary motion generation. *arXiv preprint arXiv:2411.04079*.
- Tamar Flash and Binyamin Hochner. 2005. Motor primitives in vertebrates and invertebrates. *Current opinion in neurobiology*, 15(6):660–666.
- Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.
- Chuan Guo, Yuxuan Mu, Muhammad Gohar Javed, Sen Wang, and Li Cheng. 2024. Momask: Generative masked modeling of 3d human motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1900–1910.
- Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. 2022. Generating diverse and natural 3d human motions from text. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5142–5151.
- Hendrik Heuer and Daniel Buschek. 2021. Methods for the design and evaluation of HCI+NLP systems. In *Proceedings of the First Workshop on Bridging Human-Computer Interaction and Natural Language Processing*, pages 28–33, Online. Association for Computational Linguistics.
- Han Huang, Fernanda De La Torre, Cathy Mengying Fang, Andrzej Banburski-Fahey, Judith Amores, and Jaron Lanier. 2023. Real-time animation generation and control on rigged models via large language models. *arXiv preprint arXiv:2310.17838*.
- Yiming Huang, Weilin Wan, Yue Yang, Chris Callison-Burch, Mark Yatskar, and Lingjie Liu. 2024. Como: Controllable motion generation through language guided pose code editing. In *European Conference on Computer Vision*, pages 180–196. Springer.
- J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Kunhang Li and Yansong Feng. 2024. Motion generation from fine-grained textual descriptions. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11625–11641, Torino, Italy. ELRA and ICCL.
- Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16.
- Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. 2023. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations*.
- Yin Wang, Mu Li, Jiapeng Liu, Zhiying Leng, Frederick WB Li, Ziyao Zhang, and Xiaohui Liang. 2025. Fg-t2m++: Llm-augmented fine-grained text driven human motion generation. *International Journal of Computer Vision*, pages 1–17.
- Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. 2024a. Motiondiffuse: Text-driven human motion generation

with diffusion model. *IEEE transactions on pattern analysis and machine intelligence*, 46(6):4115–4128.

Zhikai Zhang, Yitang Li, Haofeng Huang, Mingxian Lin, and Li Yi. 2024b. Freemotion: Mocap-free human motion synthesis with multimodal large language models. In *European Conference on Computer Vision*, pages 403–421. Springer.

A Appendix

A.1 Prompts

Figure A1: The list of prompt templates, with [blue] indicating the prompt type, <orange> as an illustrative label indicating the prompt function (not used in the actual LLM querying), and {red} indicating the placeholder for corresponding contents. {position}, {description} and {positions with descriptions} are all taken from Table A1.

[System Prompt]

You will be given a textual human motion instruction, followed by a sequence of clarification questions about different aspects about the motion. You should use your daily knowledge about human motions to answer the questions accurately and concisely.

[High-level Planning Prompts] (piece_by_piece)

<setup>

The human initially stands naturally with arms hanging beside the body. The textual human motion instruction is "{motion instruction}".

<movement>

What are the movements of relevant body parts in Step{step number}? The movements should be simple enough to be only **single-directional**.

<initial_state>

What are the initial states of relevant body parts in Step{step number}?

<final_state>

What are the final states of relevant body parts in Step{step number}?

<timing>

How long does Step{step number} last in the second unit?

<is_end>

Is it the end of this motion?

[High-level Planning Prompts] (in_one_go)

The human initially stands naturally with arms hanging beside the body. The textual human motion instruction is "{motion instruction}". Decompose it step-by-step with three language descriptions for each step (one for the initial state of moved body parts, one for the final state of moved body parts and one for the movement). Each

step should be simple enough to include only ****single-direction**** motions for all moved body parts. Estimate a time range in the second unit for each step (the end time of the last step should exactly be the start time of the next step).

[Low-level Planning Prompts] (hierarchical)

<step_setup>

The human initially stands naturally with arms hanging beside the body. The textual human motion instruction is "**{motion instruction}**". In the high-level plan of Step**{step number}**, the initial states of relevant body parts are "**{initial states}**", the final states of relevant body parts are "**{final states}**", and the movements of relevant body parts are "**{movements}**".

<language_description>

The last position of **{body part}** is ****{position}**** (**{description}**). Describe the movement of this body part during Step**{step number}** and final position at the end of the step in language.

<position_choice>

Details in Figure A2

<reflection_analysis>

Analyze this body part with its planned next position. Is this body part necessary for this step? If so, does the planned next position of this body part achieve the goal final state in the high-level plan?

<reflection_judgement>

Do you think there's need to replan this body part in order to achieve the goal final state in the high-level plan? Give your judgement.

<correction>

You think that: **{reflection}**. So the next position of **{body part}** should not be ****{position}****. \newline Based on the thought, replan this body part in Step**{step number}**.

[Low-level Planning Prompts] (one_by_one)

<step_setup>

The human initially stands naturally with arms hanging beside the body. The textual human motion instruction is "**{motion instruction}**". In the high-level plan of Step**{step number}**, the initial states of relevant body parts are "**{initial states}**", the final states of relevant body parts are "**{final states}**", and the movements of relevant body parts are "**{movements}**".

<language_description>

The last position of {body part} is **{position}** ({description}). Describe the movement of this body part during Step{step number} and final position at the end of the step in language.

<position_choice>

The last position of {body part} is **{position}** ({description}). Is the next position **{position}** ({description})?

<reflection_analysis>

Analyze this body part with its planned next position. Is this body part necessary for this step? If so, does the planned next position of this body part achieve the goal final state in the high-level plan?

<reflection_judgement>

Do you think there's need to replan this body part in order to achieve the goal final state in the high-level plan? Give your judgement.

<correction>

You think that: {reflection}. So the next position of {body part} should not be **{position}**.\newline Based on the thought, replan this body part in Step{step number}.

[Low-level Planning Prompts] (all)

<step_setup>

The human initially stands naturally with arms hanging beside the body. The textual human motion instruction is "{motion instruction}". In the high-level plan of Step{step number}, the initial states of relevant body parts are "{initial states}", the final states of relevant body parts are "{final states}", and the movements of relevant body parts are "{movements}".

<language_description>

The last position of {body part} is **{position}** ({description}). Describe the movement of this body part during Step{step number} and final position at the end of the step in language.

<position_choice>

There are multiple possible positions for {body part}:
{positions with descriptions}

The last position of this body part is **{position}**. Choose the next position from the options above.

<reflection_analysis>

Analyze this body part with its planned next position. Is this body part necessary for this step? If so, does the planned next

position of this body part achieve the goal final state in the high-level plan?

<reflection_judgement>

Do you think there's need to replan this body part in order to achieve the goal final state in the high-level plan? Give your judgement.

<correction>

You think that: {reflection}. So the next position of {body part} should not be **{position}**.\newline Based on the thought, replan this body part in Step{step number}.

Figure A2: The hierarchical prompts for <position_choice> of [Low-level Planning Prompts] (hierarchical). Each question comes with several options. If the value of the selected option is one position string, the hierarchical querying returns this position. Otherwise the LLM continues to ask the nested question.

```
"Head": {
  "question": "At the end of this step, is the head upright in the neutral position, tilted left, tilted right, tilted down, tilted up, turned left or turned right? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "tilted_left": {
      "question": "Is the head tilted left slightly or fully? Choose one from %s",
      "options": {
        "tilted_left_slightly": "tilted_left_slightly",
        "tilted_left_fully": "tilted_left_fully"
      }
    },
    "tilted_right": {
      "question": "Is the head tilted right slightly or fully? Choose one from %s",
      "options": {
        "tilted_right_slightly": "tilted_right_slightly",
        "tilted_right_fully": "tilted_right_fully"
      }
    },
    "tilted_down": {
      "question": "Is the head tilted down slightly or fully? Choose one from %s",
      "options": {
        "tilted_down_slightly": "tilted_down_slightly",
        "tilted_down_fully": "tilted_down_fully"
      }
    },
    "tilted_up": {
      "question": "Is the head tilted up slightly or fully? Choose one from %s",
      "options": {
        "tilted_up_slightly": "tilted_up_slightly",
        "tilted_up_fully": "tilted_up_fully"
      }
    },
    "turned_left": {
      "question": "Is the head turned left slightly or fully? Choose one from %s",
      "options": {
        "turned_left_slightly": "turned_left_slightly",
        "turned_left_fully": "turned_left_fully"
      }
    },
    "turned_right": {
      "question": "Is the head turned right slightly or fully? Choose one from %s",
      "options": {
        "turned_right_slightly": "turned_right_slightly",
        "turned_right_fully": "turned_right_fully"
      }
    }
  }
},
"Torso": {
  "question": "At the end of this step, is the torso upright in the neutral position, bent forward, bent backward, tilted left, tilted right, twisted left or twisted right? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "bent_backward": "bent_backward",
    "bent_forward": {
      "question": "Is the torso bent forward slightly or fully? Choose one from %s",
      "options": {
        "bent_forward_slightly": "bent_forward_slightly",
        "bent_forward_fully": "bent_forward_fully"
      }
    }
  }
}
```

```

    }
  },
  "tilted_left": {
    "question": "Is the torso tilted left slightly or fully? Choose one from %s",
    "options": {
      "tilted_left_slightly": "tilted_left_slightly",
      "tilted_left_fully": "tilted_left_fully"
    }
  },
  "tilted_right": {
    "question": "Is the torso tilted right slightly or fully? Choose one from %s",
    "options": {
      "tilted_right_slightly": "tilted_right_slightly",
      "tilted_right_fully": "tilted_right_fully"
    }
  },
  "twisted_left": {
    "question": "Is the torso twisted left slightly or fully? Choose one from %s",
    "options": {
      "twisted_left_slightly": "twisted_left_slightly",
      "twisted_left_fully": "twisted_left_fully"
    }
  },
  "twisted_right": {
    "question": "Is the torso twisted right slightly or fully? Choose one from %s",
    "options": {
      "twisted_right_slightly": "twisted_right_slightly",
      "twisted_right_fully": "twisted_right_fully"
    }
  }
}
},
"LeftUpperArm": {
  "question": "At the end of this step, relative to the torso, is the left upper arm neutrally resting by the side of the body, straight upward, straight forward, straight out to the side forming a right angle with the torso, or in other in-between positions? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "upward": "upward",
    "forward": {
      "question": "Is the left elbow pit facing inward to the midline of the body or upward relative to the torso? Choose one from %s",
      "options": {
        "forward_elbowpit_inward": "forward_elbowpit_inward",
        "forward_elbowpit_upward": "forward_elbowpit_upward"
      }
    },
    "side": {
      "question": "Is the left elbow pit facing forward or upward relative to the torso? Choose one from %s",
      "options": {
        "side_elbowpit_forward": "side_elbowpit_forward",
        "side_elbowpit_upward": "side_elbowpit_upward"
      }
    },
    "in_between_positions": {
      "question": "Relative to the torso, is the left upper arm out to the side, towards the midline of the body, or neither? Choose one from %s",
      "options": {
        "out_to_side": {
          "question": "Relative to the torso, is the left upper arm in front of the body, behind the body, or neither? Choose one from %s",
          "options": {
            "front": {
              "question": "Relative to the torso, is the left upper arm above the left shoulder, below the left shoulder, or neither? Choose one from %s",
              "options": {
                "above": "forward_to_upward_side",
                "below": "neutral_to_forward_side",
                "neither": "forward_to_side"
              }
            },
            "behind": {
              "question": "Relative to the torso, is the left upper arm above the left shoulder, below the left shoulder, or neither? Choose one from %s",
              "options": {
                "above": "side_to_upward_back",
                "below": "neutral_to_backward_side",
                "neither": "side_to_back"
              }
            },
            "neither": {
              "question": "Relative to the torso, is the left upper arm above the left shoulder or below the left shoulder? Choose one from %s",
              "options": {
                "above": "upward_to_side",
                "below": "neutral_to_side"
              }
            }
          }
        }
      }
    }
  }
}

```

```

    }
  }
},
"towards_midline": {
  "question": "Relative to the torso, is the left upper arm above the left shoulder, below the left shoulder, or neither? Choose one from %s",
  "options": {
    "above": "forward_to_upward_midline",
    "below": "neutral_to_forward_midline",
    "neither": "forward_to_midline"
  }
},
"neither": {
  "question": "Relative to the torso, is the left upper arm in front of the body or behind the body? Choose one from %s",
  "options": {
    "front": {
      "question": "Relative to the torso, is the left upper arm above the left shoulder or below the left shoulder? Choose one from %s",
      "options": {
        "above": "forward_to_upward",
        "below": "neutral_to_forward"
      }
    },
    "behind": "neutral_to_back"
  }
}
}
}
}
},
"RightUpperArm": {
  "question": "At the end of this step, relative to the torso, is the right upper arm neutrally resting by the side of the body, straight upward, straight forward, straight out to the side forming a right angle with the torso, or in other in-between positions? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "upward": "upward",
    "forward": {
      "question": "Is the right elbow pit facing inward to the midline of the body or upward relative to the torso? Choose one from %s",
      "options": {
        "forward_elbowpit_inward": "forward_elbowpit_inward",
        "forward_elbowpit_upward": "forward_elbowpit_upward"
      }
    },
    "side": {
      "question": "Is the right elbow pit facing forward or upward relative to the torso? Choose one from %s",
      "options": {
        "side_elbowpit_forward": "side_elbowpit_forward",
        "side_elbowpit_upward": "side_elbowpit_upward"
      }
    },
    "in_between_positions": {
      "question": "Relative to the torso, is the right upper arm out to the side, towards the midline of the body, or neither? Choose one from %s",
      "options": {
        "out_to_side": {
          "question": "Relative to the torso, is the right upper arm in front of the body, behind the body, or neither? Choose one from %s",
          "options": {
            "front": {
              "question": "Relative to the torso, is the right upper arm above the right shoulder, below the right shoulder, or neither? Choose one from %s",
              "options": {
                "above": "forward_to_upward_side",
                "below": "neutral_to_forward_side",
                "neither": "forward_to_side"
              }
            },
            "behind": {
              "question": "Relative to the torso, is the right upper arm above the right shoulder, below the right shoulder, or neither? Choose one from %s",
              "options": {
                "above": "side_to_upward_back",
                "below": "neutral_to_backward_side",
                "neither": "side_to_back"
              }
            }
          }
        },
        "neither": {
          "question": "Relative to the torso, is the right upper arm above the right shoulder or below the right shoulder? Choose one from %s",
          "options": {
            "above": "upward_to_side",
            "below": "neutral_to_side"
          }
        }
      }
    }
  }
}

```

```

    }
  }
},
"towards_midline": {
  "question": "Relative to the torso, is the right upper arm above the right shoulder, below the right shoulder, or neither? Choose one from %s",
  "options": {
    "above": "forward_to_upward_midline",
    "below": "neutral_to_forward_midline",
    "neither": "forward_to_midline"
  }
},
"neither": {
  "question": "Relative to the torso, is the right upper arm in front of the body or behind the body? Choose one from %s",
  "options": {
    "front": {
      "question": "Relative to the torso, is the right upper arm above the right shoulder or below the right shoulder? Choose one from %s",
      "options": {
        "above": "forward_to_upward",
        "below": "neutral_to_forward"
      }
    },
    "behind": "neutral_to_back"
  }
}
}
}
},
"LeftElbow": {
  "question": "At the end of this step, is the left elbow straight or bent? Choose one from %s",
  "options": {
    "straight": "neutral",
    "bent": {
      "question": "Is the left elbow slightly bent in, bent in 90 degrees or fully bent? Choose one from %s",
      "options": {
        "slightly_bent_in": "slightly_bent_in",
        "bent_in_90_degrees": "bent_in_90_degrees",
        "fully_bent": "fully_bent"
      }
    }
  }
},
"RightElbow": {
  "question": "At the end of this step, is the right elbow straight or bent? Choose one from %s",
  "options": {
    "straight": "neutral",
    "bent": {
      "question": "Is the right elbow slightly bent in, bent in 90 degrees or fully bent? Choose one from %s",
      "options": {
        "slightly_bent_in": "slightly_bent_in",
        "bent_in_90_degrees": "bent_in_90_degrees",
        "fully_bent": "fully_bent"
      }
    }
  }
},
"LeftWrist": {
  "question": "At the end of this step, is the left wrist straight in the neutral position, bent vertically, or tilted sideways? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "bent_vertically": {
      "question": "Is the left wrist bent upward so that the back of the left hand is closer to the back of the left forearm, with the muscles of the back of the left forearm contracted? Or is the left wrist bent downward so that the left palm moves towards the left forearm? Choose one from %s",
      "options": {
        "bent_upward": "bent_upward",
        "bent_downward": {
          "question": "Is the left wrist slightly bent downward or fully? Choose one from %s",
          "options": {
            "bent_slightly_downward": "bent_slightly_downward",
            "fully_bent_downward": "fully_bent_downward"
          }
        }
      }
    }
  }
},
"tilted_sideways": {
  "question": "Is the left wrist tilted laterally towards the left thumb, or the left little finger? Choose one from %s",
  "options": {

```

```

        "tilted_towards_thumb_side": "tilted_towards_thumb_side",
        "tilted_towards_pinky_side": "tilted_towards_pinky_side"
    }
}
},
"RightWrist": {
    "question": "At the end of this step, is the right wrist straight in the neutral position, bent vertically, or tilted sideways? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "bent_vertically": {
            "question": "Is the right wrist bent upward so that the back of the right hand is closer to the back of the right forearm, with the muscles of the back of the right forearm contracted? Or is the right wrist bent downward so that the right palm moves towards the right forearm? Choose one from %s",
            "options": {
                "bent_upward": "bent_upward",
                "bent_downward": {
                    "question": "Is the right wrist slightly bent downward or fully? Choose one from %s",
                    "options": {
                        "bent_slightly_downward": "bent_slightly_downward",
                        "fully_bent_downward": "fully_bent_downward"
                    }
                }
            }
        },
        "tilted_sideways": {
            "question": "Is the right wrist tilted laterally towards the right thumb, or the right little finger? Choose one from %s",
            "options": {
                "tilted_towards_thumb_side": "tilted_towards_thumb_side",
                "tilted_towards_pinky_side": "tilted_towards_pinky_side"
            }
        }
    }
},
"LeftUpperLeg": {
    "question": "At the end of this step, is the left upper leg neutrally aligned with the body midline, straight forward, straight out to the side forming a right angle with the torso, or in other in-between positions? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "forward": "forward",
        "side": "side",
        "in_between_positions": {
            "question": "Relative to the torso, is the left upper leg out to the side, towards the midline of the body, or neither? Choose one from %s",
            "options": {
                "out_to_side": {
                    "question": "Relative to the torso, is the left upper leg in front of the body, behind the body, or neither? Choose one from %s",
                    "options": {
                        "front": {
                            "question": "Relative to the torso, is the left upper leg above the left pelvis, below the left pelvis, or neither? Choose one from %s",
                            "options": {
                                "above": "forward_to_upward_side",
                                "below": "neutral_to_forward_side",
                                "neither": "forward_to_side"
                            }
                        },
                        "behind": "neutral_to_backward_side",
                        "neither": "neutral_to_side"
                    }
                },
                "towards_midline": {
                    "question": "Relative to the torso, is the left upper leg in front of the body or behind the body? Choose one from %s",
                    "options": {
                        "front": {
                            "question": "Relative to the torso, is the left upper leg above the left pelvis, below the left pelvis, or neither? Choose one from %s",
                            "options": {
                                "above": "forward_to_upward_midline",
                                "below": "neutral_to_forward_midline",
                                "neither": "forward_to_midline"
                            }
                        },
                        "behind": "neutral_to_backward_midline"
                    }
                }
            }
        },
        "neither": {
            "question": "Relative to the torso, is the left upper leg in front of the body or behind the body? Choose one from %s",
            "options": {
                "front": {

```

```

        "question": "Relative to the torso, is the left upper leg above the left
        pelvis or below the left pelvis? Choose one from %s",
        "options": {
            "above": "forward_to_upward",
            "below": "neutral_to_forward"
        }
    },
    "behind": "neutral_to_backward"
}
}
}
}
}
},
"RightUpperLeg": {
    "question": "At the end of this step, is the right upper leg neutrally aligned with the body midline,
    straight forward, straight out to the side forming a right angle with the torso, or in other in-
    between positions? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "forward": "forward",
        "side": "side",
        "in_between_positions": {
            "question": "Relative to the torso, is the right upper leg out to the side, towards the
            midline of the body, or neither? Choose one from %s",
            "options": {
                "out_to_side": {
                    "question": "Relative to the torso, is the right upper leg in front of the body,
                    behind the body, or neither? Choose one from %s",
                    "options": {
                        "front": {
                            "question": "Relative to the torso, is the right upper leg above the right
                            pelvis, below the right pelvis, or neither? Choose one from %s",
                            "options": {
                                "above": "forward_to_upward_side",
                                "below": "neutral_to_forward_side",
                                "neither": "forward_to_side"
                            }
                        },
                        "behind": "neutral_to_backward_side",
                        "neither": "neutral_to_side"
                    }
                },
                "towards_midline": {
                    "question": "Relative to the torso, is the right upper leg in front of the body or
                    behind the body? Choose one from %s",
                    "options": {
                        "front": {
                            "question": "Relative to the torso, is the right upper leg above the right
                            pelvis, below the right pelvis, or neither? Choose one from %s",
                            "options": {
                                "above": "forward_to_upward_midline",
                                "below": "neutral_to_forward_midline",
                                "neither": "forward_to_midline"
                            }
                        },
                        "behind": "neutral_to_backward_midline"
                    }
                },
                "neither": {
                    "question": "Relative to the torso, is the right upper leg in front of the body or
                    behind the body? Choose one from %s",
                    "options": {
                        "front": {
                            "question": "Relative to the torso, is the right upper leg above the right
                            pelvis or below the right pelvis? Choose one from %s",
                            "options": {
                                "above": "forward_to_upward",
                                "below": "neutral_to_forward"
                            }
                        },
                        "behind": "neutral_to_backward"
                    }
                }
            }
        }
    },
    "straight": "neutral",
    "bent": {
        "question": "Is the left knee slightly bent, bent at 90 degrees or fully bent? Choose one from
        %s",
        "options": {
            "slightly_bent": "slightly_bent",

```

```

        "bent_at_90_degrees": "bent_at_90_degrees",
        "fully_bent": "fully_bent"
    }
}
},
"RightKnee": {
    "question": "At the end of this step, is the right knee stright or bent? Choose one from %s",
    "options": {
        "straight": "neutral",
        "bent": {
            "question": "Is the right knee slightly bent, bent at 90 degrees or fully bent? Choose one from %s",
            "options": {
                "slightly_bent": "slightly_bent",
                "bent_at_90_degrees": "bent_at_90_degrees",
                "fully_bent": "fully_bent"
            }
        }
    }
},
"LeftAnkle": {
    "question": "At the end of this step, is the left ankle in a neutral position like when standing? Or is the left ankle bent vertically? Or is the left ankle tilted so that the sole moves towards the side or midline? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "bent_vertically": {
            "question": "Is the left ankle bent upward so that the top of the left foot forms an acute angle with the left shin? Or is the left ankle bent downward so that the top of the left foot forms an obtuse angle with the left shin? Choose one from %s",
            "options": {
                "bent_upward": "bent_upward",
                "bent_downward": "bent_downward"
            }
        },
        "tilted_inward_or_outward": {
            "question": "Is the left ankle tilted inwards so that the left sole moves towards the midline of the body? Or is the left ankle tilted outwards so that the left sole moves away from the midline of the body? Choose one from %s",
            "options": {
                "tilted_inward": "tilted_inward",
                "tilted_outward": "tilted_outward"
            }
        }
    }
},
"RightAnkle": {
    "question": "At the end of this step, is the right ankle in a neutral position like when standing? Or is the right ankle bent vertically? Or is the right ankle tilted so that the sole moves towards the side or midline? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "bent_vertically": {
            "question": "Is the right ankle bent upward so that the top of the right foot forms an acute angle with the right shin? Or is the right ankle bent downward so that the top of the right foot forms an obtuse angle with the right shin? Choose one from %s",
            "options": {
                "bent_upward": "bent_upward",
                "bent_downward": "bent_downward"
            }
        },
        "tilted_inward_or_outward": {
            "question": "Is the right ankle tilted inwards so that the right sole moves towards the midline of the body? Or is the right ankle tilted outwards so that the right sole moves away from the midline of the body? Choose one from %s",
            "options": {
                "tilted_inward": "tilted_inward",
                "tilted_outward": "tilted_outward"
            }
        }
    }
},
"LeftToes": {
    "question": "At the end of this step, are the left toes in the neutral position or curled? Choose one from %s",
    "options": {
        "neutral": "neutral",
        "curled": {
            "question": "Are the left toes curled up or down? Choose one from %s",
            "options": {
                "curled_up": "curled_up",
                "curled_down": "curled_down"
            }
        }
    }
}
},

```

```
"RightToes": {
  "question": "At the end of this step, are the right toes in the neutral position or curled? Choose one from %s",
  "options": {
    "neutral": "neutral",
    "curled": {
      "question": "Are the right toes curled up or down? Choose one from %s",
      "options": {
        "curled_up": "curled_up",
        "curled_down": "curled_down"
      }
    }
  }
}
```

Table A1: The language descriptions of body part positions.

Body Part	Position	Description
Head	neutral	The head is level with the spine and faces forward relative to the torso. The chin is neither raised nor lowered, forming a right angle (90 degrees) with the neck. The back of the head is aligned with the upper back, maintaining a straight, neutral posture.
	tilted_up_slightly	The chin forms a slightly obtuse angle, slightly more than 90 degrees, with the neck. The back of the head is slightly closer to the upper back, making the angle between the back of the head and the upper back slightly more than 90 degrees.
	tilted_up_fully	The chin forms a significantly obtuse angle with the neck. The back of the head almost touches the upper back.
	tilted_down_slightly	The chin forms a slightly acute angle, slightly less than 90 degrees, with the neck.
	tilted_down_fully	The chin nearly touches the chest, forming a very acute angle, often less than 45 degrees, with the neck.
	turned_left_slightly	The nose points just a bit to the left of the body's midline. The angle between the nose's direction and the body's central axis is small, resulting in a slight turn.
	turned_left_fully	The nose points directly towards or over the left shoulder, forming a nearly right angle of approximately 90 degrees with the body's midline.
	turned_right_slightly	The nose points just a bit to the right of the body's midline. The angle between the nose's direction and the body's central axis is small, resulting in a slight turn.
	turned_right_fully	The nose points directly towards or over the right shoulder, forming a nearly right angle of approximately 90 degrees with the body's midline.
	tilted_left_slightly	The left ear forms a small angle, around 45 degrees, with the left shoulder. The right ear forms a larger, obtuse angle with the right shoulder.
	tilted_left_fully	The left ear nearly touches the left shoulder, forming an angle smaller than 20 degrees. The right ear forms an almost straight angle, nearing 180 degrees, with the right shoulder.
	tilted_right_slightly	The right ear forms a small angle, around 45 degrees, with the right shoulder. The left ear forms a larger, obtuse angle with the left shoulder.
	tilted_right_fully	The right ear nearly touches the right shoulder, forming an angle smaller than 20 degrees. The left ear forms an almost straight angle, nearing 180 degrees, with the left shoulder.
Torso	neutral	The waist is in an upright position, aligned with the spine. The pelvis and torso maintain a straight posture.
	twisted_left_slightly	The left side of the waist moves slightly backward, and the right side moves slightly forward, forming a small angle to the left relative to the feet. The muscles on the right side extend mildly, while those on the left side contract slightly.
	twisted_left_fully	The left side of the waist moves backward, and the right side moves forward, causing a rotational angle to the left relative to the feet. The muscles on the right side of the lower back and abdomen extend slightly, while the muscles on the left side contract lightly.
	twisted_right_slightly	The right side of the waist moves slightly backward, and the left side moves slightly forward, forming a small angle to the right relative to the feet. The muscles on the left side extend mildly, while those on the right side contract slightly.
	twisted_right_fully	The right side of the waist moves backward, and the left side moves forward, causing a rotational angle to the right relative to the feet. The muscles on the left side of the lower back and abdomen extend slightly, while the muscles on the right side contract lightly.
	bent_forward_slightly	The waist forms a slight angle, around 45 degrees, with the thighs. The lower abdomen is closer to the thighs, and the muscles in the lower back lengthen slightly while those in the lower abdomen contract a little.
	bent_forward_fully	The waist forms a right angle, around 90 degrees, with the thighs. The muscles in the lower back stretch significantly, while those in the abdomen contract fully.
	bent_backward	The waist forms a pronounced outward curve, creating a significant arch in the lower back. The front side angle of the waist increases markedly, with the muscles in the lower back fully contracted and those in front maximally extended.
	tilted_left_slightly	The left side of the waist moves slightly downward, and the right side is slightly raised, forming a small angle. The right side muscles extend a bit, while the left muscles contract slightly.
	tilted_left_fully	The left side of the waist moves downward, forming a right angle with the pelvis, with the right side raised. The right body muscles stretch, while the left muscles contract.
	tilted_right_slightly	The right side of the waist moves slightly downward, and the left side is slightly raised, forming a small angle. The left side muscles extend a bit, while the right muscles contract slightly.
	tilted_right_fully	The right side of the waist moves downward, forming a right angle with the pelvis, with the left side raised. The left body muscles stretch, while the right muscles contract.
LeftUpperArm	neutral	The left upper arm is relaxed and hanging straight down by the side of the body, parallel to the torso.
	forward_elbowpit_inward	The left upper arm is extended straight forward relative to the torso, parallel to the ground and perpendicular to the torso. The elbow pit faces inward to the midline of the body.
	forward_elbowpit_upward	The left upper arm is extended straight forward relative to the torso, parallel to the ground and perpendicular to the torso. The elbow pit faces upward relative to the torso.
	upward	The left upper arm is lifted straight upwards, close to the ear, reaching towards the sky.
	side_elbowpit_forward	The left upper arm is extended straight out to the side, forming a right angle with the torso (horizontally aligned with the shoulders). The elbow pit faces forward.
	side_elbowpit_upward	The left upper arm is extended straight out to the side, forming a right angle with the torso (horizontally aligned with the shoulders). The elbow pit faces upward.
	neutral_to_forward	The left upper arm is raised from the neutral position towards the forward position, forming an approximate 45-degree angle in front of the body.
	forward_to_upward	The left upper arm is raised from the forward position towards the upward position, forming an approximate 45-degree angle from forward towards upwards.
	neutral_to_back	The left upper arm is raised from the neutral position towards the back, forming an approximate 45-degree angle behind the body.
	forward_to_midline	The left upper arm is raised forward and slightly crosses the midline of the body.
	forward_to_side	The left upper arm is extended from the forward position towards the side, forming an approximate 45-degree angle between forward and side.
	side_to_back	The left upper arm is extended from the side position towards the back, forming an approximate 45-degree angle between side and back.
	neutral_to_side	The left upper arm is in the neutral position but extends halfway outward to the side, forming a small angle from the torso.
	upward_to_side	The left upper arm is extended outward from the upwards position towards the side, forming an approximate 45-degree angle between upwards and side.
	forward_to_upward_side	The left upper arm is raised upwards and slightly angled towards the side from the forward position, forming a diagonal line above the shoulder.

Body Part	Position	Description
	forward_to_upward_midline	The left upper arm is raised upwards and slightly angled towards the midline from the forward position, forming a diagonal line above the shoulder.
	side_to_upward_back	The left upper arm is raised from the side position towards the back and upwards, forming a diagonal reaching upwards and backward.
	neutral_to_forward_midline	The left upper arm is raised upwards and slightly angled towards the midline from the neutral position, forming a diagonal line reaching up.
	neutral_to_forward_side	The left upper arm is raised upwards and slightly angled towards the side from the neutral position, forming a diagonal line reaching up.
	neutral_to_backward_side	The left upper arm is raised upwards and slightly angled towards the back and side from the neutral position, forming a diagonal line reaching up and behind the torso.
RightUpperArm	neutral	The right upper arm is relaxed and hanging straight down by the side of the body, parallel to the torso.
	forward_elbowpit_inward	The right upper arm is extended straight forward relative to the torso, parallel to the ground and perpendicular to the torso. The elbow pit faces inward to the midline of the body.
	forward_elbowpit_upward	The right upper arm is extended straight forward relative to the torso, parallel to the ground and perpendicular to the torso. The elbow pit faces upward relative to the torso.
	upward	The right upper arm is lifted straight upwards, close to the ear, reaching towards the sky.
	side_elbowpit_forward	The right upper arm is extended straight out to the side, forming a right angle with the torso (horizontally aligned with the shoulders). The elbow pit faces forward.
	side_elbowpit_upward	The right upper arm is extended straight out to the side, forming a right angle with the torso (horizontally aligned with the shoulders). The elbow pit faces upward.
	neutral_to_forward	The right upper arm is raised from the neutral position towards the forward position, forming an approximate 45-degree angle in front of the body.
	forward_to_upward	The right upper arm is raised from the forward position towards the upward position, forming an approximate 45-degree angle from forward towards upwards.
	neutral_to_back	The right upper arm is raised from the neutral position towards the back, forming an approximate 45-degree angle behind the body.
	forward_to_midline	The right upper arm is raised forward and slightly crosses the midline of the body.
	forward_to_side	The right upper arm is extended from the forward position towards the side, forming an approximate 45-degree angle between forward and side.
	side_to_back	The right upper arm is extended from the side position towards the back, forming an approximate 45-degree angle between side and back.
	neutral_to_side	The right upper arm is in the neutral position but extends halfway outward to the side, forming a small angle from the torso.
	upward_to_side	The right upper arm is extended outward from the upwards position towards the side, forming an approximate 45-degree angle between upwards and side.
	forward_to_upward_side	The right upper arm is raised upwards and slightly angled towards the side from the forward position, forming a diagonal line above the shoulder.
	forward_to_upward_midline	The right upper arm is raised upwards and slightly angled towards the midline from the forward position, forming a diagonal line above the shoulder.
	side_to_upward_back	The right upper arm is raised from the side position towards the back and upwards, forming a diagonal reaching upwards and backward.
	neutral_to_forward_midline	The right upper arm is raised upwards and slightly angled towards the midline from the neutral position, forming a diagonal line reaching up.
	neutral_to_forward_side	The right upper arm is raised upwards and slightly angled towards the side from the neutral position, forming a diagonal line reaching up.
	neutral_to_backward_side	The right upper arm is raised upwards and slightly angled towards the back and side from the neutral position, forming a diagonal line reaching up and behind the torso.
LeftElbow	neutral	The left elbow is extended naturally, forming a straight line from the left shoulder to the left wrist. The left upper arm and left forearm create a nearly straight alignment of about 180 degrees.
	slightly_bent_in	The left forearm forms a slightly obtuse angle with the left upper arm. The left hand moves slightly closer to the left elbow; the muscles in the left upper arm contract slightly.
	bent_in_90_degrees	The left forearm forms a right angle with the left upper arm. The muscles in the left upper arm are moderately contracted.
	fully_bent	The left forearm nearly touches or touches the left upper arm, forming a very acute angle close to zero degrees. The left hand is very close to or touching the left shoulder; the muscles in the left upper arm are fully contracted.
RightElbow	neutral	The right elbow is extended naturally, forming a straight line from the right shoulder to the right wrist. The right upper arm and right forearm create a nearly straight alignment of about 180 degrees.
	slightly_bent_in	The right forearm forms a slightly obtuse angle with the right upper arm. The right hand moves slightly closer to the right elbow; the muscles in the right upper arm contract slightly.
	bent_in_90_degrees	The right forearm forms a right angle with the right upper arm. The muscles in the right upper arm are moderately contracted.
	fully_bent	The right forearm nearly touches or touches the right upper arm, forming a very acute angle close to zero degrees. The right hand is very close to or touching the right shoulder; the muscles in the right upper arm are fully contracted.
LeftWrist	neutral	The left wrist extends straight, aligned with the left forearm, forming a continuous straight line from the left elbow to the left hand. The angle between the left forearm and left wrist is close to 180 degrees.
	bent_upward	The left wrist forms a right upward angle with the left forearm. The muscles on the back of the left forearm are fully contracted.
	bent_slightly_downward	The left wrist forms a small downward angle, less than 45 degrees. The left palm moves slightly closer to the left forearm; the muscles on the front of the left forearm contract slightly.
	fully_bent_downward	The left wrist forms a right downward angle with the left forearm. The muscles on the front of the left forearm are fully contracted.
	tilted_towards_thumb_side	The left wrist tilts laterally to form a small angle, less than 20 degrees, moving the thumb side of the left hand closer to the left forearm. The muscles on the thumb side of the left forearm contract slightly.
	tilted_towards_pinky_side	The left wrist tilts laterally to form a small angle, less than 20 degrees, moving the little finger side of the left hand closer to the left forearm. The muscles on the pinky side of the left forearm contract slightly.
RightWrist	neutral	The right wrist extends straight, aligned with the right forearm, forming a continuous straight line from the right elbow to the right hand. The angle between the right forearm and right wrist is close to 180 degrees.
	bent_upward	The right wrist forms a right upward angle with the right forearm. The muscles on the back of the right forearm are fully contracted.
	bent_slightly_downward	The right wrist forms a small downward angle, less than 45 degrees. The right palm moves slightly closer to the right forearm; the muscles on the front of the right forearm contract slightly.
	fully_bent_downward	The right wrist forms a right downward angle with the right forearm. The muscles on the front of the right forearm are fully contracted.

Body Part	Position	Description
	tilted_towards_thumb_side	The right wrist tilts laterally to form a small angle, less than 20 degrees, moving the thumb side of the right hand closer to the right forearm. The muscles on the thumb side of the right forearm contract slightly.
	tilted_towards_pinky_side	The right wrist tilts laterally to form a small angle, less than 20 degrees, moving the little finger side of the right hand closer to the right forearm. The muscles on the pinky side of the right forearm contract slightly.
LeftUpperLeg	neutral	The left upperleg is aligned with the body midline, standing straight with the foot pointing forward relative to the torso.
	forward	The left upperleg is extended forward in front of the body, perpendicular to the torso.
	side	The left upperleg is extended out to the side, perpendicular to the torso.
	forward_to_side	The left upperleg is extended fully forward and slightly to the side, forming an approximate 45-degree angle between the forward direction and the side.
	forward_to_midline	The left upperleg is extended fully forward and slightly towards the midline of the body.
	neutral_to_forward	The left upperleg is lifted from the neutral position and extended forward in front of the body, in the middle of neutral and forward positions.
	neutral_to_backward	The left upperleg is lifted from the neutral position and extended backward behind the body, in the middle of neutral and backward positions.
	forward_to_upward	The left upperleg is slightly raised upwards from the forward position.
	neutral_to_side	The left upperleg is lifted from the neutral position and extended to the side, in the middle of neutral and side positions.
	neutral_to_forward_side	The left upperleg is lifted from the neutral position and extended partly forward and slightly to the side, forming a diagonal line.
	neutral_to_forward_midline	The left upperleg is lifted from the neutral position and extended partly forward and slightly towards the midline of the body.
	neutral_to_backward_side	The left upperleg is lifted from the neutral position and extended partly backward and slightly to the side, forming a diagonal line.
	neutral_to_backward_midline	The left upperleg is lifted from the neutral position and extended partly backward and slightly towards the midline of the body.
	forward_to_upward_side	The left upperleg is extended fully forward and raised partly upwards while slightly moving towards the side.
	forward_to_upward_midline	The left upperleg is extended fully forward and raised partly upwards while slightly moving towards the midline of the body.
RightUpperLeg	neutral	The right upperleg is aligned with the body midline, standing straight with the foot pointing forward relative to the torso.
	forward	The right upperleg is extended forward in front of the body, perpendicular to the torso.
	side	The right upperleg is extended out to the side, perpendicular to the torso.
	forward_to_side	The right upperleg is extended fully forward and slightly to the side, forming an approximate 45-degree angle between the forward direction and the side.
	forward_to_midline	The right upperleg is extended fully forward and slightly towards the midline of the body.
	neutral_to_forward	The right upperleg is lifted from the neutral position and extended forward in front of the body, in the middle of neutral and forward positions.
	neutral_to_backward	The right upperleg is lifted from the neutral position and extended backward behind the body, in the middle of neutral and backward positions.
	forward_to_upward	The right upperleg is slightly raised upwards from the forward position.
	neutral_to_side	The right upperleg is lifted from the neutral position and extended to the side, in the middle of neutral and side positions.
	neutral_to_forward_side	The right upperleg is lifted from the neutral position and extended partly forward and slightly to the side, forming a diagonal line.
	neutral_to_forward_midline	The right upperleg is lifted from the neutral position and extended partly forward and slightly towards the midline of the body.
	neutral_to_backward_side	The right upperleg is lifted from the neutral position and extended partly backward and slightly to the side, forming a diagonal line.
	neutral_to_backward_midline	The right upperleg is lifted from the neutral position and extended partly backward and slightly towards the midline of the body.
	forward_to_upward_side	The right upperleg is extended fully forward and raised partly upwards while slightly moving towards the side.
	forward_to_upward_midline	The right upperleg is extended fully forward and raised partly upwards while slightly moving towards the midline of the body.
LeftKnee	neutral	The left leg is straight, the left knee fully extended, forming a continuous line from the left thigh to the left ankle. The angle between the left thigh and calf is close to 180 degrees. Both the front and back thigh muscles maintain a neutral length.
	slightly_bent	The left calf forms a small angle, less than 45 degrees, with the back of the left thigh. The left knee is slightly flexed; the front thigh muscles contract slightly, and the back thigh muscles lengthen slightly.
	bent_at_90_degrees	The left calf forms a right angle with the left back thigh. The front thigh muscles are moderately contracted, and the back thigh muscles are moderately stretched.
	fully_bent	The left calf forms a nearly zero-degree angle with the back of the left thigh, with the left heel nearly or fully touching the buttocks. The front thigh muscles are extremely contracted, and the back thigh muscles are maximally stretched.
RightKnee	neutral	The right leg is straight, the right knee fully extended, forming a continuous line from the right thigh to the right ankle. The angle between the right thigh and calf is close to 180 degrees. Both the front and back thigh muscles maintain a neutral length.
	slightly_bent	The right calf forms a small angle, less than 45 degrees, with the back of the right thigh. The right knee is slightly flexed; the front thigh muscles contract slightly, and the back thigh muscles lengthen slightly.
	bent_at_90_degrees	The right calf forms a right angle with the right back thigh. The front thigh muscles are moderately contracted, and the back thigh muscles are moderately stretched.
	fully_bent	The right calf forms a nearly zero-degree angle with the back of the right thigh, with the right heel nearly or fully touching the buttocks. The front thigh muscles are extremely contracted, and the back thigh muscles are maximally stretched.
LeftAnkle	neutral	The left toes point straight ahead, aligned with the left foot. The angle between the top of the left foot and the left shin is around 90 degrees. Both the front and back muscles of the left lower leg maintain a neutral length.
	bent_upward	The top of the left foot forms an acute angle, less than 90 degrees, with the left shin. The left toes point closer to the left shin; the muscles on the front of the left lower leg contract to lift the left toes.
	bent_downward	The top of the left foot forms a steep obtuse angle, around 180 degrees, with the left shin. The left toes point significantly downward; the calf muscles are fully contracted.
	tilted_inward	The sole of the left foot moves towards the midline of the body, forming a small inward angle with the left ankle. The muscles on the inside of the left lower leg contract slightly.

Body Part	Position	Description
	tilted_outward	The sole of the left foot moves away from the midline of the body, forming a small outward angle with the left ankle. The muscles on the outside of the left lower leg contract slightly.
RightAnkle	neutral	The right toes point straight ahead, aligned with the right foot. The angle between the top of the right foot and the right shin is around 90 degrees. Both the front and back muscles of the right lower leg maintain a neutral length.
	bent_upward	The top of the right foot forms an acute angle, less than 90 degrees, with the right shin. The right toes point closer to the right shin; the muscles on the front of the right lower leg contract to lift the right toes.
	bent_downward	The top of the right foot forms a steep obtuse angle, around 180 degrees, with the right shin. The right toes point significantly downward; the calf muscles are fully contracted.
	tilted_inward	The sole of the right foot moves towards the midline of the body, forming a small inward angle with the right ankle. The muscles on the inside of the right lower leg contract slightly.
	tilted_outward	The sole of the right foot moves away from the midline of the body, forming a small outward angle with the right ankle. The muscles on the outside of the right lower leg contract slightly.
LeftToes	neutral	The left toes point straight ahead, aligned with the left foot. The left toes and the left foot form a straight line from the base to the tips, making a right angle.
	curled_up	The left toes lift upwards, forming small upward angles between the left toe tips and the left foot. The muscles on the top of the left foot and left toes contract to lift the left toes.
	curled_down	The left toes curl downward, forming small downward angles with the sole of the left foot. The muscles on the bottom of the left foot and left toes contract to curl the left toes.
RightToes	neutral	The right toes point straight ahead, aligned with the right foot. The right toes and the right foot form a straight line from the base to the tips, making a right angle.
	curled_up	The right toes lift upwards, forming small upward angles between the right toe tips and the right foot. The muscles on the top of the right foot and right toes contract to lift the right toes.
	curled_down	The right toes curl downward, forming small downward angles with the sole of the right foot. The muscles on the bottom of the right foot and right toes contract to curl the right toes.

A.2 Details of Preset Body Parts and Predefined Positions

Table A2: Preset body parts with their predefined positions and the corresponding coordinates of relevant SMPL joints (Rules).

Body Part	Position	Coordinates of the Relevant SMPL Joints (Rules)
Head	neutral	{m_avg_Head: [0.0, 0.0, 0.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	tilted_up_slightly	{m_avg_Head: [-30.0, 0.0, 0.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	tilted_up_fully	{m_avg_Head: [-45.0, 0.0, 0.0], m_avg_Neck: [-10.0, 0.0, 0.0]}
	tilted_down_slightly	{m_avg_Head: [30.0, 0.0, 0.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	tilted_down_fully	{m_avg_Head: [45.0, 0.0, 0.0], m_avg_Neck: [30.0, 0.0, 0.0]}
	turned_left_slightly	{m_avg_Head: [0.0, -45.0, 0.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	turned_left_fully	{m_avg_Head: [0.0, -60.0, 0.0], m_avg_Neck: [0.0, -30.0, 0.0]}
	turned_right_slightly	{m_avg_Head: [0.0, 45.0, 0.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	turned_right_fully	{m_avg_Head: [0.0, 60.0, 0.0], m_avg_Neck: [0.0, 30.0, 0.0]}
	tilted_left_slightly	{m_avg_Head: [0.0, 0.0, 20.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	tilted_left_fully	{m_avg_Head: [0.0, 0.0, 45.0], m_avg_Neck: [0.0, 0.0, 20.0]}
	tilted_right_slightly	{m_avg_Head: [0.0, 0.0, -20.0], m_avg_Neck: [0.0, 0.0, 0.0]}
	tilted_right_fully	{m_avg_Head: [0.0, 0.0, -45.0], m_avg_Neck: [0.0, 0.0, -20.0]}
Torso	neutral	{m_avg_Spine1: [0.0, 0.0, 0.0], m_avg_Spine2: [0.0, 0.0, 0.0], m_avg_Spine3: [0.0, 0.0, 0.0]}
	twisted_left_slightly	{m_avg_Spine1: [0.0, -20.0, 0.0], m_avg_Spine2: [0.0, -10.0, 0.0], m_avg_Spine3: [0.0, -5.0, 0.0]}
	twisted_left_fully	{m_avg_Spine1: [0.0, -30.0, 0.0], m_avg_Spine2: [0.0, -20.0, 0.0], m_avg_Spine3: [0.0, -10.0, 0.0]}
	twisted_right_slightly	{m_avg_Spine1: [0.0, 20.0, 0.0], m_avg_Spine2: [0.0, 10.0, 0.0], m_avg_Spine3: [0.0, 5.0, 0.0]}
	twisted_right_fully	{m_avg_Spine1: [0.0, 30.0, 0.0], m_avg_Spine2: [0.0, 20.0, 0.0], m_avg_Spine3: [0.0, 10.0, 0.0]}
	bent_forward_slightly	{m_avg_Spine1: [30.0, 0.0, 0.0], m_avg_Spine2: [5.0, 0.0, 0.0], m_avg_Spine3: [5.0, 0.0, 0.0]}
	bent_forward_fully	{m_avg_Spine1: [60.0, 0.0, 0.0], m_avg_Spine2: [10.0, 0.0, 0.0], m_avg_Spine3: [10.0, 0.0, 0.0]}
	bent_backward	{m_avg_Spine1: [-20.0, 0.0, 0.0], m_avg_Spine2: [0.0, 0.0, 0.0], m_avg_Spine3: [0.0, 0.0, 0.0]}
	tilted_left_slightly	{m_avg_Spine1: [0.0, 0.0, 10.0], m_avg_Spine2: [0.0, 0.0, 5.0], m_avg_Spine3: [0.0, 0.0, 5.0]}
	tilted_left_fully	{m_avg_Spine1: [0.0, 0.0, 20.0], m_avg_Spine2: [0.0, 0.0, 10.0], m_avg_Spine3: [0.0, 0.0, 10.0]}
	tilted_right_slightly	{m_avg_Spine1: [0.0, 0.0, -10.0], m_avg_Spine2: [0.0, 0.0, -5.0], m_avg_Spine3: [0.0, 0.0, -5.0]}
	tilted_right_fully	{m_avg_Spine1: [0.0, 0.0, -20.0], m_avg_Spine2: [0.0, 0.0, -10.0], m_avg_Spine3: [0.0, 0.0, -10.0]}
LeftUpperArm	neutral	{m_avg_L_Collar: [0.0, 0.0, 20.0], m_avg_L_Shoulder: [0.0, 0.0, 70.0]}
	forward_elbowpit_inward	{m_avg_L_Collar: [0.0, 0.0, 0.0], m_avg_L_Shoulder: [0.0, 90.0, 0.0]}
	forward_elbowpit_upward	{m_avg_L_Collar: [0.0, 0.0, 20.0], m_avg_L_Shoulder: [-70.0, 90.0, 0.0]}
	upward	{m_avg_L_Collar: [0.0, 0.0, -45.0], m_avg_L_Shoulder: [0.0, 0.0, -40.0]}
	side_elbowpit_forward	{m_avg_L_Collar: [0.0, 0.0, 0.0], m_avg_L_Shoulder: [0.0, 0.0, 0.0]}
	side_elbowpit_upward	{m_avg_L_Collar: [-45.0, 0.0, 0.0], m_avg_L_Shoulder: [-45.0, 0.0, 0.0]}
	neutral_to_forward	{m_avg_L_Collar: [0.0, 0.0, 10.0], m_avg_L_Shoulder: [-60.0, 20.0, 70.0]}
	forward_to_upward	{m_avg_L_Collar: [0.0, 0.0, -10.0], m_avg_L_Shoulder: [0.0, 90.0, -35.0]}
	forward_to_back	{m_avg_L_Collar: [0.0, 0.0, 10.0], m_avg_L_Shoulder: [10.0, 90.0, 120.0]}
	forward_to_midline	{m_avg_L_Collar: [0.0, 10.0, 0.0], m_avg_L_Shoulder: [0.0, 120.0, 0.0]}
	forward_to_side	{m_avg_L_Collar: [0.0, 10.0, 0.0], m_avg_L_Shoulder: [0.0, 30.0, 0.0]}
	side_to_back	{m_avg_L_Collar: [0.0, -10.0, 0.0], m_avg_L_Shoulder: [0.0, -30.0, 0.0]}
	neutral_to_side	{m_avg_L_Collar: [0.0, 0.0, 10.0], m_avg_L_Shoulder: [0.0, 0.0, 45.0]}
	upward_to_side	{m_avg_L_Collar: [0.0, 0.0, -10.0], m_avg_L_Shoulder: [0.0, 0.0, -45.0]}
	forward_to_upward_side	{m_avg_L_Collar: [0.0, 0.0, -10.0], m_avg_L_Shoulder: [0.0, 45.0, -30.0]}
	forward_to_upward_midline	{m_avg_L_Collar: [0.0, 15.0, -10.0], m_avg_L_Shoulder: [0.0, 110.0, -30.0]}
	side_to_upward_back	{m_avg_L_Collar: [0.0, -15.0, -10.0], m_avg_L_Shoulder: [0.0, -30.0, -30.0]}
	neutral_to_forward_midline	{m_avg_L_Collar: [0.0, 0.0, 10.0], m_avg_L_Shoulder: [-45.0, 0.0, 100.0]}
	neutral_to_forward_side	{m_avg_L_Collar: [0.0, 0.0, 10.0], m_avg_L_Shoulder: [45.0, 90.0, 45.0]}
	neutral_to_backward_side	{m_avg_L_Collar: [20.0, 0.0, 10.0], m_avg_L_Shoulder: [45.0, 90.0, 100.0]}
RightUpperArm	neutral	{m_avg_R_Collar: [0.0, 0.0, -20.0], m_avg_R_Shoulder: [0.0, 0.0, -70.0]}
	forward_elbowpit_inward	{m_avg_R_Collar: [0.0, 0.0, 0.0], m_avg_R_Shoulder: [0.0, -90.0, 0.0]}
	forward_elbowpit_upward	{m_avg_R_Collar: [0.0, 0.0, -20.0], m_avg_R_Shoulder: [-70.0, -90.0, 0.0]}
	upward	{m_avg_R_Collar: [0.0, 0.0, 45.0], m_avg_R_Shoulder: [0.0, 0.0, 40.0]}
	side_elbowpit_forward	{m_avg_R_Collar: [0.0, 0.0, 0.0], m_avg_R_Shoulder: [0.0, 0.0, 0.0]}
	side_elbowpit_upward	{m_avg_R_Collar: [-45.0, 0.0, 0.0], m_avg_R_Shoulder: [-45.0, 0.0, 0.0]}
	neutral_to_forward	{m_avg_R_Collar: [0.0, 0.0, -10.0], m_avg_R_Shoulder: [-60.0, -20.0, -70.0]}
	forward_to_upward	{m_avg_R_Collar: [0.0, 0.0, 10.0], m_avg_R_Shoulder: [0.0, -90.0, 35.0]}
	neutral_to_back	{m_avg_R_Collar: [0.0, 0.0, -10.0], m_avg_R_Shoulder: [10.0, -90.0, -120.0]}
	forward_to_midline	{m_avg_R_Collar: [0.0, -10.0, 0.0], m_avg_R_Shoulder: [0.0, -120.0, 0.0]}
	forward_to_side	{m_avg_R_Collar: [0.0, -10.0, 0.0], m_avg_R_Shoulder: [0.0, -30.0, 0.0]}
	side_to_back	{m_avg_R_Collar: [0.0, 10.0, 0.0], m_avg_R_Shoulder: [0.0, 30.0, 0.0]}
	neutral_to_side	{m_avg_R_Collar: [0.0, 0.0, -10.0], m_avg_R_Shoulder: [0.0, 0.0, -45.0]}
	upward_to_side	{m_avg_R_Collar: [0.0, 0.0, 10.0], m_avg_R_Shoulder: [0.0, 0.0, 45.0]}
	forward_to_upward_side	{m_avg_R_Collar: [0.0, 0.0, 10.0], m_avg_R_Shoulder: [0.0, -45.0, 30.0]}
	forward_to_upward_midline	{m_avg_R_Collar: [0.0, -15.0, 10.0], m_avg_R_Shoulder: [0.0, -110.0, 30.0]}
	side_to_upward_back	{m_avg_R_Collar: [0.0, 15.0, 10.0], m_avg_R_Shoulder: [0.0, 30.0, 30.0]}
	neutral_to_forward_midline	{m_avg_R_Collar: [0.0, 0.0, -10.0], m_avg_R_Shoulder: [-45.0, 0.0, -100.0]}
	neutral_to_forward_side	{m_avg_R_Collar: [0.0, 0.0, -10.0], m_avg_R_Shoulder: [45.0, -90.0, -45.0]}
	neutral_to_backward_side	{m_avg_R_Collar: [20.0, 0.0, -10.0], m_avg_R_Shoulder: [45.0, -90.0, -100.0]}
LeftElbow	neutral	{m_avg_L_Elbow: [0.0, 0.0, 0.0]}
	slightly_bent_in	{m_avg_L_Elbow: [0.0, 45.0, 0.0]}
	bent_in_90_degrees	{m_avg_L_Elbow: [0.0, 90.0, 0.0]}
RightElbow	fully_bent	{m_avg_L_Elbow: [0.0, 135.0, 0.0]}
	neutral	{m_avg_R_Elbow: [0.0, 0.0, 0.0]}
	slightly_bent_in	{m_avg_R_Elbow: [0.0, -45.0, 0.0]}
	bent_in_90_degrees	{m_avg_R_Elbow: [0.0, -90.0, 0.0]}
	fully_bent	{m_avg_R_Elbow: [0.0, -135.0, 0.0]}

Body Part	Position	Coordinates of the Relevant SMPL Joints (Rules)
LeftWrist	neutral	{m_avg_L_Wrist: [0.0, 0.0, 0.0]}
	bent_upward	{m_avg_L_Wrist: [0.0, 0.0, -45.0]}
	bent_slightly_downward	{m_avg_L_Wrist: [0.0, 0.0, 45.0]}
	fully_bent_downward	{m_avg_L_Wrist: [0.0, 0.0, 80.0]}
	tilted_towards_thumb_side	{m_avg_L_Wrist: [0.0, 30.0, 0.0]}
RightWrist	neutral	{m_avg_R_Wrist: [0.0, 0.0, 0.0]}
	bent_upward	{m_avg_R_Wrist: [0.0, 0.0, -45.0]}
	bent_slightly_downward	{m_avg_R_Wrist: [0.0, 0.0, 45.0]}
	fully_bent_downward	{m_avg_R_Wrist: [0.0, 0.0, -80.0]}
	tilted_towards_thumb_side	{m_avg_R_Wrist: [0.0, -30.0, 0.0]}
LeftUpperLeg	neutral	{m_avg_L_Hip: [0.0, 0.0, 0.0]}
	forward	{m_avg_L_Hip: [-90.0, 0.0, 0.0]}
	side	{m_avg_L_Hip: [0.0, 0.0, -80.0]}
	forward_to_side	{m_avg_L_Hip: [-90.0, 0.0, -45.0]}
	forward_to_midline	{m_avg_L_Hip: [-90.0, 0.0, 45.0]}
	neutral_to_forward	{m_avg_L_Hip: [-45.0, 0.0, 0.0]}
	neutral_to_backward	{m_avg_L_Hip: [45.0, 0.0, 0.0]}
	forward_to_upward	{m_avg_L_Hip: [-120.0, 0.0, 0.0]}
	neutral_to_side	{m_avg_L_Hip: [0.0, 0.0, -45.0]}
	neutral_to_forward_side	{m_avg_L_Hip: [-45.0, 0.0, -45.0]}
	neutral_to_forward_midline	{m_avg_L_Hip: [-45.0, 0.0, 45.0]}
	neutral_to_backward_side	{m_avg_L_Hip: [45.0, 0.0, -45.0]}
	neutral_to_backward_midline	{m_avg_L_Hip: [45.0, 0.0, 45.0]}
	forward_to_upward_side	{m_avg_L_Hip: [-120.0, 0.0, -45.0]}
	forward_to_upward_midline	{m_avg_L_Hip: [-120.0, 0.0, 45.0]}
RightUpperLeg	neutral	{m_avg_R_Hip: [0.0, 0.0, 0.0]}
	forward	{m_avg_R_Hip: [-90.0, 0.0, 0.0]}
	side	{m_avg_R_Hip: [0.0, 0.0, 80.0]}
	forward_to_side	{m_avg_R_Hip: [-90.0, 0.0, 45.0]}
	forward_to_midline	{m_avg_R_Hip: [-90.0, 0.0, -45.0]}
	neutral_to_forward	{m_avg_R_Hip: [-45.0, 0.0, 0.0]}
	neutral_to_backward	{m_avg_R_Hip: [45.0, 0.0, 0.0]}
	forward_to_upward	{m_avg_R_Hip: [-120.0, 0.0, 0.0]}
	neutral_to_side	{m_avg_R_Hip: [0.0, 0.0, 45.0]}
	neutral_to_forward_side	{m_avg_R_Hip: [-45.0, 0.0, 45.0]}
	neutral_to_forward_midline	{m_avg_R_Hip: [-45.0, 0.0, -45.0]}
	neutral_to_backward_side	{m_avg_R_Hip: [45.0, 0.0, 45.0]}
	neutral_to_backward_midline	{m_avg_R_Hip: [45.0, 0.0, -45.0]}
	forward_to_upward_side	{m_avg_R_Hip: [-120.0, 0.0, 45.0]}
	forward_to_upward_midline	{m_avg_R_Hip: [-120.0, 0.0, -45.0]}
LeftKnee	neutral	{m_avg_L_Knee: [0.0, 0.0, 0.0]}
	slightly_bent	{m_avg_L_Knee: [45.0, 0.0, 0.0]}
	bent_at_90_degrees	{m_avg_L_Knee: [90.0, 0.0, 0.0]}
	fully_bent	{m_avg_L_Knee: [135.0, 0.0, 0.0]}
RightKnee	neutral	{m_avg_R_Knee: [0.0, 0.0, 0.0]}
	slightly_bent	{m_avg_R_Knee: [45.0, 0.0, 0.0]}
	bent_at_90_degrees	{m_avg_R_Knee: [90.0, 0.0, 0.0]}
	fully_bent	{m_avg_R_Knee: [135.0, 0.0, 0.0]}
LeftAnkle	neutral	{m_avg_L_Ankle: [0.0, 0.0, 0.0]}
	bent_upward	{m_avg_L_Ankle: [-20.0, 0.0, 0.0]}
	bent_downward	{m_avg_L_Ankle: [45.0, 0.0, 0.0]}
	tilted_inward	{m_avg_L_Ankle: [0.0, 0.0, 30.0]}
	tilted_outward	{m_avg_L_Ankle: [0.0, 0.0, -10.0]}
RightAnkle	neutral	{m_avg_R_Ankle: [0.0, 0.0, 0.0]}
	bent_upward	{m_avg_R_Ankle: [-20.0, 0.0, 0.0]}
	bent_downward	{m_avg_R_Ankle: [45.0, 0.0, 0.0]}
	tilted_inward	{m_avg_R_Ankle: [0.0, 0.0, -30.0]}
	tilted_outward	{m_avg_R_Ankle: [0.0, 0.0, 10.0]}
LeftToes	neutral	{m_avg_L_Foot: [0.0, 0.0, 0.0]}
	curled_up	{m_avg_L_Foot: [-30.0, 0.0, 0.0]}
	curled_down	{m_avg_L_Foot: [30.0, 0.0, 0.0]}
RightToes	neutral	{m_avg_R_Foot: [0.0, 0.0, 0.0]}
	curled_up	{m_avg_R_Foot: [-30.0, 0.0, 0.0]}
	curled_down	{m_avg_R_Foot: [30.0, 0.0, 0.0]}

A.3 Details of Human Evaluation

We conduct human evaluation as part of our research methodology. The nine human evaluators are graduate students, technical staff or researchers working on artificial intelligence at the same university, and participate voluntarily with above-average wage compensation. Tasks are designed to be safe and unbiased, with clear instructions and reasonable time commitments. Participants are informed of the study’s purpose and withdrawal rights. Prior to evaluation, we explicitly explain how evaluators’ responses and feedback will be used in our research, including potential publication in academic venues, and obtain written consent from all evaluators. No personal data are collected. The protocol is approved by our institution’s ethics review board and adheres to human subject research guidelines.

A.3.1 High-level Planning

Score	Judgement
5	The high-level plan follows the motion instruction well and specifies all important details.
4	The high-level plan generally follows the motion instruction (80–90%), but contains some minor errors.
3	The high-level plan follows the motion instruction 50–70% and contains one or two major errors that prevent it from achieving the goal
2	The high-level plan shows some sign of following the motion instruction (20–40%), but contains so many errors that it is far from the goal state
1	The high-level plan does not follow the motion instruction at all

Table A3: Rubric for High-level Plan Score (HPS). The instruction to the annotators: Given the motion instruction, to what extent do you think the high-level plan of body part movements appropriately specifies the instructed motion? Score from 1 (poor) to 5 (excellent). Possible shortcomings usually include wrong and incomplete action descriptions.

Illustration of Evaluation Rubrics (High-level Planning)

Thank you for taking part in the evaluation! In this document, I will illustrate the rubrics of scoring the high-level plans and show some examples.

Introduction

In this study, we aim to explore the extent to which Large Language Models (LLMs) understand the knowledge involved in human motions. Your task is to evaluate the LLM-generated high-level motion plans in the middle of the animation generation, following our rubrics. Each high-level plan is composed of several steps, with each step specifying the movements, initial states and final states of moved body parts. The number of high-level plans is 108.

Figure A3: The illustrative document for each human evaluator to read. It is followed by evaluation rubrics and four examples covering different scores.

A.3.2 Complete Generation

Score	Judgement
5	The animation follows the motion instruction well without redundant or strange movements
4	The animation generally follows the motion instruction (70–90%), but contains some minor errors (e.g., redundant or strange movements)
3	The animation follows the motion instruction 40–60% and contains one or two major errors that prevent it from achieving the goal
2	The animation shows some sign of following the motion instruction (20–30%), but contains so many errors that it is far from the goal state
1	The animation does not follow the motion instruction at all

Table A4: Rubric for Whole Body Score (WBS). The instruction to the annotators: Given the motion instruction, to what extent do you think the animation is appropriately following the instructed motion? Score from 1 (poor) to 5 (excellent).

Label	Judgement
Good	The body part follows the given motion instruction well
Partially Good	The body part follows the given motion instruction partially, but has errors
Bad	The body part does not follow the given motion instruction at all. Or, the body part is not absolutely necessary to this motion but is ridiculously moved
Not Relevant	The body part is not absolutely necessary to this motion and is not ridiculously moved

Table A5: Rubric for Body Part Quality (BPQ). The instruction to the annotators: For each body part, choose one from “Good”, “Partially Good”, “Bad” and “Not Relevant”.

Illustration of Evaluation Rubrics (Complete Generation)

Thank you for taking part in the evaluation! In this document, I will illustrate the rubrics of scoring the animations and show some examples.

Introduction

In this study, we aim to explore the extent to which Large Language Models (LLMs) understand the knowledge involved in human motions. In particular, we seek to exploit LLMs' knowledge of human motions to generate animations on a 3D human model, given the motion instruction. Your task is to evaluate the generated animations following our rubrics.

Here are two example animations:

Motion Instruction: water a 30-centimeter-tall plant using the watering can in the right hand

■ example_0.mp4

Motion Instruction: slide the window open from the center to the sides with both hands

■ example_1.mp4



The 3D human model in Unity

Important Preliminaries (consult with me if you are unclear with these)

All your judgements should only be based on **how well the human model changes the body part positions**. Please **DO NOT** give negative judgement because of the following phenomena as they are not relevant to this evaluation:

1. The generated animation is not really realistic, and looks like a robot.
 - Explanation: We linearly interpolate positions of body parts across steps, and the time range of each step is a very rough estimate by LLMs.
2. The human model stays in the same position like a dangling puppet.
 - Explanation: We do not model root movements so the human model does not globally move. For example, when the human model walks, it simply moves body parts locally without globally moving to other positions.
3. The body parts might overlap with each other.
 - Explanation: We do not set collision detection, so body parts might sink into each other.
4. The human model's fingers are not moving at all.
 - Explanation: The human model I'm using does not involve finger joints.

Figure A4: The illustrative document for each human evaluator to read. This page of the document is followed by evaluation rubrics and 6 examples covering different WBS and BPQ.

Animation ID: sample

Motion Instruction: slide the window open from the center to the sides with both hands

Overall Score (1 - poor, 5 - excellent) *

1

2

3

4

5

☐

☐

☐

☐

☐

Head *

☐ Good

☐ Partially Good

☐ Bad

☐ Not Relevant

Torso *

☐ Good

☐ Partially Good

Figure A5: Sample form in human evaluation of the animations from the complete generation. At the header of each form, we provide a link to the illustrative document shown in Figure A4. We ensure that each annotator evaluates a balanced mix of animations from different settings including LLMs and querying strategies.

A.4 Motion Instructions

We create twenty motion instructions for the main experiments. They are freely available for research purposes only. Researchers working on language and human motion are welcome to access and use these instructions for future investigation, which aligns with our intended use case. Derivative works or applications should remain within research contexts. The motion instructions are carefully curated to be neutral in nature, containing no personally identifiable information or controversial content.

ID	Motion Instruction
1	Slide the window open from the center to the sides with both hands.
2	Water a 30-centimeter-tall plant using the watering can in the right hand.
3	Look down to check the time of the watch on the left wrist.
4	Pat a 30-centimeter-tall dog in front of you on the head with the right hand.
5	Lean back fully and toss the ball into the air at a 45-degree angle using both hands.
6	Wipe down the 1-meter-high table in front of you with a cloth in the left hand.
7	Hold the glass with the left hand and pour the juice with the right hand.
8	Put a book on the 2-meter-high shelf with both hands.
9	Lift a 20-centimeter-high box from the ground to the table on your left with both hands.
10	Swing the golf club from right to left.
11	Close the 2-meter-high store shutter door from top to bottom.
12	Squat to pick up litter by the right foot with the right hand.
13	Lift the right shoe with both hands and put it on in the air.
14	Perform a left-leg high side kick in Karate.
15	Kneel in a traditional Japanese bow.
16	Roll out a yoga mat on the ground.
17	Crouch to check a car tyre.
18	Arch the back 60 degrees to relieve tension in the lower back muscles with two hands on the waist.
19	Bend to the left to reach for an item by the left foot without moving or bending the left leg.
20	Walk through while ducking under a low-hanging branch.

Table A6: The motion instructions for main experiments. Each instruction specifies necessary contextual elements to eliminate ambiguity while testing LLMs’ ability to infer implicit motion details. To show the potential of application, we devise each instruction to be related to a practical scene, while deliberately avoiding common game animations to focus on challenging scenarios requiring genuine motion understanding.

A.5 Hyperparameters and Computational Costs

Hyperparameter	Value
temperature	1
max_tokens	4095
timeout	60
max_retries	3

Table A7: Hyperparameter configuration for LLMs

The average cost per animation generation ranges from \$2.70 (Claude 3.5 Sonnet) to \$0.07 (GPT-4o-mini), with GPT-4o at \$1.20 and GPT-3.5-turbo at \$0.25. The open-source Llama-3.1-70B requires 1.5-2 hours on one 48GB NVIDIA RTX A6000.

A.6 Self-reflection Analysis

We report the reflection statistics in Figure A6. To investigate the effect of self-reflection, we calculate the percentage of corrections among all body parts of all steps (**Correction Percentage**), the percentage of body parts with finally correct positions among all corrected body parts (**Success Rate**), and the percentage of corrected body parts where the last selected position is correct and previous selected positions are wrong, among all corrected body parts (**Perfect Correction Rate**).

Llama-3.1-70B has an extraordinarily high correction percentage, while other models seldom correct after reflection. While Llama-3.1-70B's Success Rate is the highest, its Perfect Correction Rate is the lowest relative to Success Rate. This phenomenon might be attributed to Llama-3.1-70B's lack of proper instruction following capabilities, i.e., it tends to reflect and correct when asked, no matter whether the selected position should be corrected.

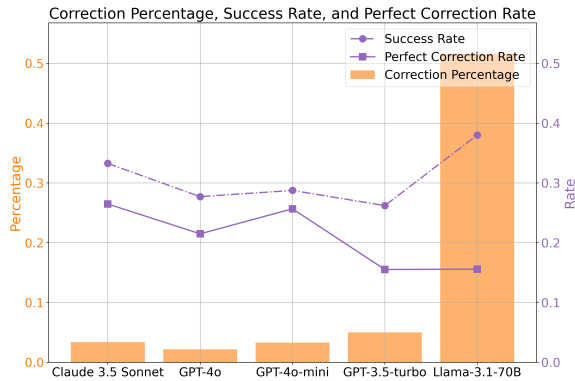


Figure A6: Reflection Statistics.

A.7 Complementary Agreement Information

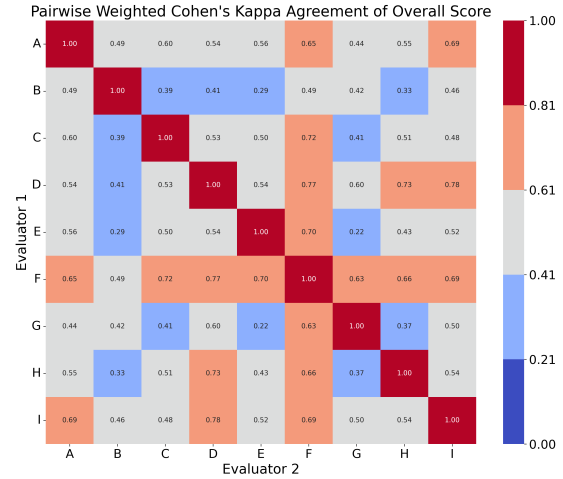


Figure A7: Pairwise weighted kappa scores of nine evaluators for WBS agreement. Based on the interpretation of Landis and Koch (1977), the ranges 0.21-0.40, 0.41-0.60, and 0.61-0.80 respectively correspond to fair, moderate, and substantial levels of agreement.

Body Part	APA
Head	0.510 / 0.758
Torso	0.536 / 0.667
Left Arm	0.569 / 0.558
Right Arm	0.553 / 0.550
Left Leg	0.637 / 0.792
Right Leg	0.638 / 0.717

Table A8: Average pairwise agreement (APA) on BPQ for each body part. Each score pair contains the agreement among human annotators (left), and between human and Gemini 2.5 Pro judgement (right). Human agreement is computed as the mean percentage of matching categories across annotator pairs; human-Gemini agreement reflects the percentage of matched categories between the majority votes of humans and Gemini 2.5 Pro.

A.8 Statistical Measures

A.8.1 HPS Scores

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B
1	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.67 (0.47, 0.22)	2.33 (0.47, 0.22)	5.00 (0.00, 0.00)
2	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	4.33 (0.47, 0.22)	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)
3	5.00 (0.00, 0.00)	3.33 (0.47, 0.22)	5.00 (0.00, 0.00)	3.33 (0.47, 0.22)	3.00 (0.00, 0.00)
4	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	2.33 (0.47, 0.22)	2.33 (0.47, 0.22)
5	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.33 (0.47, 0.22)	4.67 (0.47, 0.22)
6	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.33 (0.47, 0.22)
7	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)
8	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	2.67 (0.47, 0.22)	4.67 (0.47, 0.22)
9	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	2.33 (0.47, 0.22)
10	3.67 (0.47, 0.22)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	3.33 (0.47, 0.22)	3.67 (0.47, 0.22)
11	4.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)
12	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)
13	3.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	1.67 (0.47, 0.22)	3.33 (0.47, 0.22)
14	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.67 (0.47, 0.22)
15	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	2.67 (0.47, 0.22)
16	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	3.67 (0.47, 0.22)	4.00 (0.00, 0.00)	4.67 (0.47, 0.22)
17	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.33 (0.47, 0.22)
18	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.33 (0.47, 0.22)	4.33 (0.47, 0.22)
19	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	2.67 (0.47, 0.22)	4.33 (0.47, 0.22)
20	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	2.67 (0.47, 0.22)	4.67 (0.47, 0.22)

Table A9: Human-scored HPS (piece_by_piece) by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B
1	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	2.67 (0.47, 0.22)	3.00 (0.00, 0.00)	4.67 (0.47, 0.22)
2	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
3	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	3.00 (0.00, 0.00)
4	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	3.67 (0.47, 0.22)
5	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	3.00 (0.00, 0.00)
6	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.00 (0.00, 0.00)
7	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.00 (0.00, 0.00)
8	2.67 (0.47, 0.22)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	2.00 (0.00, 0.00)	3.00 (0.00, 0.00)
9	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	2.00 (0.00, 0.00)
10	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	3.67 (0.47, 0.22)
11	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
12	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)
13	3.67 (0.47, 0.22)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	2.00 (0.00, 0.00)	3.00 (0.00, 0.00)
14	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.00 (0.00, 0.00)
15	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.00 (0.00, 0.00)	3.33 (0.47, 0.22)	3.67 (0.47, 0.22)
16	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.67 (0.47, 0.22)
17	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)
18	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.33 (0.47, 0.22)	4.00 (0.00, 0.00)
19	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	5.00 (0.00, 0.00)
20	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	5.00 (0.00, 0.00)

Table A10: GPT-4.1-scored HPS (piece_by_piece) by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo
1	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)
2	3.67 (0.47, 0.22)	4.67 (0.47, 0.22)	3.33 (0.47, 0.22)	4.33 (0.47, 0.22)
3	4.67 (0.47, 0.22)	3.00 (0.00, 0.00)	3.00 (0.00, 0.00)	3.33 (0.47, 0.22)
4	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)
5	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)
6	3.67 (0.47, 0.22)	5.00 (0.00, 0.00)	3.67 (0.47, 0.22)	3.67 (0.47, 0.22)
7	3.33 (0.47, 0.22)	4.67 (0.47, 0.22)	3.33 (0.47, 0.22)	3.67 (0.47, 0.22)
8	3.67 (0.47, 0.22)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	3.00 (0.00, 0.00)
9	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	2.00 (0.00, 0.00)	3.33 (0.47, 0.22)
10	3.33 (0.47, 0.22)	4.67 (0.47, 0.22)	3.00 (0.00, 0.00)	2.00 (0.82, 0.67)
11	4.33 (0.47, 0.22)	3.33 (0.47, 0.22)	3.67 (0.47, 0.22)	4.67 (0.47, 0.22)
12	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	2.33 (0.47, 0.22)
13	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	4.00 (0.00, 0.00)	2.67 (0.47, 0.22)
14	4.67 (0.47, 0.22)	3.33 (0.47, 0.22)	4.67 (0.47, 0.22)	3.00 (0.00, 0.00)
15	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	2.67 (0.47, 0.22)	2.67 (0.47, 0.22)
16	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	2.67 (0.47, 0.22)	1.00 (0.00, 0.00)
17	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)
18	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	3.33 (0.47, 0.22)
19	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	3.67 (0.47, 0.22)
20	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.67 (0.47, 0.22)	2.00 (0.00, 0.00)

Table A11: Human-scored HPS (in_one_go) by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo
1	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)
2	4.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
3	5.00 (0.00, 0.00)	2.00 (0.00, 0.00)	3.00 (0.00, 0.00)	2.33 (0.47, 0.22)
4	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)
5	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.67 (0.47, 0.22)	3.00 (0.00, 0.00)
6	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	3.33 (0.47, 0.22)	3.00 (0.00, 0.00)
7	4.33 (0.47, 0.22)	4.67 (0.47, 0.22)	3.00 (0.00, 0.00)	3.00 (0.00, 0.00)
8	4.00 (0.00, 0.00)	3.67 (0.47, 0.22)	3.00 (0.00, 0.00)	2.00 (0.00, 0.00)
9	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	2.67 (0.47, 0.22)	3.33 (0.47, 0.22)
10	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	2.00 (0.00, 0.00)
11	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	3.00 (0.00, 0.00)	4.33 (0.47, 0.22)
12	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
13	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	2.00 (0.00, 0.00)	2.67 (0.47, 0.22)
14	5.00 (0.00, 0.00)	3.67 (0.47, 0.22)	4.00 (0.00, 0.00)	3.00 (0.00, 0.00)
15	4.67 (0.47, 0.22)	4.33 (0.47, 0.22)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
16	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.00 (0.00, 0.00)	1.00 (0.00, 0.00)
17	4.33 (0.47, 0.22)	5.00 (0.00, 0.00)	4.33 (0.47, 0.22)	4.00 (0.00, 0.00)
18	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	5.00 (0.00, 0.00)	2.00 (0.00, 0.00)
19	5.00 (0.00, 0.00)	4.67 (0.47, 0.22)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)
20	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	4.00 (0.00, 0.00)	2.00 (0.00, 0.00)

Table A12: GPT-4.1-scored HPS (in_one_go) by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

A.8.2 BPPA Scores

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B
1	0.7812 (0.03, 0.00)	0.7188 (0.00, 0.00)	0.7188 (0.03, 0.00)	0.7031 (0.02, 0.00)	0.5781 (0.02, 0.00)
2	0.9000 (0.00, 0.00)	0.9125 (0.01, 0.00)	0.9062 (0.02, 0.00)	0.8250 (0.01, 0.00)	0.6750 (0.04, 0.00)
3	0.9531 (0.02, 0.00)	0.9062 (0.03, 0.00)	0.8750 (0.00, 0.00)	0.9531 (0.02, 0.00)	0.6719 (0.08, 0.01)
4	0.7396 (0.03, 0.00)	0.5312 (0.03, 0.00)	0.5938 (0.07, 0.01)	0.6562 (0.01, 0.00)	0.4896 (0.01, 0.00)
5	0.7812 (0.00, 0.00)	0.7812 (0.00, 0.00)	0.6328 (0.01, 0.00)	0.6797 (0.05, 0.00)	0.5079 (0.09, 0.01)
6	0.8938 (0.01, 0.00)	0.8625 (0.01, 0.00)	0.7875 (0.01, 0.00)	0.7875 (0.00, 0.00)	0.7375 (0.01, 0.00)
7	0.7916 (0.02, 0.00)	0.8021 (0.01, 0.00)	0.7396 (0.01, 0.00)	0.8646 (0.01, 0.00)	0.7812 (0.01, 0.00)
8	0.8750 (0.00, 0.00)	0.9375 (0.00, 0.00)	0.8125 (0.03, 0.00)	0.7891 (0.01, 0.00)	0.7500 (0.02, 0.00)
9	0.4375 (0.02, 0.00)	0.4141 (0.04, 0.00)	0.4219 (0.00, 0.00)	0.5000 (0.03, 0.00)	0.3125 (0.05, 0.00)
10	0.7266 (0.05, 0.00)	0.6172 (0.01, 0.00)	0.7188 (0.02, 0.00)	0.7422 (0.01, 0.00)	0.4141 (0.07, 0.00)
11	0.7708 (0.00, 0.00)	0.7084 (0.02, 0.00)	0.6562 (0.01, 0.00)	0.6875 (0.04, 0.00)	0.4583 (0.00, 0.00)
12	0.6875 (0.00, 0.00)	0.5156 (0.05, 0.00)	0.5000 (0.03, 0.00)	0.5312 (0.00, 0.00)	0.3281 (0.08, 0.01)
13	0.4531 (0.00, 0.00)	0.4062 (0.03, 0.00)	0.6406 (0.02, 0.00)	0.5859 (0.01, 0.00)	0.3359 (0.01, 0.00)
14	0.7604 (0.01, 0.00)	0.6771 (0.03, 0.00)	0.7500 (0.00, 0.00)	0.6875 (0.02, 0.00)	0.4062 (0.03, 0.00)
15	0.6000 (0.08, 0.01)	0.7125 (0.10, 0.01)	0.4938 (0.04, 0.00)	0.4313 (0.02, 0.00)	0.4188 (0.03, 0.00)
16	0.5547 (0.01, 0.00)	0.5000 (0.08, 0.01)	0.5391 (0.04, 0.00)	0.3750 (0.03, 0.00)	0.5312 (0.05, 0.00)
17	0.6146 (0.03, 0.00)	0.6666 (0.02, 0.00)	0.5833 (0.00, 0.00)	0.4166 (0.02, 0.00)	0.3854 (0.05, 0.00)
18	0.7344 (0.02, 0.00)	0.7344 (0.02, 0.00)	0.7656 (0.08, 0.01)	0.7188 (0.03, 0.00)	0.5625 (0.06, 0.00)
19	0.8958 (0.06, 0.00)	0.8541 (0.08, 0.01)	0.8750 (0.02, 0.00)	0.7396 (0.09, 0.01)	0.6250 (0.02, 0.00)
20	0.8125 (0.00, 0.00)	0.8125 (0.00, 0.00)	0.6562 (0.09, 0.01)	0.8188 (0.04, 0.00)	0.4875 (0.00, 0.00)

Table A13: BPPA (hierarchical) scores by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B
1	0.8438 (0.03, 0.00)	0.7812 (0.00, 0.00)	0.7656 (0.08, 0.01)	0.7344 (0.02, 0.00)	0.6250 (0.02, 0.00)
2	0.8750 (0.01, 0.00)	0.8562 (0.01, 0.00)	0.8625 (0.00, 0.00)	0.7937 (0.03, 0.00)	0.5813 (0.08, 0.01)
3	0.9062 (0.03, 0.00)	0.9062 (0.03, 0.00)	0.9062 (0.00, 0.00)	0.8125 (0.09, 0.01)	0.7500 (0.03, 0.00)
4	0.7188 (0.01, 0.00)	0.6562 (0.01, 0.00)	0.6041 (0.02, 0.00)	0.5938 (0.01, 0.00)	0.5625 (0.06, 0.00)
5	0.6719 (0.02, 0.00)	0.7734 (0.01, 0.00)	0.6406 (0.00, 0.00)	0.6641 (0.01, 0.00)	0.4766 (0.07, 0.00)
6	0.8562 (0.02, 0.00)	0.8687 (0.02, 0.00)	0.8438 (0.02, 0.00)	0.8188 (0.02, 0.00)	0.6687 (0.07, 0.00)
7	0.7812 (0.01, 0.00)	0.8438 (0.01, 0.00)	0.7396 (0.01, 0.00)	0.8021 (0.01, 0.00)	0.6562 (0.01, 0.00)
8	0.8672 (0.01, 0.00)	0.8906 (0.02, 0.00)	0.8281 (0.02, 0.00)	0.7578 (0.10, 0.01)	0.7812 (0.02, 0.00)
9	0.5078 (0.02, 0.00)	0.4766 (0.01, 0.00)	0.5000 (0.02, 0.00)	0.4766 (0.05, 0.00)	0.3672 (0.01, 0.00)
10	0.7422 (0.01, 0.00)	0.7109 (0.01, 0.00)	0.7266 (0.01, 0.00)	0.6641 (0.01, 0.00)	0.5312 (0.03, 0.00)
11	0.7084 (0.08, 0.01)	0.7292 (0.00, 0.00)	0.6666 (0.02, 0.00)	0.6562 (0.01, 0.00)	0.6459 (0.08, 0.01)
12	0.6250 (0.00, 0.00)	0.5937 (0.06, 0.00)	0.4688 (0.03, 0.00)	0.4688 (0.03, 0.00)	0.2812 (0.12, 0.02)
13	0.3438 (0.02, 0.00)	0.4062 (0.02, 0.00)	0.5156 (0.02, 0.00)	0.5079 (0.04, 0.00)	0.3047 (0.07, 0.00)
14	0.7291 (0.02, 0.00)	0.6979 (0.05, 0.00)	0.6354 (0.01, 0.00)	0.6250 (0.02, 0.00)	0.5209 (0.04, 0.00)
15	0.7000 (0.06, 0.00)	0.6562 (0.04, 0.00)	0.4937 (0.02, 0.00)	0.3562 (0.01, 0.00)	0.3938 (0.06, 0.00)
16	0.4609 (0.04, 0.00)	0.5468 (0.03, 0.00)	0.5312 (0.03, 0.00)	0.4063 (0.06, 0.00)	0.5469 (0.02, 0.00)
17	0.6250 (0.08, 0.01)	0.5625 (0.00, 0.00)	0.5625 (0.02, 0.00)	0.5625 (0.02, 0.00)	0.3438 (0.01, 0.00)
18	0.6250 (0.03, 0.00)	0.6562 (0.00, 0.00)	0.6563 (0.06, 0.00)	0.5156 (0.08, 0.01)	0.5312 (0.09, 0.01)
19	0.8333 (0.00, 0.00)	0.7604 (0.03, 0.00)	0.8229 (0.01, 0.00)	0.6875 (0.12, 0.02)	0.5104 (0.03, 0.00)
20	0.7750 (0.04, 0.00)	0.8625 (0.04, 0.00)	0.7250 (0.06, 0.00)	0.6250 (0.04, 0.00)	0.5625 (0.07, 0.01)

Table A14: BPPA (one_by_one) scores by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B
1	0.7656 (0.02, 0.00)	0.7578 (0.01, 0.00)	0.8125 (0.00, 0.00)	0.2188 (0.03, 0.00)	0.6250 (0.08, 0.01)
2	0.8062 (0.02, 0.00)	0.9062 (0.01, 0.00)	0.9000 (0.04, 0.00)	0.3187 (0.07, 0.00)	0.6125 (0.05, 0.00)
3	0.8906 (0.02, 0.00)	0.7969 (0.02, 0.00)	0.8594 (0.02, 0.00)	0.2344 (0.02, 0.00)	0.7344 (0.02, 0.00)
4	0.6562 (0.01, 0.00)	0.4896 (0.01, 0.00)	0.6146 (0.05, 0.00)	0.2083 (0.02, 0.00)	0.4584 (0.04, 0.00)
5	0.7109 (0.04, 0.00)	0.7188 (0.00, 0.00)	0.6875 (0.02, 0.00)	0.1641 (0.09, 0.01)	0.4062 (0.03, 0.00)
6	0.8187 (0.01, 0.00)	0.7688 (0.01, 0.00)	0.8000 (0.06, 0.00)	0.2062 (0.03, 0.00)	0.4750 (0.04, 0.00)
7	0.8021 (0.03, 0.00)	0.8021 (0.01, 0.00)	0.6979 (0.01, 0.00)	0.1354 (0.05, 0.00)	0.5104 (0.07, 0.01)
8	0.8594 (0.02, 0.00)	0.8672 (0.01, 0.00)	0.8046 (0.02, 0.00)	0.3125 (0.02, 0.00)	0.5859 (0.02, 0.00)
9	0.5546 (0.02, 0.00)	0.5157 (0.03, 0.00)	0.3516 (0.05, 0.00)	0.2500 (0.02, 0.00)	0.3516 (0.02, 0.00)
10	0.6718 (0.03, 0.00)	0.5938 (0.00, 0.00)	0.6718 (0.03, 0.00)	0.3203 (0.01, 0.00)	0.4219 (0.02, 0.00)
11	0.6562 (0.01, 0.00)	0.7084 (0.02, 0.00)	0.6771 (0.01, 0.00)	0.1562 (0.01, 0.00)	0.4271 (0.01, 0.00)
12	0.6406 (0.02, 0.00)	0.5781 (0.05, 0.00)	0.5000 (0.06, 0.00)	0.1094 (0.05, 0.00)	0.2812 (0.03, 0.00)
13	0.4765 (0.04, 0.00)	0.4453 (0.01, 0.00)	0.6328 (0.01, 0.00)	0.1875 (0.00, 0.00)	0.3203 (0.01, 0.00)
14	0.6875 (0.02, 0.00)	0.6146 (0.05, 0.00)	0.5104 (0.01, 0.00)	0.2188 (0.01, 0.00)	0.4688 (0.03, 0.00)
15	0.6875 (0.01, 0.00)	0.6313 (0.02, 0.00)	0.5125 (0.02, 0.00)	0.1313 (0.06, 0.00)	0.3750 (0.03, 0.00)
16	0.5312 (0.03, 0.00)	0.5547 (0.01, 0.00)	0.6562 (0.02, 0.00)	0.2266 (0.04, 0.00)	0.4454 (0.02, 0.00)
17	0.5416 (0.02, 0.00)	0.5938 (0.05, 0.00)	0.4062 (0.09, 0.01)	0.2292 (0.00, 0.00)	0.2916 (0.02, 0.00)
18	0.7344 (0.02, 0.00)	0.7344 (0.02, 0.00)	0.6250 (0.06, 0.00)	0.2188 (0.03, 0.00)	0.4531 (0.02, 0.00)
19	0.8750 (0.02, 0.00)	0.6250 (0.04, 0.00)	0.7709 (0.04, 0.00)	0.2083 (0.06, 0.00)	0.4583 (0.00, 0.00)
20	0.7812 (0.01, 0.00)	0.7063 (0.04, 0.00)	0.5000 (0.03, 0.00)	0.2062 (0.02, 0.00)	0.4625 (0.04, 0.00)

Table A15: BPPA (all) scores by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

A.8.3 WBS Scores

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B	Oracle Annotation
1	1.80 (0.75, 0.56)	4.40 (0.49, 0.24)	1.80 (0.75, 0.56)	1.80 (0.75, 0.56)	2.20 (0.75, 0.56)	5.00 (0.00, 0.00)
2	4.80 (0.40, 0.16)	4.20 (0.40, 0.16)	2.80 (0.75, 0.56)	1.80 (0.40, 0.16)	1.60 (0.80, 0.64)	5.00 (0.00, 0.00)
3	4.00 (0.00, 0.00)	1.00 (0.00, 0.00)	2.60 (0.49, 0.24)	1.00 (0.00, 0.00)	2.40 (1.02, 1.04)	5.00 (0.00, 0.00)
4	2.60 (1.50, 2.24)	3.20 (0.40, 0.16)	3.60 (1.02, 1.04)	1.00 (0.00, 0.00)	1.40 (0.49, 0.24)	4.60 (0.49, 0.24)
5	3.40 (1.20, 1.44)	3.20 (0.40, 0.16)	3.40 (0.80, 0.64)	2.60 (0.49, 0.24)	1.60 (0.80, 0.64)	4.80 (0.40, 0.16)
6	4.20 (0.40, 0.16)	2.60 (0.80, 0.64)	2.80 (1.17, 1.36)	3.20 (0.40, 0.16)	1.60 (0.49, 0.24)	4.80 (0.40, 0.16)
7	3.20 (0.75, 0.56)	2.80 (0.75, 0.56)	2.00 (0.89, 0.80)	2.40 (0.49, 0.24)	3.20 (0.98, 0.96)	5.00 (0.00, 0.00)
8	3.00 (0.63, 0.40)	3.40 (0.80, 0.64)	4.00 (0.63, 0.40)	3.00 (0.63, 0.40)	2.40 (0.80, 0.64)	4.20 (0.40, 0.16)
9	3.40 (1.02, 1.04)	2.80 (0.75, 0.56)	1.80 (0.75, 0.56)	1.60 (0.80, 0.64)	1.40 (0.49, 0.24)	4.80 (0.40, 0.16)
10	1.60 (0.80, 0.64)	4.20 (0.75, 0.56)	2.20 (0.75, 0.56)	1.20 (0.40, 0.16)	1.20 (0.40, 0.16)	3.00 (1.10, 1.20)
11	3.40 (1.02, 1.04)	3.20 (0.75, 0.56)	3.20 (0.75, 0.56)	2.80 (0.40, 0.16)	1.80 (0.75, 0.56)	4.00 (0.63, 0.40)
12	3.20 (1.17, 1.36)	3.20 (0.75, 0.56)	2.60 (0.49, 0.24)	2.80 (0.75, 0.56)	1.80 (0.75, 0.56)	5.00 (0.00, 0.00)
13	2.80 (0.75, 0.56)	2.40 (0.80, 0.64)	1.80 (0.75, 0.56)	1.20 (0.40, 0.16)	1.60 (0.49, 0.24)	4.40 (0.80, 0.64)
14	3.80 (0.40, 0.16)	3.80 (0.40, 0.16)	4.00 (0.63, 0.40)	3.40 (0.49, 0.24)	4.00 (0.63, 0.40)	3.80 (0.75, 0.56)
15	5.00 (0.00, 0.00)	3.40 (0.80, 0.64)	2.80 (0.75, 0.56)	2.20 (0.75, 0.56)	2.60 (1.36, 1.84)	4.40 (0.49, 0.24)
16	2.80 (0.75, 0.56)	3.00 (0.63, 0.40)	3.60 (0.49, 0.24)	3.40 (0.49, 0.24)	2.20 (0.98, 0.96)	4.60 (0.49, 0.24)
17	3.80 (0.75, 0.56)	3.60 (1.02, 1.04)	2.60 (0.49, 0.24)	1.80 (1.17, 1.36)	2.20 (0.75, 0.56)	4.40 (0.49, 0.24)
18	2.60 (1.02, 1.04)	2.20 (1.17, 1.36)	2.60 (1.02, 1.04)	2.40 (1.02, 1.04)	2.20 (0.40, 0.16)	4.80 (0.40, 0.16)
19	3.40 (0.49, 0.24)	3.00 (0.63, 0.40)	4.20 (0.40, 0.16)	1.20 (0.40, 0.16)	3.40 (1.02, 1.04)	4.80 (0.40, 0.16)
20	3.00 (0.63, 0.40)	3.00 (0.63, 0.40)	3.00 (0.00, 0.00)	2.00 (0.63, 0.40)	1.80 (0.40, 0.16)	5.00 (0.00, 0.00)

Table A16: Human-scored WBS by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

Motion ID	Claude 3.5 Sonnet	GPT-4o	GPT-4o-mini	GPT-3.5-turbo	Llama-3.1-70B	Oracle Annotation
1	4.20 (1.17, 1.36)	5.00 (0.00, 0.00)	3.20 (0.98, 0.96)	2.20 (0.75, 0.56)	1.80 (0.75, 0.56)	5.00 (0.00, 0.00)
2	4.20 (0.75, 0.56)	5.00 (0.00, 0.00)	1.00 (0.00, 0.00)	1.00 (0.00, 0.00)	1.20 (0.40, 0.16)	5.00 (0.00, 0.00)
3	4.20 (0.98, 0.96)	1.00 (0.00, 0.00)	2.60 (1.20, 1.44)	1.80 (1.60, 2.56)	1.40 (0.49, 0.24)	3.40 (1.96, 3.84)
4	4.00 (1.26, 1.60)	1.60 (0.49, 0.24)	2.60 (0.80, 0.64)	1.00 (0.00, 0.00)	1.00 (0.00, 0.00)	4.60 (0.80, 0.64)
5	2.60 (0.49, 0.24)	4.60 (0.49, 0.24)	2.60 (1.02, 1.04)	2.20 (0.40, 0.16)	1.80 (0.75, 0.56)	1.20 (0.40, 0.16)
6	4.00 (0.89, 0.80)	1.20 (0.40, 0.16)	2.20 (0.40, 0.16)	4.40 (0.80, 0.64)	2.60 (0.49, 0.24)	1.40 (0.49, 0.24)
7	4.60 (0.49, 0.24)	2.60 (0.49, 0.24)	1.00 (0.00, 0.00)	4.00 (1.55, 2.40)	4.60 (0.49, 0.24)	4.60 (0.80, 0.64)
8	4.60 (0.80, 0.64)	2.80 (0.40, 0.16)	5.00 (0.00, 0.00)	2.00 (0.63, 0.40)	1.40 (0.49, 0.24)	4.80 (0.40, 0.16)
9	2.80 (0.75, 0.56)	4.20 (0.98, 0.96)	3.40 (0.80, 0.64)	3.20 (0.75, 0.56)	3.00 (0.00, 0.00)	5.00 (0.00, 0.00)
10	1.00 (0.00, 0.00)	3.00 (0.63, 0.40)	2.20 (0.40, 0.16)	1.00 (0.00, 0.00)	2.00 (0.00, 0.00)	4.00 (1.26, 1.60)
11	4.00 (1.10, 1.20)	3.20 (1.17, 1.36)	4.00 (0.00, 0.00)	2.00 (0.00, 0.00)	3.20 (0.75, 0.56)	3.60 (0.80, 0.64)
12	4.60 (0.80, 0.64)	4.40 (0.80, 0.64)	2.20 (0.75, 0.56)	1.80 (0.40, 0.16)	1.40 (0.49, 0.24)	5.00 (0.00, 0.00)
13	3.80 (1.17, 1.36)	3.40 (0.49, 0.24)	1.20 (0.40, 0.16)	1.00 (0.00, 0.00)	2.60 (0.49, 0.24)	2.80 (0.40, 0.16)
14	2.00 (0.00, 0.00)	2.80 (1.17, 1.36)	2.40 (0.49, 0.24)	2.00 (0.00, 0.00)	5.00 (0.00, 0.00)	4.40 (0.80, 0.64)
15	3.80 (1.17, 1.36)	4.60 (0.80, 0.64)	2.20 (0.40, 0.16)	1.80 (0.40, 0.16)	2.60 (0.49, 0.24)	4.60 (0.80, 0.64)
16	2.00 (0.63, 0.40)	3.00 (1.10, 1.20)	1.80 (0.40, 0.16)	3.20 (1.17, 1.36)	1.40 (0.49, 0.24)	4.40 (0.49, 0.24)
17	5.00 (0.00, 0.00)	4.20 (0.98, 0.96)	4.00 (0.89, 0.80)	2.60 (0.49, 0.24)	3.20 (0.98, 0.96)	4.00 (1.26, 1.60)
18	3.00 (0.00, 0.00)	1.00 (0.00, 0.00)	1.60 (0.49, 0.24)	1.00 (0.00, 0.00)	1.20 (0.40, 0.16)	5.00 (0.00, 0.00)
19	4.20 (0.98, 0.96)	3.80 (0.75, 0.56)	4.80 (0.40, 0.16)	1.80 (1.60, 2.56)	1.60 (0.49, 0.24)	2.80 (0.40, 0.16)
20	4.40 (0.80, 0.64)	3.00 (0.00, 0.00)	4.60 (0.49, 0.24)	4.00 (0.89, 0.80)	2.80 (0.75, 0.56)	3.80 (0.98, 0.96)

Table A17: Gemini 2.5 Pro-scored WBS by motion ID for each LLM. Values represent the mean score with associated standard deviation and variance.

A.9 High-level Planning Errors

Figure A8: Example of wrong action descriptions, generated by GPT-3.5-turbo. In this example, the plan only specifies the process of lifting the foot and grabbing the shoe, deviating from the intention of lifting the shoe and then putting it on.

```
[motion instruction]
lift the right shoe with both hands and put it on in the air

[high-level plan]
[
  {
    "step_number": 1,
    "time_range": [0, 3],
    "movement": "In Step 1, the relevant body parts and their
single-directional movements are:\n1. Right foot: Lifted
towards the hands\n2. Hands: Reach down and grasp the right
shoe",
    "initial_state": "In Step 1, the initial states of relevant
body parts are:\n1. Right foot: Initially on the ground\n2.
Hands: Initially hanging beside the body",
    "final_state": "In Step 1, the final states of relevant body
parts are:\n1. Right foot: Lifted in the air\n2. Hands: Holding
the right shoe"
  }
]
```

Figure A9: Example of incomplete action descriptions, generated by GPT-4o. In this example, the plan does not specify the action of the left arm in order for the eyes to see the watch.

```
[motion instruction]
look down to check the time of the watch on the left wrist

[high-level plan]
[
  {
    "step_number": 1,
    "time_range": [0, 2],
    "movement": "In Step 1, the relevant movements are:\n\n1. The
head tilts downward.\n2. The eyes move downward to focus on the
wrist.",
    "initial_state": "The initial states of the relevant body parts
in Step 1 are:\n\n1. The head is upright and facing forward.\n
2. The eyes are looking straight ahead.\n3. The left arm is
hanging naturally beside the body.\n4. The left wrist is
aligned with the left arm, facing inward.",
    "final_state": "The final states of the relevant body parts in
Step 1 are:\n\n1. The head is tilted downward.\n2. The eyes are
directed downward, focusing on the left wrist.\n3. The left
arm remains hanging naturally beside the body.\n4. The left
wrist remains aligned with the left arm, facing inward."
  }
]
```

A.10 Complexity Analysis

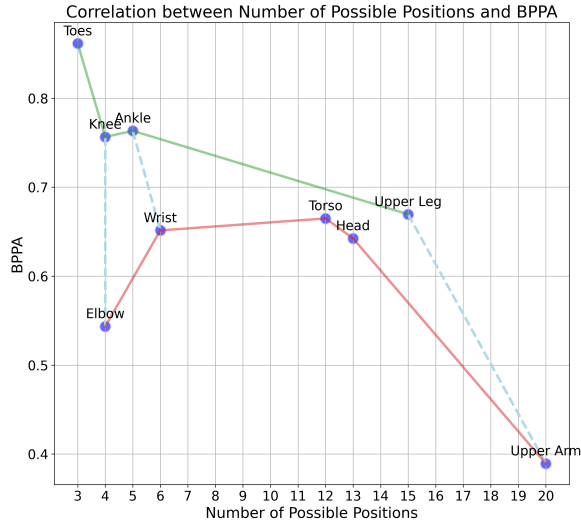


Figure A10: The body-part-wise correlation between the number of possible positions and BPPA. We average the BPPA for paired body parts. Comparison of the lower body performance (green line) and upper body performance (red line) demonstrates that LLMs achieve higher accuracy for lower body parts versus their upper body counterparts.

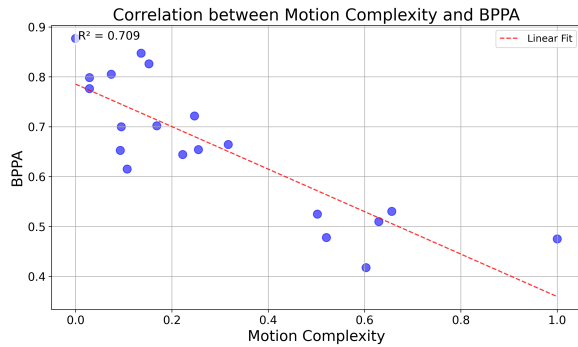


Figure A11: The motion-wise correlation between the motion complexity and BPPA. The motion complexity is calculated as $\sum_{s=1}^N \frac{|moved_s|}{|unmoved_s|}$, where s denotes the step number and $|\cdot|$ represents the count of body parts.

A.11 Generating Raw Avatar Control Parameters

To generate the raw avatar control parameters, we extend our approach by incorporating the SMPL model specifications into the textual summarization, including the initial joint coordinates and descriptions of the joint rotational directions. For querying strategies, we query the LLM to generate an animation plan by sequentially determining relevant joints, rotational directions, and movement quantities which can directly be executed.

Figure A12: Example high-level plan to show imprecise timing (in whole seconds), generated by GPT-4o.

```
[motion instruction]
look down to check the time of
the watch on the left wrist

[high-level plan]
[
  {
    "step_number": 1,
    "time_range": [0, 1],
    "movement": "Tilt the head
downward.",
    "initial_state": "Head
facing forward.",
    "final_state": "Head tilted
downward."
  },
  {
    "step_number": 2,
    "time_range": [1, 2],
    "movement": "Move eyes
downward to check the time
on the watch on the left
wrist.",
    "initial_state": "Eyes
looking forward.",
    "final_state": "Eyes
looking at the left wrist
watch."
  }
]
```

Figure A13: Example high-level plan to show imprecise timing (in whole and half seconds), generated by GPT-4o.

```
[motion instruction]
swing the golf club from right
to left

[high-level plan]
[
  {
    "step_number": 1,
    "time_range": [0, 1],
    "movement": "Raise the golf
club upwards with a
clockwise twist from behind
the right shoulder.",
    "initial_state": "The
```

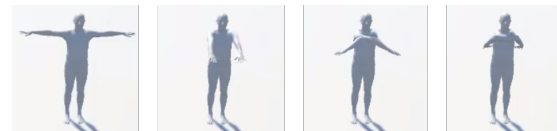
```

    initial position of the
    golf club is behind the
    right shoulder, with the
    right arm extended back.",
    "final_state": "The golf
    club is raised upwards and
    begins moving forward."
  },
  {
    "step_number": 2,
    "time_range": [1, 2.5],
    "movement": "Swing the golf
    club downwards and across
    the front body, moving
    towards the left side.",
    "initial_state": "The golf
    club is raised upwards and
    twisting forward.",
    "final_state": "The golf
    club is positioned over the
    left shoulder, with arms
    crossed in front of the
    body."
  },
  {
    "step_number": 3,
    "time_range": [2.5, 3],
    "movement": "Complete the
    follow-through of the swing
    by bringing the club above
    and behind the left
    shoulder.",
    "initial_state": "The golf
    club is over the left
    shoulder with arms crossed
    .",
    "final_state": "The golf
    club has completed its
    swing and is resting over
    the left shoulder."
  }
]

```



(a) "Toss a ball in the air."



(b) "Hug a person."

Figure A14: Key frames from example animations generated by GPT-4o that struggle to handle SMPL joint rotations, i.e., direction recognition and movement quantity generation. The avatar starts with a T-pose extending arms straight to sides. While the movement directions of joints often deviate from the intended goals, the rotation angles are frequently exaggerated beyond anatomical limits, as shown in the third frame of (a) and the last two frames of (b).

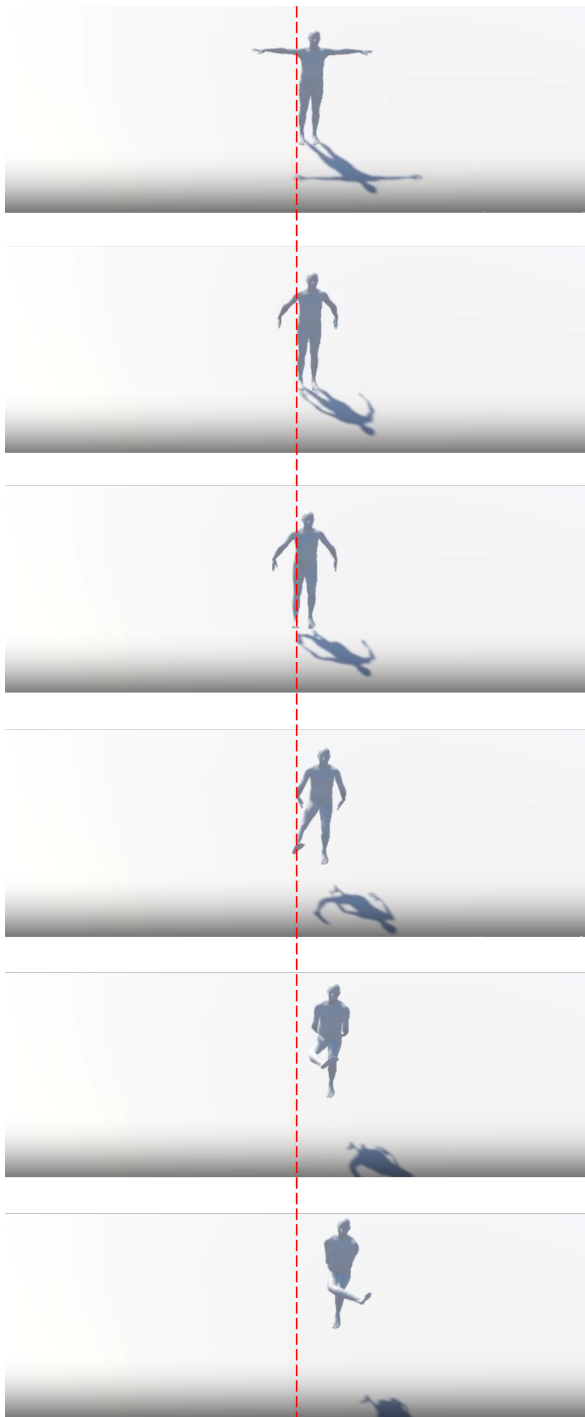


Figure A15: Key frames from an example animation generated by GPT-4o that struggles to handle root movements. The avatar starts with a T-pose extending arms straight to sides. Motion instruction: “Mount a horse.” The generated root movements are linearly approximated, resulting in unrealistic global movements that largely deviate from expected real-world values.