Wearable Intelligence for Healthcare Robotics: from Brain Activity to Body Movements

Jiankai Sun[1]
[1]Stanford University
450 Jane Stanford Way
Stanford, CA 94305
jksun@stanford.edu

Abstract

My research aims to **pioneer efficient and reliable wearable intelligence algorithms that transform healthcare robotics into adaptive, patient-centered systems**. I take a four-step approach: (1) design multimodal wearable sensing platforms to capture human and biometric signals; (2) train a foundation model that learns from these rich datasets to reason about human behaviors and health states; (3) validate the model through large-scale simulation and principled uncertainty quantification; and (4) deploy it in rehabilitation and assistive robots for intelligent, personalized care. This research not only advances fundamental understanding of multimodal human behavior, but also opens new pathways for early disease diagnosis, adaptive treatment, and accessible digital health. By bridging AI, wearables, and robotics, my work aspires to lay the groundwork for the next generation of healthcare technologies that are proactive, trustworthy, and deeply aligned with human well-being.

Vision, Motivation, and Introduction

What will be the dominant personal computing platform in the next decade? We anticipate that it will not be smartphones, but wearable devices such as smartwatches and smart glasses. The convergence of AI, wearable technologies, and robotics marks a paradigm shift in healthcare, opening the door to personalized, continuous, and pervasive digital health solutions[1]. Our vision is to bridge the gap between AI insights, wearable intelligence, and robotic precision, thereby creating healthcare systems seamlessly tailored to individual needs.

Recent years have witnessed remarkable progress in robotics for manufacturing (Sun et al. 2021a), healthcare (Qiu et al. 2022, 2023), and service delivery. Within this landscape, wearable intelligence is emerging as a crucial enabler in healthcare robotics: decoding brain signals for prosthetic control, monitoring physiological signals for personalized care, and providing real-time assessments for decision support. For instance, we developed an algorithm that forecasts trajectories of egocentric camera wearers in crowded spaces. This ability not only supports navigation assistance for the visually impaired but also transfers human navigation behavior to mobile robots, fostering safer and more socially compliant human-robot interactions (Qiu et al. 2022). Nonetheless, deploying such systems in real-world healthcare remains challenging (Sun et al. 2023b), underscoring the need for a generalizable multimodal foundation model capable of modeling complex human behaviors and biometric signals.

[1]I am co-organizing a workshop dedicated to this theme: https://icrawearable.github.io/.

Progress to Date

Multimodal Wearable Sensing Reliable wearable intelligence requires diverse and complementary data sources. I developed a multimodal sensing system using Project Aria, capable of collecting RGB video, eye-tracking, audio, barometric data, GPS, Wi-Fi, Bluetooth, SLAM, and IMU signals (Sun et al. 2023a). This pioneering dataset effort lays the foundation for future research in wearable technologies and their integration with healthcare robotics (Sun et al. 2025b; Qiu et al. 2023; Firoozi et al. 2023; Sun et al. 2024b).

Foundation Models for Wearable Intelligence Building on my prior work in human motion analysis (LocATe (Sun et al. 2024a)) and learning robot policies from human play data (Wang et al. 2023), I am constructing a multimodal foundation model that captures human behavior from wearable data. This model overcomes the limitations of unimodal representations (Sun et al. 2021b) through attention-based architectures (Zheng, Sun et al. 2025a; Zheng et al. 2025; Zheng, Sun et al. 2025b; Ren et al. 2025), and provides a reasoning backbone for human-robot collaboration (Zheng et al. 2024; Zhang et al. 2025; Hong et al. 2025b,a; **?**).

Human Dynamics and Simulation Simulation is essential for testing and transferring wearable intelligence to real-world healthcare applications. We contributed to BEHAVIOR-1K (Li et al. 2023), which models 1,000 everyday activities in household and office environments, covering 5,000 objects with physical and semantic properties. Building on this, I extend simulators to better capture human dynamics and integrate uncertainty quantification (Sun et al. 2023b; Mandyam et al. 2025) for reliable deployment in healthcare scenarios.

Integration into Healthcare Robotics The validated foundation model is deployed in real-world healthcare robots, including exoskeletons and robotic prostheses, with a focus on intelligent rehabilitation and personalized assistance (Qiu et al. 2022).

Planned Work

To further enrich multimodal data, I plan to incorporate biometric signals such as EEG, EOG, and EMG, alongside body motion capture using Rokoko and standard human body models (SMPL, SMPL-X).

Contributions

My primary contributions include: (1) robotics, including ARCH (Sun et al. 2025a, 2023c), PlanCP (Sun et al. 2023b); (2) world models, including PlaTe (Sun et al. 2022), MAIRL (Sun et al. 2021a), GSplat (Sun et al. 2023a), and methods for multimodal representation (Sun et al. 2021b, 2023d, 2025b); and (3) human performance modeling with LocATe (Sun et al. 2024a). Collaborator-led projects include 3D world models (Sun et al. 2023e; Shorinwa, Sun et al. 2025; Huang et al. 2025; Li et al. 2025) and healthcare robotics (Qiu et al. 2022, 2023).

Significance

This research stands at the intersection of robotics, AI, and healthcare, addressing an area that remains largely unexplored: integrating multimodal foundation models into wearable intelligence for healthcare robotics.

References

Firoozi, R.; Tucker, J.; Tian, S.; Majumdar, A.; Sun, J.; et al. 2023. Foundation Models in Robotics: Applications, Challenges, and the Future. *IJRR*.

Hong, J.; et al. 2025a. Benchmarking the Thinking Mode of Multimodal Large Language Models in Clinical Tasks. *arXiv preprint arXiv:2511.03328*.

Hong, J.; et al. 2025b. Diagnosing Shoulder Disorders Using Multimodal Large Language Models and Consumer-Grade Cameras. *arXiv preprint arXiv:2510.09230*.

Huang, S.; Chen, Q.; Zhang, X.; Sun, J.; et al. 2025. Particle-Former: A 3D Point Cloud World Model for Multi-Object, Multi-Material Robotic Manipulation. *CoRL*.

Li, C.; Zhang, R.; Wong, J.; Gokmen, C.; Srivastava, S.; Martín-Martín, R.; Wang, C.; Levine, G.; Lingelbach, M.; Sun, J.; et al. 2023. Behavior-1k: A benchmark for embodied ai with 1,000 everyday activities and realistic simulation. In *CoRL*, 80–93. PMLR.

Li, W.; Xu, Y.; Yao, M.; Liang, F.; Sun, J.; Wang, M.; Zhang, G.; Huang, L.; and Li, H. 2025. NopeRoomGS: Indoor 3D Gaussian Splatting Optimization without Camera Pose Input. In *NeurIPS*.

Mandyam, A.; Meng, J.; Gao, G.; Sun, J.; Schwager, M.; Engelhardt, B. E.; and Brunskill, E. 2025. Perry: Policy evaluation with confidence intervals using auxiliary data. *arXiv preprint arXiv:2507.20068*.

Qiu, J.; Chen, L.; Gu, X.; Lo, F. P.-W.; Tsai, Y.-Y.; Sun, J.; et al. 2022. Egocentric human trajectory forecasting with a wearable camera and multi-modal fusion. *RA-L*.

Qiu, J.; Li, L.; Sun, J.; et al. 2023. Large AI Models in Health Informatics: Applications, Challenges, and the Future. *JBHI*, 27(12): 6074–6087.

Ren, L.; Chen, C.; Xu, H.; Kim, Y. J.; Atkinson, A.; Zhan, Z.; Sun, J.; et al. 2025. Decoder-hybrid-decoder architecture for efficient reasoning with long generation. *NeurIPS*.

Shorinwa, O.; Sun, J.; et al. 2025. Siren: Semantic, initialization-free registration of multi-robot gaussian splatting maps. *CoRL*.

Sun, J.; Huang, D.-A.; Lu, B.; Liu, Y.-H.; Zhou, B.; and Garg, A. 2022. PlaTe: Visually-grounded planning with transformers in procedural tasks. *RA-L*.

Sun, J.; et al. 2021a. Adversarial Inverse Reinforcement Learning With Self-Attention Dynamics Model. *RA-L*.

Sun, J.; et al. 2021b. HiABP: Hierarchical Initialized ABP for Unsupervised Representation Learning. In *AAAI*, volume 35, 9747–9755.

Sun, J.; et al. 2023a. Aria-NeRF: Multimodal Egocentric View Synthesis. *ICRA-WIHR Workshop*.

Sun, J.; et al. 2023b. Conformal prediction for uncertainty-aware planning with diffusion dynamics model. In *NeurIPS*.

Sun, J.; et al. 2023c. Connected Autonomous Vehicle Motion Planning with Video Predictions from Smart, Self-Supervised Infrastructure. In *IEEE ITSC*.

Sun, J.; et al. 2023d. NeRF-Loc: Transformer-Based Object Localization Within Neural Radiance Fields. *RA-L*, 8(8): 5244–5250.

Sun, J.; et al. 2023e. NeRF-Loc: Transformer-based object localization within neural radiance fields. *RA-L*, 5244–5250.

Sun, J.; et al. 2024a. Localization and recognition of human action in 3D using transformers. *Nature Communications Engineering*, 3(1): 125.

Sun, J.; et al. 2024b. Open x-embodiment: Robotic learning datasets and rt-x models. *ICRA*.

Sun, J.; et al. 2025a. Hierarchical hybrid learning for long-horizon contact-rich robotic assembly. *CoRL*.

Sun, J.; et al. 2025b. A Survey of Reasoning with Foundation Models: Concepts, Methodologies, and Outlook. *ACM Comput. Surv.*, 57(11).

Wang, C.; Fan, L.; Sun, J.; Zhang, R.; Fei-Fei, L.; Xu, D.; Zhu, Y.; and Anandkumar, A. 2023. Mimicplay: Long-horizon imitation learning by watching human play. *CoRL*.

Zhang, W.; et al. 2025. SynapseRoute: An Auto-Route Switching Framework on Dual-State Large Language Model. *arXiv preprint arXiv:2507.02822*.

Zheng, C.; Gao, Y.; Shi, H.; Xiong, J.; Sun, J.; et al. 2025. Dape v2: Process attention score as feature map for length extrapolation. *ACL*.

Zheng, C.; Sun, J.; et al. 2025a. SAS: Simulated Attention Score. *NeurIPS*.

Zheng, C.; Sun, J.; et al. 2025b. Understanding the mixture-of-experts with nadaraya-watson kernel. *arXiv preprint arXiv:2509.25913*.

Zheng, C.; Wang, H.; Xie, E.; Liu, Z.; Sun, J.; et al. 2024. Lyra: Orchestrating dual correction in automated theorem proving. *TMLR*.