

Multi-Probe Trait-State Latent Modeling for Cross-Context Clinical Motor Assessment

Jiahui Yu¹, Yanxin Cui¹, Xiaowen Lou², You Wu², Yaping Zhang², Ting Pang¹, and Xin Xu^{1,3,4}(✉)

¹ Department of Psychiatry, School of Public Health, The Second Affiliated Hospital of School of Medicine, Zhejiang University, Hangzhou 310058, China

² School of Public Health, Zhejiang University, Hangzhou 310058, China

³ The State Key Lab of Brain-Machine Intelligence, MOE Frontier Science Center for Brain Science and Brain-Machine Integration, Zhejiang University School of Medicine, Hangzhou 310003, China

⁴ Nanhu Brain-Computer Interface Institute, Hangzhou 311100, China
xuxinsummer@zju.edu.cn

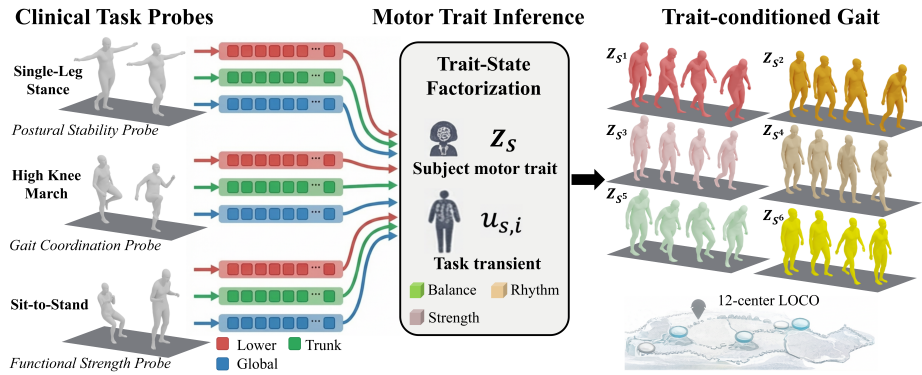


Fig. 1. Overview of cross-context clinical motor assessment. Standardized probes yield context-dependent motor observations. Aggregated multi-probe evidence estimates a subject-level trait, separated from probe-specific transients, to condition gait synthesis for cross-context consistency.

Abstract. A central challenge in clinical motor assessment is to infer stable motor capacity from task observations shaped by context. We propose a multi-probe trait-state latent framework that models each standardized task as a partial observation of a subject-level motor trait. A shared latent captures cross-context regularities while probe-specific latents absorb transient effects, and the trait is inferred via precision-weighted aggregation. Resting-state gait is decoded from the shared

J. Yu and Y. Cui contributed equally.

trait alone, enabling cross-context consistency evaluation. Under subject-disjoint 12-center leave-one-center-out evaluation on 1,196 subjects, the proposed method achieves a 0.60 correlation with held-out real-gait phenotype descriptors, improves biomechanical realism, and outperforms strong conditional baselines. These results support the feasibility of trait-state latent modeling for clinical motor assessment from standardized task probes.

Keywords: Clinical motor assessment · Human motion modeling · Cross-context inference · Motor phenotyping.

1 Introduction

Clinical motor assessment relies on short standardized tasks to probe motor function. Each task provides a context-dependent view shaped by instructions, effort, and environment, whereas clinicians ultimately seek stable motor capacities that persist across contexts [1, 8, 17]. Clinicians implicitly integrate heterogeneous observations to distinguish transient responses from stable motor characteristics [13]. Such cross-context reasoning remains largely qualitative, limiting reproducibility across clinicians, settings, and centers [2, 7, 23].

Recent advances in parametric human motion modeling and generative learning enable rich representation and synthesis of human movement [11, 15, 16, 19, 24]. However, most existing approaches remain context-bound: they recognize or generate motion within a fixed task setting and implicitly treat observed performance as a direct proxy for ability [3–5, 10, 20–22, 26]. This assumption is insufficient in clinical assessment, where the same subject may show task-induced variability while retaining stable motor characteristics [9], and where fatigue, instructions, and environment can shift observed performance [6, 14]. This context-bound modeling constrains cross-center reproducibility and cross-context comparability [18], motivating the trait-state framework in Figure 1 to separate shared subject traits from task-specific transients.

Specifically, we cast clinical motor assessment as multi-probe latent trait inference. Each standardized task is treated as a partial, context-dependent probe of motor capacity. The model uses a shared subject latent to capture cross-context motor regularities and probe-specific latents to absorb execution effects; only the shared latent conditions target-domain gait generation. Precision-weighted aggregation fuses probe-wise evidence, making gait synthesis a structural test of cross-context consistency rather than a standalone generation goal.

We make three contributions. (1) **Trait-state formulation.** We model clinical motor assessment as latent trait inference that disentangles stable motor traits from task-specific transients. (2) **Cross-context validation paradigm.** We use trait-conditioned gait synthesis as a principled test of trait generalization rather than direct task-to-task translation. (3) **Multi-center evaluation.** Under subject-disjoint 12-center LOCO evaluation, our method consistently outperforms strong conditional generation baselines across all primary metrics.

2 Method

2.1 Problem Setup and Motion Representation

For each subject s , we collect a set of clinically guided task motions $\mathcal{X}_s^{\text{task}} = \{\mathbf{x}_{s,i}^{\text{task}}\}_{i=1}^{N_s}$, where $N_s = 3$ denotes the three standardized probes: single-leg stance (SLS), sit-to-stand (STS), and high knees (HK). When available, we also collect a resting-state gait sequence $\mathbf{x}_s^{\text{gait}}$, which is used as the target domain for cross-context consistency evaluation. Each motion is represented as a parametric trajectory using the neutral SMPL-X [12] body model. For subject s , a motion sequence of T frames is represented as:

$$\mathbf{x} = \{(\boldsymbol{\theta}_t, \mathbf{r}_t, \mathbf{t}_t, \boldsymbol{\beta}_s)\}_{t=1}^T, \quad (1)$$

where $\boldsymbol{\theta}_t \in \mathbb{R}^{165}$ denotes pose parameters (joint rotations in axis-angle representation), $\mathbf{r}_t \in \mathbb{R}^3$ denotes root orientation, $\mathbf{t}_t \in \mathbb{R}^3$ denotes root translation, and $\boldsymbol{\beta}_s \in \mathbb{R}^{300}$ is the subject-specific body shape parameter vector, which remains constant across time for subject s . The temporal motion input is therefore formed by stacking pose, root orientation, and root translation into $\mathbf{x} \in \mathbb{R}^{T \times D}$ with $D = 171$. The shape vector $\boldsymbol{\beta}_s$ is kept for SMPL-X reconstruction but is not included in the temporal token input.

2.2 Compositional Motion Tokenization

Task and gait motions are tokenized into a shared discrete space using component-wise VQ-VAE codebooks trained on training-center sequences. Each motion $\mathbf{x}$ is factorized into three components, $\mathbf{x}^{\text{lower}}$, $\mathbf{x}^{\text{trunk}}$, and $\mathbf{x}^{\text{global}}$. For each component $k \in \{\text{lower}, \text{trunk}, \text{global}\}$, a VQ-VAE [25] produces a token stream $\mathbf{q}_{1:T}^{(k)}$.

To feed multi-stream tokens into a single transformer, we interleave the three streams per frame:

$$\mathbf{q}_{1:3T} = \text{Flat}\left(\{(q_t^{\text{lower}}, q_t^{\text{trunk}}, q_t^{\text{global}})\}_{t=1}^T\right). \quad (2)$$

Learned embeddings encode stream and time identities. Given generated gait tokens $\hat{\mathbf{q}}_s^{\text{gait}}$, we decode each component and compose them into SMPL-X trajectories,

$$\tilde{\mathbf{x}}_s^{\text{gait}} = \text{Compose}(\{D^{(k)}(\hat{\mathbf{q}}_s^{(k)})\}_k). \quad (3)$$

2.3 Trait-State Latent Factorization

We model each task motion as a controlled probe that reveals a stable, subject-specific motor trait while containing task-induced transients. To separate these factors, we introduce (i) a subject-level latent trait $\mathbf{z}_s$ shared across task probes and (ii) a task-level latent state $\mathbf{u}_{s,i}$ capturing task-specific transients, with

Gaussian priors $p(\mathbf{z}_s) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $p(\mathbf{u}_{s,i}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. We parameterize the token likelihoods with autoregressive sequence models:

$$p_\theta(\mathbf{q}_{s,i}^{\text{task}} \mid \mathbf{z}_s, \mathbf{u}_{s,i}, c_{s,i}) = \prod_{t=1}^{L_i} p_\theta(q_{s,i,t}^{\text{task}} \mid \mathbf{q}_{s,i,<t}^{\text{task}}, \mathbf{z}_s, \mathbf{u}_{s,i}, c_{s,i}), \quad (4)$$

$$p_\theta(\mathbf{q}_s^{\text{gait}} \mid \mathbf{z}_s) = \prod_{t=1}^{L_g} p_\theta(q_{s,t}^{\text{gait}} \mid \mathbf{q}_{s,<t}^{\text{gait}}, \mathbf{z}_s), \quad (5)$$

where $c_{s,i}$ denotes the task context label (clinical protocol identifier). Equivalently, the gait decoder follows the factorization $p_\theta(\mathbf{q}_s^{\text{gait}} \mid \mathbf{z}_s, \mathbf{u}_{s,i}) = p_\theta(\mathbf{q}_s^{\text{gait}} \mid \mathbf{z}_s)$, so $\mathbf{u}_{s,i}$ has no direct conditioning path to gait synthesis.

We perform amortized variational inference with posteriors $q_\phi(\mathbf{z}_s \mid \mathcal{Q}_s^{\text{task}})$ and $q_\phi(\mathbf{u}_{s,i} \mid \mathbf{q}_{s,i}^{\text{task}})$. Figure 2 shows the full pipeline.

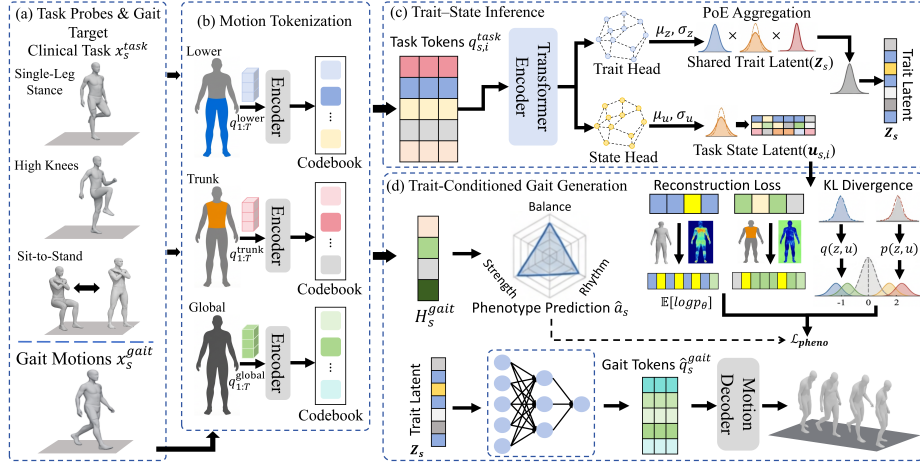


Fig. 2. Trait-state modeling for clinical motor assessment. (a) Standardized task probes and resting-state gait define the cross-context setting. (b) SMPL-X task and gait motions are tokenized by shared component-wise VQ-VAE codebooks into lower-body, trunk, and global tokens. (c) Task tokens produce probe-wise trait and state posteriors, with trait posteriors fused by product-of-experts aggregation into the subject-level latent $\mathbf{z}_s$. (d) The aggregated trait latent conditions gait-token generation and SMPL-X reconstruction, supervised by token reconstruction, KL regularization, and phenotype alignment.

2.4 Probe-wise Trait Inference and Product-of-Experts Aggregation

For each task probe, we use a transformer encoder f_{enc} to map the task-token sequence to a fixed-dimensional representation $h_{s,i} = f_{\text{enc}}(q_{s,i}^{\text{task}}) \in \mathbb{R}^m$. Two

MLP heads estimate the trait and state posteriors:

$$q_\phi(z_s | q_{s,i}^{\text{task}}) = \mathcal{N}(\mu_{s,i}^z, \text{diag}((\sigma_{s,i}^z)^2)), \quad (\mu_{s,i}^z, \log \sigma_{s,i}^z) = f_{\mu\sigma}^z(h_{s,i}), \quad (6)$$

$$q_\phi(u_{s,i} | q_{s,i}^{\text{task}}) = \mathcal{N}(\mu_{s,i}^u, \text{diag}((\sigma_{s,i}^u)^2)), \quad (\mu_{s,i}^u, \log \sigma_{s,i}^u) = f_{\mu\sigma}^u(h_{s,i}). \quad (7)$$

The trait posterior estimates subject-level information recoverable from one probe, while the state posterior captures probe-specific execution patterns.

The three probes provide complementary evidence about the same subject-level trait. To obtain a robust subject-level trait posterior, we aggregate the per-probe trait posteriors $\{q_\phi(\mathbf{z}_s | \mathbf{q}_{s,i}^{\text{task}})\}_{i=1}^3$ using a product-of-experts (PoE) rule:

$$q_\phi(z_s | \mathcal{Q}_s^{\text{task}}) = \mathcal{N}(\mu_s, \Sigma_s), \quad \Sigma_s = \left(I + \sum_{i=1}^3 \Lambda_{s,i} \right)^{-1}, \quad \mu_s = \Sigma_s \sum_{i=1}^3 \Lambda_{s,i} \mu_{s,i}^z, \quad (8)$$

where $\Lambda_{s,i} = \text{diag}((\sigma_{s,i}^z)^{-2})$ is the precision matrix. This precision-weighted fusion gives higher weight to lower-uncertainty probes. The aggregated posterior $q_\phi(z_s | \mathcal{Q}_s^{\text{task}})$ is used to condition gait generation.

2.5 Variational Objective and Clinical Alignment

The training objective combines variational reconstruction and phenotype-guided regularization.

Variational objective. For each subject s , we maximize an evidence lower bound consisting of task-token reconstruction conditioned on $(z_s, u_{s,i})$, gait-token reconstruction conditioned on z_s when gait is available, and KL regularization on both latents:

$$\begin{aligned} \mathcal{L}_{\text{ELBO}}(s) = & \sum_i \mathbb{E}_{q_\phi(z_s, u_{s,i})} [\log p_\theta(q_{s,i}^{\text{task}} | z_s, u_{s,i}, c_{s,i})] \\ & + \delta_s \mathbb{E}_{q_\phi(z_s)} [\log p_\theta(q_s^{\text{gait}} | z_s)] \\ & - \text{KL}(q_\phi(z_s) \| p(z_s)) - \sum_i \text{KL}(q_\phi(u_{s,i}) \| p(u_{s,i})). \end{aligned} \quad (9)$$

Phenotype-guided regularization. We compute nine phenotype indicators $a_s \in \mathbb{R}^9$ describing balance, strength, and rhythm. The nine descriptors quantify stride length, cadence, gait-cycle frequency, stride-time variability, step-width variability, trunk sway, foot clearance, pelvis velocity, and left-right asymmetry.

Given gait-model hidden states $H_s^{\text{gait}} \in \mathbb{R}^{L \times d}$, a linear head predicts:

$$\hat{a}_s = g_\psi \left(\frac{1}{L} \sum_{t=1}^L H_{s,t}^{\text{gait}} \right), \quad (10)$$

The phenotype loss is:

$$\mathcal{L}_{\text{pheno}} = \sum_{s: \delta_s=1} \|\hat{a}_s - a_s\|_2^2. \quad (11)$$

Final objective. The overall training objective is:

$$\mathcal{L} = \sum_s \mathcal{L}_{\text{ELBO}}(s) + \lambda_p \mathcal{L}_{\text{pheno}}. \quad (12)$$

3 Experiments and Results

3.1 Dataset and Protocol

We curate a 12-center clinical motion dataset with 1,196 subjects performing three clinician-guided probes: single-leg stance, sit-to-stand, and high knee lifting. When available, each subject also provides up to two resting gait clips. Videos are recorded at 30 FPS and converted to SMPL-X trajectories through a unified fitting pipeline, yielding approximately 8,000 motion sequences totaling 63.1 hours and spanning ages 19 to 93, with 639 male and 557 female participants.

Resting gait is not available for all subjects; we denote gait availability by $\delta_s \in \{0, 1\}$ and report gait-based evaluation on subjects with $\delta_s = 1$ only. We evaluate cross-center generalization using subject-disjoint leave-one-center-out (LOCO) over 12 centers. To prevent leakage, both the tokenizer (VQ-VAE codebooks) and all model components are trained using training-center data only. Within each LOCO fold, held-out center data are never used for tokenizer training, model fitting, or hyperparameter selection.

3.2 Baselines

Uncond. A transformer trained on gait tokens only, without any conditioning. **Direct $\mathbf{T} \rightarrow \mathbf{G}$.** A paired transformer mapping task tokens to gait tokens without an explicit subject-level trait latent. **CondGen (Diff.).** A conditional diffusion model generating gait tokens from the same task embedding. **Concat (no latent).** A transformer that concatenates the three task token streams and directly generates gait tokens, without $\mathbf{z}_s/\mathbf{u}_{s,i}$ factorization or PoE aggregation. **Ours w/o $\mathcal{L}_{\text{pheno}}$.** Our model without phenotype-guided regularization. **Fairness.** All methods share the same tokenizer/decoder, LOCO splits, and training budget.

3.3 Evaluation Metrics

Following GAITGen and clinical gait analysis practice, we evaluate generated gait by descriptor fidelity, biomechanical realism, and subject consistency.

Descriptor Fidelity. For each paired test subject with $\delta_s = 1$, we extract a 9-dimensional descriptor vector from real and generated gait, denoted by $\mathbf{d}_s^{\text{real}}$ and $\mathbf{d}_s^{\text{gen}}$. When multiple gait clips are available, descriptors are averaged at the subject level. We report Pearson correlation (Corr.) and mean absolute error (MAE) between subject-matched descriptors.

Table 1. Main LOCO results, reported as fold-averaged means over 12 held-out centers. Descriptor fidelity is the primary objective; retrieval is reported as optional reference. $\uparrow$ higher is better, except MAE/Foot/Cycle ($\downarrow$). UB denotes the empirical upper bound from real gait clips.

Method	Descriptor fidelity		Biomechanics		Consistency
	Corr. $\uparrow$	MAE $\downarrow$	Foot $\downarrow$	Cycle $\downarrow$	Acc@1 $\uparrow$
Real gait (UB)	–	–	0.03	0.07	–
Uncond.	0.23	0.79	0.09	0.17	0.18
Direct T $\rightarrow$ G	0.35	0.65	0.07	0.14	0.32
Concat (no latent)	0.39	0.61	0.07	0.13	0.37
CondGen (Diff.)	0.45	0.55	0.05	0.11	0.52
Ours w/o $\mathcal{L}_{pheno}$	0.52	0.47	0.05	0.10	0.58
Ours (Full)	0.60	0.38	0.04	0.09	0.72

Biomechanical Realism. *Foot Sliding* is the average foot velocity during estimated ground-contact frames, normalized by body scale. *Cycle Error* is the relative deviation of dominant gait frequency, $|f_{\text{gen}} - f_{\text{real}}|/f_{\text{real}}$, where f is estimated from pelvis forward velocity.

Subject Consistency. We train a gait encoder $E(\cdot)$ on training subjects only and freeze it to retrieve the matched real gait from the test gallery, reporting Top-1 accuracy.

Table 2. Robustness under incomplete clinical observations (fold-averaged means over LOCO splits). Top: missing task probes at inference. Bottom: limited paired supervision during training.

Setting	Foot $\downarrow$	Top-1 $\uparrow$	Corr. $\uparrow$
<i>Missing task probes (inference)</i>			
1 task	0.050	0.48	0.34
2 tasks	0.043	0.57	0.45
3 tasks	0.038	0.72	0.60
<i>Limited paired subjects (training)</i>			
100% paired	0.038	0.72	0.60
50% paired	0.041	0.60	0.49
20% paired	0.046	0.54	0.42
10% paired	0.051	0.49	0.35

3.4 Main Results

Table 1 shows that our full model achieves the best descriptor fidelity and biomechanical realism. Compared with CondGen (Diff.), it improves Corr. by 0.15 and

Table 3. Trait diagnostics under LOCO (avg. over folds). For $\mathbf{z}$, lower task prediction indicates weaker task leakage; for $\mathbf{u}$, higher task prediction indicates stronger probe-specific encoding. Lower is better for Intra-subj Var.

Method	Task pred. from $\mathbf{z} \downarrow$	Task pred. from $\mathbf{u} \uparrow$	Intra-subj Var(μ^z) $\downarrow$
Direct T $\rightarrow$ G	0.61	–	0.083
CondGen (Diff.)	0.57	–	0.072
Ours (Full)	0.27	0.92	0.035

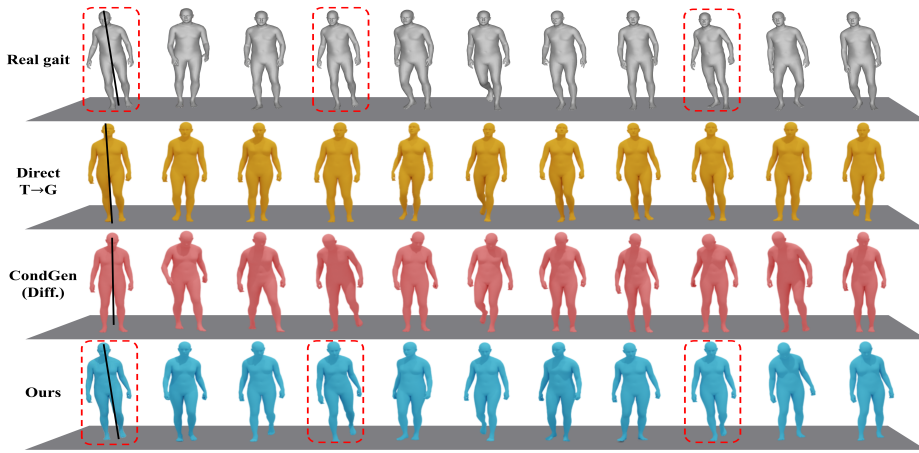


Fig. 3. Qualitative comparison. Phase-aligned frames from one gait cycle of a representative subject. Our method preserves the subject-specific rightward trunk lean more faithfully than CondGen (Diff.). Black lines indicate the estimated trunk axis.

Top-1 by 0.20, while reducing Foot and Cycle errors. Removing $\mathcal{L}_{\text{pheno}}$ lowers Corr. and Top-1, and Concat remains below the factorized model.

Table 2 shows monotonic gains as more probes are observed and graceful degradation under reduced paired supervision, with usable performance even at 10% paired subjects.

3.5 Trait Disentanglement and Clinical Relevance

Table 3 diagnoses the intended factorization. Our shared latent $\mathbf{z}$ is less task-predictive than baseline representations, while $\mathbf{u}$ remains strongly task-predictive. The intra-subject variance of per-probe trait estimates also decreases to 0.035, compared with 0.083 and 0.072 in the baselines.

Clinically, cross-protocol comparability demands a shared latent that is stable and minimally task-entangled; our lower-variance, less task-predictive $\mathbf{z}$ better supports aggregation across heterogeneous probes. Figure 3 further shows that the inferred trait preserves a phase-consistent trunk-lean pattern for the same

subject, providing qualitative evidence of subject-specific motor preservation across contexts.

4 Conclusion

We presented a multi-probe trait–state framework for cross-context clinical motor assessment. By conditioning gait generation on the shared trait latent alone, the model improves descriptor fidelity, biomechanical realism, and subject consistency under subject-disjoint 12-center LOCO evaluation.

Acknowledgments. This work was partly supported by the Zhejiang Provincial Key Research and Development Program of China under Grant 2025C02108 and the National Natural Science Foundation of China under Grant 62406280. This work was also partly supported by the Nanhu Brain-computer Interface Institute, Hangzhou, China.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cho, H.J., Kim, S., Yu, H., Jeong, S., Kang, K.: Quantitative analysis of gait and balance using deep learning on monocular videos and the timed up and go test in idiopathic normal-pressure hydrocephalus. *Frontiers in Aging Neuroscience* **17**, 1644543 (2025)
2. Clement, A., Willoughby, J., Voiculescu, I.: Confidence in angle predictions for clinical decision support. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 116–124. Springer (2025)
3. Feng, Y., Lin, J., Dwivedi, S.K., Sun, Y., Patel, P., Black, M.J.: Chatpose: Chatting about 3d human pose. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 2093–2103 (2024)
4. Jiang, B., Chen, X., Liu, W., Yu, J., Yu, G., Chen, T.: Motiongpt: Human motion as a foreign language. *Advances in Neural Information Processing Systems* **36**, 20067–20079 (2023)
5. Jin, Y., Haworth, B.: Back to basics: Motion representation matters for human motion generation using diffusion model. *arXiv preprint arXiv:2512.04499* (2025)
6. Jones, C., Castro, D.C., De Sousa Ribeiro, F., Oktay, O., McCradden, M., Glocker, B.: A causal perspective on dataset bias in machine learning for medical imaging. *Nature Machine Intelligence* **6**(2), 138–146 (2024)
7. Kenny, L., Azizi, Z., Moore, K., Alcock, M., Heywood, S., Jonsson, A., McGrath, K., Foley, M.J., Sweeney, B., O’Sullivan, S., et al.: Inter-rater reliability of hand motor function assessment in parkinson’s disease: Impact of clinician training. *Clinical Parkinsonism & Related Disorders* **11**, 100278 (2024)
8. Kirk, C., Packer, E., Polhemus, A., MacLean, M.K., Bailey, H., Kluge, F., Gaßner, H., Rochester, L., Del Din, S., Yarnall, A.J.: A systematic review of real-world gait-related digital mobility outcomes in parkinson’s disease. *npj Digital Medicine* **8**(1), 585 (2025)
9. Kowalski, K.L., Boolani, A., Christie, A.D.: Sex differences in the impact of state and trait fatigue on gait variability. *Human Movement Science* **80**, 102884 (2021)

10. Li, J., Cao, J., Zhang, H., Rempe, D., Kautz, J., Iqbal, U., Yuan, Y.: Genmo: A generalist model for human motion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11766–11776 (2025)
11. Meng, Z., Xie, Y., Peng, X., Han, Z., Jiang, H.: Rethinking diffusion for text-driven human motion generation: Redundant representations, evaluation, and masked autoregression. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 27859–27871 (2025)
12. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3d hands, face, and body from a single image. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10975–10985 (2019)
13. Rodríguez-Molinero, A., Pérez-López, C., Caballol, N., Buongiorno, M., Ávila Rivera, M.A., López Ariztegui, N., López-Manzanares, L., Hernández-Vara, J., Bayes-Rusiñol, À., Gironell Carreró, A., et al.: Parkinson’s disease medication adjustments based on wearable device information compared to other methods: randomized clinical trial. *npj Parkinson’s Disease* **11**(1), 249 (2025)
14. Romijnders, R., Atrsaei, A., Rehman, R.Z.U., Strehlow, L., Massoud, J., Hinchliffe, C., Macrae, V., Emmert, K., Reilmann, R., Janneke van der Woude, C., et al.: Association of real life postural transitions kinematics with fatigue in neurodegenerative and immune diseases. *NPJ digital medicine* **8**(1), 12 (2025)
15. Ruiz-Ponce, P., Barquero, G., Palmero, C., Escalera, S., García-Rodríguez, J.: Mix-ermdm: Learnable composition of human motion diffusion models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12380–12390 (2025)
16. Ruth, P.S., Uhlrich, S.D., de Monts, C., Falisse, A., Muccini, J., Covitz, S., Vogt-Domke, S., Day, J., Duong, T., Delp, S.L.: Video-based biomechanical analysis captures disease-specific movement signatures of different neuromuscular diseases. *NEJM AI* **2**(9), A10a2401137 (2025)
17. Segado, M., Parodi, F., Matelsky, J.K., Platt, M.L., Dyer, E.B., Kording, K.P.: Grounding intelligence in movement. *arXiv preprint arXiv:2507.02771* (2025)
18. Souza, R., Stanley, E.A., Gulve, V., Moore, J., Kang, C., Camicioli, R., Monchi, O., Ismail, Z., Wilms, M., Forkert, N.D.: Harmonytm: multi-center data harmonization applied to distributed learning for parkinson’s disease classification. *Journal of Medical Imaging* **11**(5), 054502–054502 (2024)
19. Tan, T., Van Wouwe, T., Werling, K.F., Liu, C.K., Delp, S.L., Hicks, J.L., Chaudhari, A.S.: Gaitdynamics: A generative foundation model for analyzing human walking and running. *Nature Biomedical Engineering* pp. 1–13 (2026)
20. Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., Bermano, A.H.: Human motion diffusion model. *arXiv preprint arXiv:2209.14916* (2022)
21. Van Wouwe, T., Lee, S., Falisse, A., Delp, S., Liu, C.K.: Diffusionposer: Real-time human motion reconstruction from arbitrary sparse sensors using autoregressive diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2513–2523 (2024)
22. Wang, Z., Wang, J., Li, Y., Lin, D., Dai, B.: Intercontrol: Zero-shot human interaction generation by controlling every joint. *Advances in Neural Information Processing Systems* **37**, 105397–105424 (2024)
23. Werling, K., Bianco, N.A., Raitor, M., Stingel, J., Hicks, J.L., Collins, S.H., Delp, S.L., Liu, C.K.: Addbiomechanics: Automating model scaling, inverse kinematics, and inverse dynamics from human motion data through sequential optimization. *Plos one* **18**(11), e0295152 (2023)

24. Xiang, L., Gao, Z., Yu, P., Fernandez, J., Gu, Y., Wang, R., Gutierrez-Farewik, E.M.: Explainable artificial intelligence for gait analysis: advances, pitfalls, and challenges-a systematic review. *Frontiers in Bioengineering and Biotechnology* **13**, 1671344 (2025)
25. Zhang, J., Zhang, Y., Cun, X., Zhang, Y., Zhao, H., Lu, H., Shen, X., Shan, Y.: Generating human motion from textual descriptions with discrete representations. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 14730–14740 (2023)
26. Zhou, Z., Liang, J., Peng, Z., Fan, C., An, F., Yu, S.: Gait patterns as biomarkers: A video-based approach for classifying scoliosis. In: *International conference on medical image computing and computer-assisted intervention*. pp. 284–294. Springer (2024)