Embodying Human-Like Modes of Balance Control Through Human-In-the-Loop Dyadic Learning

Sheikh Mannan[1], Vivekanand Pandey Vimal[2], Paul DiZio[2], Nikhil Krishnaswamy[1]

[1]Colorado State University, Fort Collins, CO USA
[2]Brandeis University, Waltham, MA USA
{sheikh.mannan,nkrishna}@colostate.edu

Abstract

In this paper, we explore how humans and AIs trained to perform a virtual inverted pendulum (VIP) balancing task converge and differ in their learning and performance strategies. We create a visual analogue of disoriented IP balancing, as may be experienced by pilots suffering from spatial disorientation, and train AI models on data from human subjects performing a real-world disoriented balancing task. We then place the trained AI models in a dyadic human-in-the-loop (HITL) training setting. Episodes in which human subjects disagreed with AI actions were logged and used to fine-tune the AI model. Human subjects then performed the task while being given guidance from pretrained and dyadically fine-tuned versions of an AI model. We examine the effects of HITL training on AI performance, AI guidance on human performance, and the behavior patterns of human subjects and AI models during task performance. We find that in many cases, HITL training improves AI performance, AI guidance improves human performance, and after dyadic training the two converge on similar behavior patterns.

Introduction

In domains like piloting, spaceflight, and even driving, maintaining spatial awareness and orientation through visual, somatosensory, and vestibular signals is critical for humans. AI systems can use numerical signals to maintain vehicle position in space, and could in principle track the human and vehicle's positioning in the relevant orientational plane(s), detect if there is a risk of losing control (Daiker et al. 2018; Zgonnikova, Zgonnikov, and Kanemoto 2016; Wang et al. 2022), and even alert the operator to make corrective maneuvers. However, due to differences in training method and data (e.g., through exposure to environmental physics or sensorimotor data from humans in the task), AIs may learn to perform the task in very different ways from humans due to differing *embodiment* of the relationship between actions and the problem space.

Fig. 1 shows a relevant example in our use case: a 30-sec. sample of a human balancing a multi-axis rotation system (MARS) device, with the angular position shown in black and the subject's joystick deflections shown in red, along with the deflections predicted by a DDPG model (blue) and an LSTM trained over actual human motions from other MARS trials (green). While both models learn to perform the task, the model trained on human data *embodies* the problem space similarly to a human, making it a superior predictor of novel humans' actions while solving the same problem.

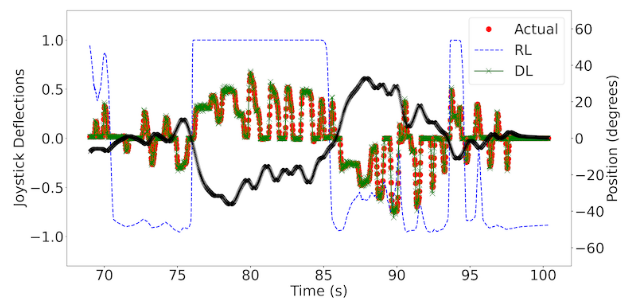


Figure 1: Actions in an IP balancing task (described below) predicted by a reinforcement learning (RL) model (blue) and a deep learning (DL) model trained over human data (green) compared to an actual 30-sec. participant trial sample (participant actions in red and angular position in black).

This has significant implications for human-AI collaboration in disorienting scenarios. We present an experimental protocol that allows us to quantify such differences, and a dyadic human-in-the-loop (HITL) training paradigm through which we train AIs to better align their decisions with humans, relative to AIs embodying their own ideal characteristics (e.g. zero reaction-time delay, high stiffness), while still maintaining proficiency.

Task Background

The *multi-axis rotation system* (MARS) paradigm is a documented, realistic simulation of vehicle control in helicopter hovering and spaceflight (Panic et al. 2017; Vimal, DiZio, and Lackner 2019), in which subjects self-balance while seated in a device programmed with inverted pendulum (IP) dynamics.

In the MARS *Upright Roll* paradigm, blindfolded subjects use a joystick to balance themselves upright in a vertical roll plane, where naturalistic gravitational reference cues (from

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Figure 2: MARS device in the supine roll condition.

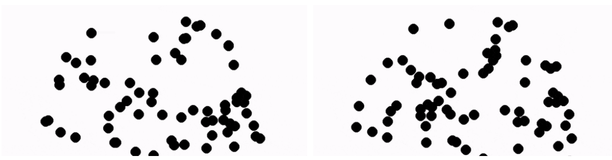


Figure 3: Two consecutive frames of the VIP 50% coherent RDK display.

otoliths) and angular rate cues (from semi-circular canals) are available. The MARS *Supine Roll* self-balancing task (Fig. 2) is a disorienting spaceflight analogue because it denies subjects task-relevant gravitational position cues by placing the body perpendicular to the gravitational vertical where subjects are no longer tilting relative to it.

The *visual inverted pendulum* (VIP) is a high-throughput alternative for studying manual object balancing. In the standard VIP paradigm, analogous to the MARS Upright Roll, subjects balance a visually-simulated circular array of dots (random dot kinematogram, RDK) which rolls in the plane of the display screen with 100% coherence in every frame, providing global configural positional information and low-level motion cues. The disorienting visual analogue is the VIP 50% coherent condition, where alternating halves of the RDK dots displace coherently across two frames while the other half jump randomly, eliminating configural displacement cues relative to the upright direction of balance (DOB) while providing low-level retinal motion cues (see Fig. 3).

Both MARS and VIP can be observed and actuated by an AI and by a human using a joystick or keyboard. Both paradigms can be configured in challenging but non-disorienting modes providing standard sensory information or difficult and disorienting modes providing degraded information. The VIP and MARS tasks possess similar underlying physical models of instability, governed by $\ddot{\theta} = k_P sin\theta$ (θ is degrees deviation from the DOB, pendulum constant $k_P = 600°/s^2$), with an RK4 integrator. In both standard paradigms, human subjects initially show angular excursions exceeding programmed "crash" boundaries ($\pm 60°$) which vanish with practice as the subject learns to maintain angular position around the upright DOB at low velocity. In the disorienting modes of both the MARS and VIP tasks, performance deteriorates in parallel ways, with positionally-drifting oscillations leading to frequent "crashes" and minimal amounts and rates of learning.

Data and Initial Training

Data presented in Vimal et al. (2020) and Wang et al. (2022) consists of 34 healthy adult subjects performing the MARS Supine Roll (disorienting) task. Each subject experienced two experimental sessions on consecutive days, each consisting of 20 100-sec. trials where, while blindfolded, they attempted to balance themselves with minimal oscillations. The data contains angular positions and velocities, and joystick deflections for each trial at a sampling rate of 50 Hz. Vimal et al. (2020) clustered participant performance into *Good*, *Medium*, and *Bad* groups based on proficiency characteristics such as number of crashes and tendency to destabilize and oscillate.

We used data from the *Good* group to train a long short-term memory (LSTM) and data from the *Bad* group to train a multi-layer perceptron (MLP) neural network model that predicted future joystick actions (direction and magnitude of joystick deflections) using sliding window approaches where inputs consisted of past angular positions, velocities, and deflections. The models exhibited performance in the VIP task that closely approximated the performance of *Good* and *Bad* humans, respectively. This resemblance was determined according to 3 variables that were strongly correlated with crash frequency in the human subjects (root mean squared velocity, velocity stabilogram diffusion function diffusion coefficient, and velocity total power). We also trained a third model on data from human subjects of all proficiencies (the *All Proficiency* model).

Human-In-the-Loop (HITL) Training

Each AI model was then placed in co-performance with human subjects in an instance of the VIP task. This setup allowed us to test both how an AI that was proficient at the task would cue the subjects to make corrective deflections if the subjects were in danger of crashing, and how a suboptimal AI performer could be made to improve through learning from interaction with a human.

Nine subjects were recruited. Each subject performed 3 30-second trials of the VIP at 50% coherence to establish baseline solo performance. The joystick control mode available to subjects followed DiZio et al. (2023); the subject supplied only direction information and the input to the simulated pendulum was always of full magnitude. Subjects engaged in two additional sessions with AIs:

1) Dyadic human-in-the-loop **(HITL) AI training**: Subjects performed the task again with VIP motion controlled by each of the three AI models in turn, while human actions served as potential corrections to AI-driven balancing. Episodes (consisting of the input window and predicted action) where the direction of AI-predicted deflection conflicted with the direction of human deflection were stored. These disagreement samples were combined with the original training data by flipping the direction of predicted deflections in the corresponding training samples. The AI model was then fine-tuned on the updated data. **2**) **AI guidance** of human performance: Control of the VIP was then swapped again, with the human performing the VIP task while receiving visual cues according to the predictions of the *Good* AI. Cues were rendered as red arrows in the top left/right of the window for the frame in which the AI predicted the human was in danger of crashing. Subjects performed 3 30-second trials each with the AI as retrained after Part 1 and with the original pretrained AI.

HITL training assesses change to AI performance after retraining it with human corrections. **AI guidance** assesses human-AI co-performance with the pretrained and dyadically fine-tuned *Good* AI model to see if human performance improves with either version providing cues. Performance metrics include the number of crashes, percentage of destabilizing actions (actions that accelerate the IP away from the DOB), average angular distance from the DOB, average velocity magnitude, standard deviation of angular position and velocity (indicating level of oscillation), and root mean square velocity. Distance/theta is in degrees. Velocity metrics are in degrees/second. Lower values mean better performance and more time spent around the DOB at lower velocity.

Results

Fig. 4 shows aggregate results from both experiments. Fig. 4 (right) shows results averaged over all nine human participants.

HITL training Preliminary results show that the *Good* AI displays minimal improvement after HITL—and even sometimes degradation—in important metrics like number of crashes and destabilizing deflections. The *Bad* AI shows little if any change across the board ($<$10% difference in all metrics whose values declined). The *All proficiency* model, which was trained on data from humans of multiple skill levels, shows substantial performance improvements on multiple metrics, especially those reflecting oscillation. The *All proficiency* model baseline metric values are also lower, indicating a more proficient starting point (i.e., this model never crashed when performing the task alone, but HITL training helped it maintain lower velocity).

AI guidance When placed in co-performance with human subjects to provide guidance, the pretrained *Good* AI model helped participants reduce destabilizing actions, crashes, and velocity/oscillation. After fine-tuning with HITL, the updated model helped further reduce most of these metrics, and aligned average human performance very closely with the solo performance of the *Good* model. However, the level of human improvement with the fine-tuned model over the pretrained model was narrower than the improvement over baseline using the pretrained model.

AI vs. Human Behavior Patterns Figure 5a shows VIP angular velocity vs. angular position for two representative participants in the preliminary trials. Red dots represent destabilizing deflections while blue dots represent “anticipatory” deflections (where VIP is tilted away from the balance point, and subject action and VIP velocity go in opposite directions—usually done to slow the VIP down when velocity is perceived as being too high). We can see that the participants display a common pattern of behavior where they oscillate around a center. The level of participant improvement between the first and the last trial also appears to be qualitatively reflected in the co-performance scenario with the AI before and after HITL. For example, in Fig. 5a subject 007’s increased destabilizing deflections are reflected in behavior of their assistant during HITL training assistant their HITL trained with, but assistance still keeps them from crashing. Subject 011 is initially unable to maintain control with AI assistance, but after HITL training, the AI assistant provides cues that keeps them closely balanced around the center. Figure 5b displays similar velocity-position scatter plots for AI agents alone, namely the good and bad pilots and an MLP-based assistant. For each model, we display the version of the AI pretrained on the original MARS dataset, the dyadic HITL training with each participant, and the final performance after updating the AI model with the data from the HITL trials. For the AI models, we see similar phase oscillations in the velocity-position scatter plots, particularly in the how the behavior of the bad and good pilot AIs during HITL training with subjects 007 and 011 reflect those subjects’ solo performance behaviors in the first and last trials, respectively.

Conclusion

If an AI is already capable of maintaining balance itself, why not have the AI override pilot inputs and take control of a vehicle directly if it detects an imminent loss of control? In a real-life piloting scenario, there may be external factors that still require human value judgments, such as engagement with other vehicles. This necessitates that a human remains in control, but also that the human be assured that the signals received from the AI assistant are informed by their own modes of balance control, to better inform just-in-time decision making.

One compelling way of ensuring this is to have the AI learn to do the task not just proficiently, but *in the manner of a proficient human*. Our results indicate the potential for dyadic HITL training and AI provided guidance to respectively improve AI and human performance in disoriented balancing, provided the underlying model is already proficient in the task (compare improvements shown by the *Good* model after HITL to the *Bad* model), and has been exposed to a wide variety of human behaviors, even suboptimal ones. In this, the *All Proficiency* model and highly-performing RL models are objects of ongoing study.

A future study can also investigate transfer to more complicated conditions, like orientation in multiple roll planes or flight simulators. There also remains the question of *how* to deliver an AI assistant’s cues to a human pilot. Possibilities include aurally rendered tones or visual indicators on the screen to indicate the direction and magnitude of the corrective action, or linguistic instructions (for instance, Mannan and Krishnaswamy (2022) present evidence toward the utility of language understanding in task performance). Precisely when and how to deliver corrective information is another avenue of future study.

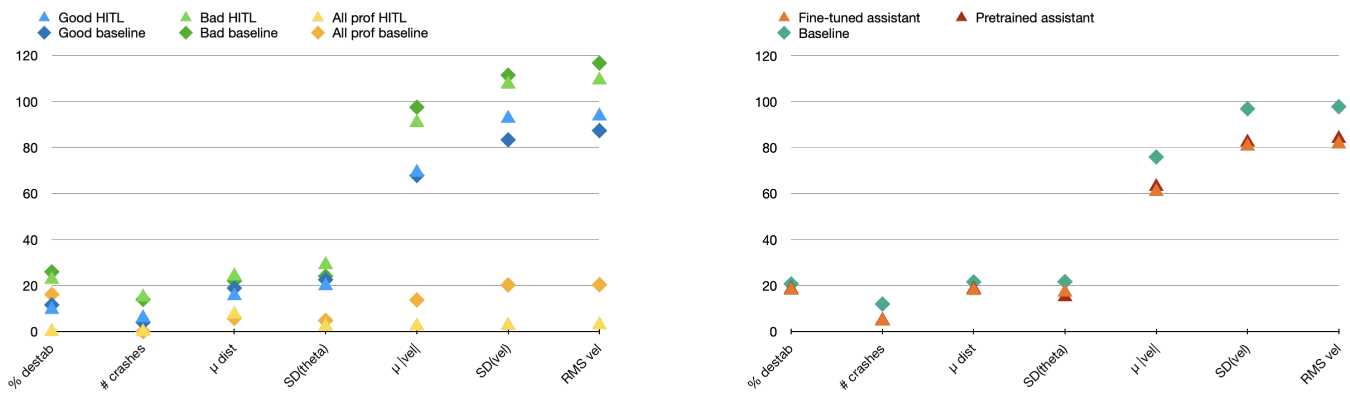
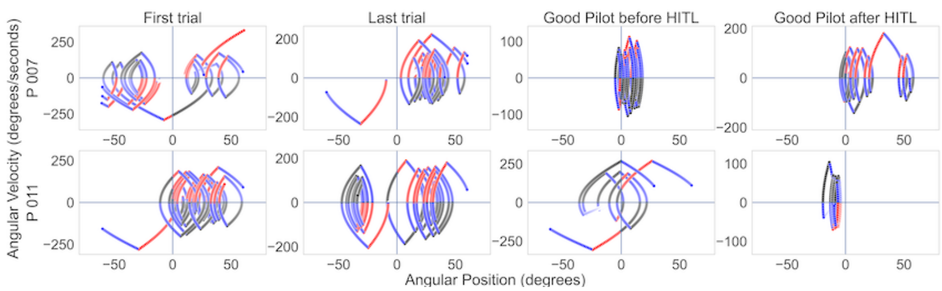
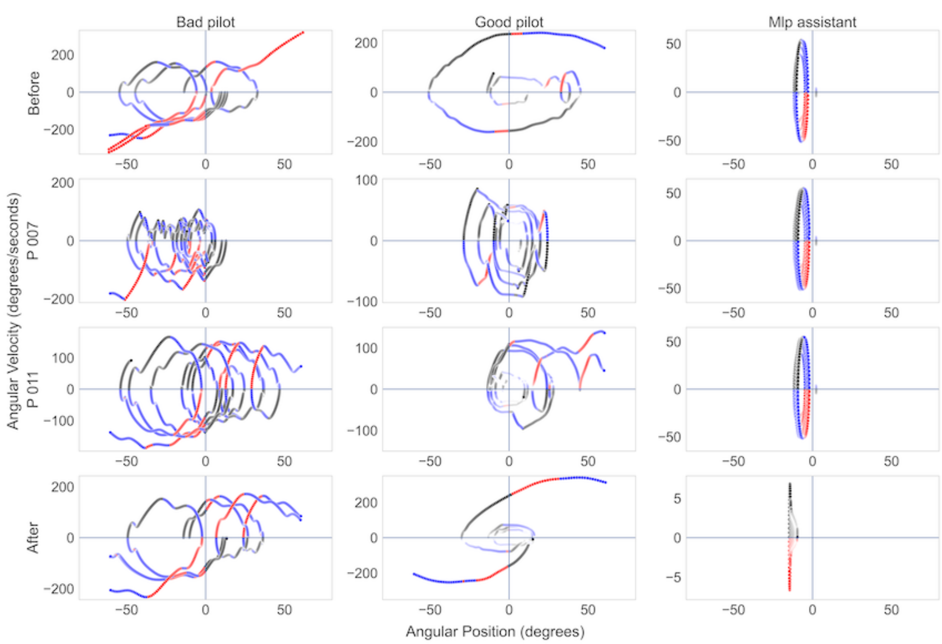


Figure 4: Aggregate results of HITL Training (left) and AI Guidance (right).



(a) Angular velocity vs. angular position for sample participants' first and last trials, and while receiving assistance from an AI before and after the AI was fine-tuned with HITL training.



(b) Angular velocity vs. angular position for each model before, during, and after HITL trials with representative participants.

Figure 5: Velocity-position scatter plots. Red dots represent destabilizing deflections while blue dots represent "anticipatory" deflections.

References

Daiker, R.; Ellis, K.; Mathan, S.; and Redmond, W. 2018. Use of real-time, predictive human modeling for spatial disorientation detection and mitigation. In *Proceedings of the Modsim World 2018 Conference*.

DiZio, P.; Krishnaswamy, N.; Mannan, S.; and Hansen, P. 2023. Manual balancing of a visual inverted pendulum by quantized versus proportional joystick commands. In *Neuroscience*.

Mannan, S.; and Krishnaswamy, N. 2022. Where Am I and Where Should I Go? Grounding Positional and Directional Labels in a Disoriented Human Balancing Task. In *Proceedings of the 2022 CLASP Conference on (Dis)embodiment*, 70–79. Gothenburg, Sweden: Association for Computational Linguistics.

Panic, A. S.; Panic, H.; DiZio, P.; and Lackner, J. R. 2017. Gravitational and somatosensory influences on control and perception of roll balance. *Aerospace medicine and human performance*, 88(11): 993–999.

Vimal, V. P.; DiZio, P.; and Lackner, J. R. 2019. Learning and long-term retention of dynamic self-stabilization skills. *Experimental brain research*, 237: 2775–2787.

Vimal, V. P.; Zheng, H.; Hong, P.; Fakharzadeh, L. N.; Lackner, J. R.; and DiZio, P. 2020. Characterizing individual differences in a dynamic stabilization task using machine learning. *Aerospace medicine and human performance*, 91(6): 479–488.

Wang, Y.; Tang, J.; Vimal, V. P.; Lackner, J. R.; DiZio, P.; and Hong, P. 2022. Crash prediction using deep learning in a disorienting spaceflight analog balancing task. *Frontiers in physiology*, 13: 51.

Zgonnikova, I.; Zgonnikov, A.; and Kanemoto, S. 2016. Stick must fall: Using machine learning to predict human error in virtual balancing task. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 173–177. IEEE.