From Assistance to Skill Transfer in Tightly Coupled Human–AI Control

Wei Wang[1], Enlin Gu[1], Antonio Loquercio[1], Haimin Hu[1,2], and Rahul Mangharam[1]

Abstract— AI copilots can substantially boost human performance through shared control, but excessive assistance can induce over-reliance and skill atrophy. This paper studies how an embodied AI agent can act as a coach that accelerates human motor-skill development. We argue that effective coaching requires strategic scaffolding and stepping back that are aligned with the learner's capability, allowing productive failures that drive learning. We formalize the interactive AI coaching process as a non-cooperative dynamic game in which the learner optimizes task performance while the coach targets the learner's independent competence. Building on this formalism, we develop a reinforcement learning framework combining adaptive shared control with probabilistic models of the coach's causal influence on skill evolution. A user study ($N = 33$) on first-person-view drone racing shows significant gains in human learning outcomes over state-of-the-art AI coaching baselines. We discuss what this objective implies for control transfer in physically coupled teams.

I. Introduction

Mastering a motor skill, from racing a car at the limit, to piloting a drone through tight gates, requires deliberate practice at the boundary of one's current ability. Human coaches accelerate this process not by performing the task for the learner, but by orchestrating the right challenges at the right moments [1], [2]. Can an embodied AI agent assume this role?

Recent advances in robot learning have produced embodied AI agents that can outperform world champions [3], [4]. A natural use of such expert AI is as a shared-control copilot: a system that fuses the human's commands with the AI's corrective inputs to keep the team performing well despite human error [5], [6], [7], [8], [9]. This is the dominant paradigm in modern human–AI collaboration, where AI acts as a "guardian angel" that prevents failures and smooths execution. However, helping humans succeed now is fundamentally different from helping them succeed alone in the future. Persistent AI assistance can induce over-reliance and skill stagnation in human users [10], [11], depriving them of the productive failures [1] that drive skill development and retention. Conversely, a coach that withdraws assistance indiscriminately exposes the learner to failures that are uninformative or frustrating, neither of which benefits human skill learning.

Our key insight is that an effective AI coach must neither overprotect nor disengage from the learner; it should strategically step back when doing so turns mistakes into catalysts of long-term independent competence. We formalize this through the Human–AI Coaching Game, a non-cooperative dynamic game in which a learner optimizes task performance while a coach targets the learner's long-term independent competence. We develop a reinforcement learning framework combining an adaptive, closed-loop shared control rule, a probabilistic automaton that captures the coach's causal influence on learner skill, and a tractable skill-improvement reward for scalable coaching policy training. We deploy our AI coach in a user study ($N = 33$) of first-person-view (FPV) drone racing with a high-fidelity simulator, demonstrating significant gains in human learning over state-of-the-art AI coaches.

Our experiments study one human learner and one embodied AI coach, not a multi-robot team. The direct result is therefore about heterogeneous shared control. We use it to motivate a narrower question for physically coupled teams: whether assistance leaves the other member more capable when control must be transferred.

II. The Human–AI Coaching Game

At each moment, the coach must decide how much it scaffolds the learner toward task completion, or steps back to let them struggle, fail, and learn. Intervening aggressively risks coddling, leaving the learner over-reliant on AI; withdrawing too early yields uninformative failures and frustration. Moreover, the learner's true skill is never directly observed and must be inferred from behavior as it evolves.

We cast the AI coaching problem as a two-player partially observable stochastic game. The learner seeks to maximize task performance, while the coach optimizes a pedagogical objective, namely the learner's independent competence. These distinct incentives are critical for coaching success: the coach may modulate assistance when doing so accelerates long-term human skill development, even at the cost of short-term task performance. Intuitively, the coach's Value of Independence asks: "If the coach were to walk away right now, how well would the learner perform?"

The AI coach implements shared control by fusing the learner's action with an expert baseline via a linear blending rule:

$$a_t = \lambda_t \odot \pi^*(o_t^C) + (\mathbf{1} - \lambda_t) \odot a_t^H, \tag{1}$$

where $\lambda_t \in [0,1]^{\dim(\mathcal{A})}$ is a per-axis blending vector that modulates the assistance level, from zero (no assistance) to one (full override), and π^* is a pretrained expert

[1]University of Pennsylvania, Philadelphia, PA, USA.
[2]Johns Hopkins University, Baltimore, MD, USA.
Correspondence: wang100@engineering.upenn.edu.

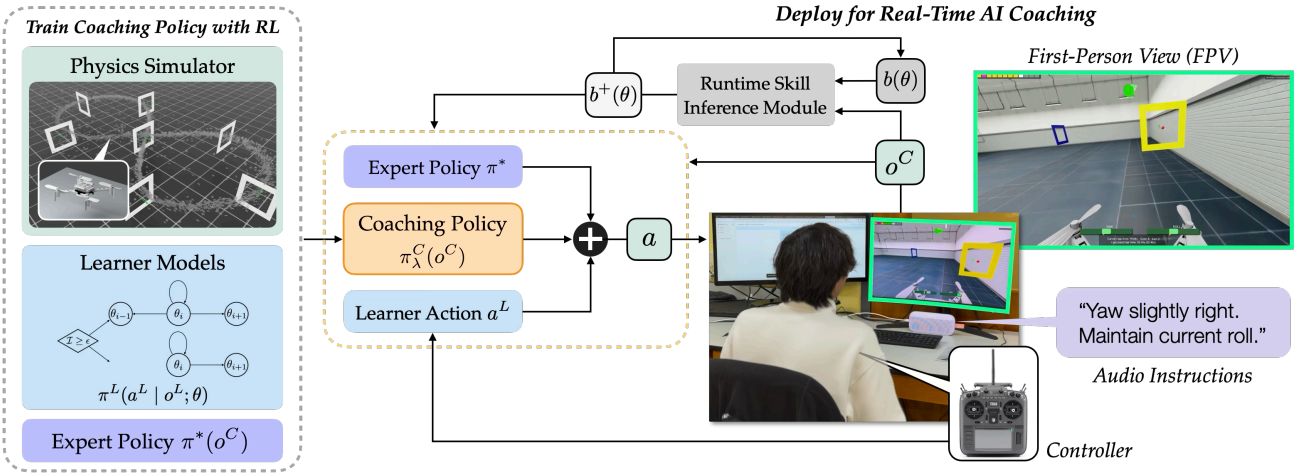


Fig. 1. Our AI coach accelerates human motor-skill development through strategic scaffolding and stepping back. The coach adaptively blends the expert action with the learner's command. At deployment, the AI maintains a belief over the learner's latent skill and conveys feedback through shared control, audio instructions, and visual cues.

policy. Anchoring on a pretrained expert baseline decouples task competence from coaching strategy: the coach only learns to modulate assistance rather than synthesize full control from scratch. Unlike fixed modulation rules, our blending factor is observation-dependent and adapts in real time to both the physical environment and the learner's action.

III. Learning to Coach

To train a coaching policy, we augment the robot's transition dynamics with two learner-side components: a skill-conditioned control policy, and a model of how the learner's latent skill θ evolves in response to coaching events such as successful or failed task attempts. Given an expert value function $Q^*(o^H, a)$, we model learners of different skill as Boltzmann noisily-rational decision-makers whose action variance shrinks with proficiency:

$$\pi^H(a^H \mid o^H; \theta) \propto \exp\big(\beta(\theta) Q^*(o^H, a^H)\big), \qquad (2)$$

where $\beta(\theta)$ increases monotonically in θ. A simulated novice samples actions broadly, while more skilled learners concentrate their actions more sharply around the expert action. Our framework is designed to be modular with respect to the learner model: it only requires a θ-parameterized policy class. The Boltzmann model is one tractable choice that is simple and interpretable.

We model the learner's skill change as probabilistic transitions triggered by success and failure events. The learner is deemed to have failed when a task-level performance index falls below a prescribed threshold. When the learner succeeds, they upskill, downskill, or remain unchanged with skill-dependent probabilities; when the learner fails, they may still upskill. The transition rates capture pedagogically meaningful trends: small wins matter most early on, skill becomes harder to lose once consolidated, and skilled learners extract more from their mistakes. To cover a diverse learner population, we randomize the transition parameters at the start of each training episode.

We train the coaching policy with proximal policy optimization, leveraging privileged access to the learner's skill level in simulation. The coach receives the tractable surrogate reward for skill improvement

$$r_t^C = \theta_t - \theta_{t-1}. \qquad (3)$$

Under the assumption that independent task value is nondecreasing in skill, a policy that induces a stochastically higher skill trajectory also obtains a higher discounted independent-value return. At deployment, skill is unobservable. The coach maintains a belief b_t over θ and updates it through runtime inference from observed learner behavior.

In drone racing, we partition the track into per-gate segments and maintain an independent belief over the learner's skill at each gate. The per-gate likelihood is a Gaussian on realized segment passage time, centered at a skill-dependent target time calibrated from a pilot study.

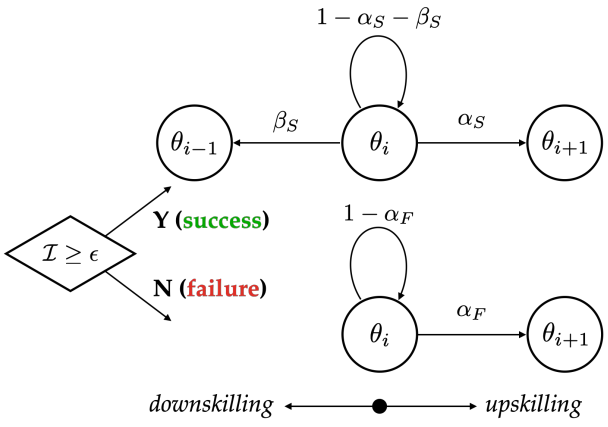


Fig. 2. Hybrid probabilistic finite-state automaton for simulating skill change triggered by success or failure events.

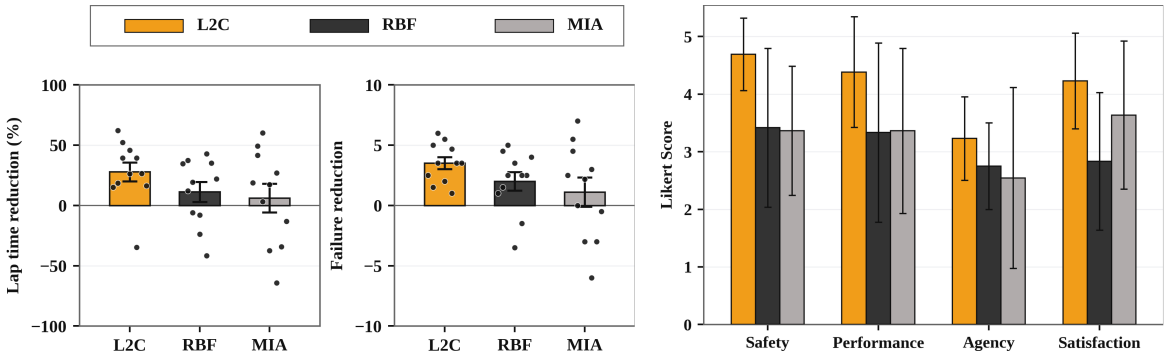


Fig. 3. After coaching, learners trained with L2C show significant reductions in lap time and failure count and report higher safety, performance, agency, and satisfaction than both baselines.

Faster passages shift belief mass toward higher skill, slower passages toward lower. Upon failure, the learner restarts one gate back and re-attempts the failed segment with targeted guidance from the coaching system.

IV. User Study

We conducted a user study with $N = 33$ participants from diverse drone-flying backgrounds in a high-fidelity FPV drone-racing simulator. Each trial has three phases: two unassisted pre-test laps, 15 coached training laps (around 40 minutes), and two unassisted post-test laps. Participants were randomly assigned to L2C, rule-based fading (RBF) [12], or minimally invasive assistance (MIA). All three methods use the same simulator, track, expert policy, language instructions, and visual cues.

For L2C, a paired-samples test yielded a mean lap-time reduction of 27.9% with $p = 0.005$ and Cohen's $d_z = -1.08$. The mean failure count decreased by 3.52 failures per lap with $p < 0.001$ and $d_z = -2.13$. RBF reduced lap time by 11.3% and failures by 2.00 per lap; MIA reduced lap time by 6.2% and failures by 1.11. L2C is the only method to reach within-subject significance on both outcomes. The learned intervention is not a global fading schedule: assistance changes with estimated skill and varies sharply around gates that are critical to the task.

V. From Human Coaching to Physical Teams

Our experiments contain one human learner and one embodied AI coach. The AI blends its expert action with the learner's command, and the learner's skill changes through repeated practice. The results therefore do not establish a multi-robot method. They show a tradeoff that also appears in physically coupled teams: stronger assistance can improve the joint trajectory while leaving the weaker member less able to act without the expert.

The Human–AI Coaching Game separates these two objectives. The learner seeks task performance under the combined action. The coach is evaluated by the learner's independent continuation value, or how well the learner would perform if assistance were removed. In a team, the same distinction is between joint return during normal coordination and a member's return when control must be transferred.

VI. Control Transfer as an Evaluation

A first extension can keep the asymmetric setting used in this paper: one frozen expert and one adapting partner. Each episode has a collaboration phase, an independent probe with assistance disabled, and a transfer phase in which the adapting partner takes a role previously handled by the expert. Every method should use matched initial states and probe times.

The comparison should include persistent expert assistance, minimally invasive assistance, a fixed fading schedule, and capability-conditioned coaching. Joint return should be reported with independent probe return, transfer completion, recovery time, and cumulative assistance, avoiding gains caused by an unusably difficult collaboration phase.

The unassisted laps in our study instantiate this measurement by removing the expert action. A multi-robot experiment must additionally show that an artificial partner improves through interaction and enables faster or safer role transfer.

VII. Capability and Credit Assignment

When only one partner adapts, its independent probes provide a direct measure of capability change. When both partners learn, an improved team trajectory is harder to interpret. It may result from one policy learning, from better coordination, or from the stronger member silently compensating for the other.

For simultaneous adaptation, the observation model should condition on each member's contribution. In L2C, the human action, expert action, blending vector, and executed action are all logged. Related signals in a robot team may include force contribution, subtask ownership, or a counterfactual rollout with one policy removed. Without such signals, a successful coupled trajectory may be credited to the member being evaluated even when most control came from the expert policy.

Capability can also affect task allocation. A coordinator can assign safety-critical work to the stronger member and use recoverable subtasks as practice for

the weaker member. The benefit should appear in later independent probes and an expert-unavailability test, not only in normal joint performance.

VIII. Safety Boundary

Stepping back does not mean allowing arbitrary physical failure. Our drone-racing study is conducted in simulation, where a collision triggers a reset. A real physically coupled system needs a safety mechanism that enforces collision, force, or workspace constraints. The coaching decision acts only within the assistance levels that this safety mechanism allows.

Independent probes must begin in recoverable states under the same safety envelope. Reports should distinguish requested from executed assistance and pair shield activations, recovery time, and task failure with capability gain.

This layered design matches the original method. The expert provides a corrective reference, while the coach learns only how strongly to blend it with the partner's action. A safety filter can override the blend without changing the independent-competence objective.

IX. Scope of the Claim

The current evidence is about human motor-skill development in a high-fidelity simulator. It does not show that the same probabilistic skill model describes robot learning, nor that the learned human-coaching policy transfers directly to a robot team. The proposed extension requires a new capability model and an experiment that tests adaptation by an artificial partner.

What transfers is the evaluation question. Two team policies may have the same joint success because one expert continually compensates for another member. Removing or reassigning the expert distinguishes permanent compensation from capability that remains after collaboration. For long-lived physical teams, that retained capability is a useful complement to immediate task return.

This matters when the strongest agent becomes unavailable or must change roles. A transfer benchmark should report the weakest member's result and repeat the test after role reassignment, since high mean performance can conceal a remaining single point of failure.

Persistent shared-control assistance can improve task performance while inducing over-reliance. Our method instead closes the loop on evolving skill and physical context. This distinction motivates control-transfer evaluation: independent value is measured under a fixed starting-state distribution of unassisted trials, and the same distribution should be used when control is transferred between team members.

X. Limitations and Outlook

Although our probabilistic skill model appeared sufficient for drone racing, the simulated learners may not capture the full variability of real human learning in more complex motor-skill training settings. Future work could explore data-driven alternatives such as behavior cloning, world models, or mixtures of policies. Our current system learns only the assistance modulation of physical control, while verbal and visual cues are predefined and triggered by fixed rules. Future work could jointly optimize a multimodal coaching policy with a foundation-model backbone.

XI. Conclusion

As AI systems grow more capable, the question is no longer whether they can replace human skill, but how they can help build it. We have argued, and shown empirically, that optimizing for independent human competence gives both a principled definition of human-empowering AI and a practical recipe for building one. In a comprehensive drone-racing user study, our game-informed RL coach measurably accelerates the development of human skill.

References

[1] M. Kapur, "Productive failure," Cognition and instruction, vol. 26, no. 3, pp. 379–424, 2008.

[2] J. Metcalfe, "Learning from errors," Annual review of psychology, vol. 68, no. 1, pp. 465–489, 2017.

[3] P. R. Wurman, S. Barrett, K. Kawamoto, J. MacGlashan, K. Subramanian, T. J. Walsh, R. Capobianco, A. Devlic, F. Eckert, F. Fuchs, et al., "Outracing champion Gran Turismo drivers with deep reinforcement learning," Nature, vol. 602, no. 7896, pp. 223–228, 2022.

[4] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, "Champion-level drone racing using deep reinforcement learning," Nature, vol. 620, no. 7976, pp. 982–987, 2023.

[5] S. Reddy, A. D. Dragan, and S. Levine, "Shared autonomy via deep reinforcement learning," in Proc. Robotics:ScienceandSystems, 2018.

[6] J. DeCastro, A. Silva, D. Gopinath, E. Sumner, T. M. Balch, L. Dees, and G. Rosman, "Dreaming to assist: Learning to align with human objectives for shared control in high-speed racing," in Conf. RobotLearning, 2024. [Online]. Available: https://proceedings.mlr.press/v270/decastro25a.html

[7] M. Srivastava, R. Iranmanesh, Y. Cui, D. Gopinath, E. S. Sumner, A. Silva, L. Dees, G. Rosman, and D. Sadigh, "Shared autonomy for proximal teaching," in 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2025, pp. 232–241.

[8] D. D. Oh, J. Lidard, H. Hu, H. Sinhmar, E. Lazarski, D. Gopinath, E. S. Sumner, J. A. DeCastro, G. Rosman, N. E. Leonard, et al., "Safety with Agency: Human-Centered Safety Filter with Application to AI-Assisted Motorsports," Proc. Robotics:ScienceandSystems, 2025.

[9] S. Sha, Y. Wang, B. Huang, A. Loquercio, and Y. Li, "Efficient and reliable teleoperation through real-to-sim-to-real shared autonomy," arXiv preprint arXiv:2603.17016, 2026.

[10] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI can harm learning," The Wharton School Research Paper, 2024.

[11] B. N. Macnamara, I. Berber, M. C. Çavuşoğlu, E. A. Krupinski, N. Nallapareddy, N. E. Nelson, P. J. Smith, A. L. Wilson-Delfosse, and S. Ray, "Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness?" Cognitive Research: Principles and Implications, vol. 9, no. 1, p. 46, 2024.

[12] C. Shen, S. Yu, Y. Weng, H. Ma, C. Li, H. Yasuda, J. Dallas, M. Thompson, J. Subosits, and T. Ersal, "Cyber racing coach: A haptic shared control framework for teaching advanced driving skills," arXiv preprint arXiv:2509.20653, 2025.