REPRESENTATION OR OBJECTIVE? SYNERGY ACTION SPACES FOR HUMAN-LIKE DEXTEROUS CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning dexterous motor control over redundant musculoskeletal systems promises biologically grounded behavior. Still, reinforcement learning tends to discover non-human, high-effort activations. On a dexterous manipulation task, standard policy search learns to complete the task by de-humanizing its muscle use, roughly doubling effort and drifting off the human activation manifold as performance rises. A natural remedy is to constrain where the policy may act by confining it to a low-dimensional coordination manifold, a muscle-synergy action space. We ask whether this suffices to recover human-like dexterous control, and whether the source of the manifold matters, comparing synergies derived from human motion capture, system dynamics, and learned predictive and generative models. We find that a coordination bottleneck restores human-like muscle use in aggregate, and that different sources buy different strengths in task reliability, zero-shot transfer, or coordination, but no action-space prior recovers human-like control at the level of individual movements. Constraining where actions may live does not determine which the policy picks. What does is the objective: a behavioral imitation reward recovers per-movement human-likeness even with no coordination prior at all. We conclude that human-like dexterous motor control is a property of the objective, not the action space, and release the benchmark, synergy bases, and human-data pipeline.

1 INTRODUCTION

The human body has many more muscles than mechanical degrees of freedom, so numerous activation patterns produce the same movement, a challenge Bernstein (1967) described as selecting among redundant solutions. A long tradition in neuroscience suggests that the nervous system addresses this using a small set of muscle synergies, fixed patterns of coactivation driven by low-dimensional commands (d'Avella et al., 2003; Ting & McKay, 2007; Tresch et al., 2006), making biological control both efficient and stereotyped: individuals reach, grasp, and type with coordinated, low-effort, and recognizably human muscle activity.

Musculoskeletal simulators such as MyoSuite (Caggiano et al., 2022) let us study this computationally: the bimanual MyoHand actuates 46 joints with 78 muscle-tendon units (23 joints, 39 muscles per hand). A reinforcement learning policy trained directly over these muscles, however, does not recover human-like coordination. Since the reward measures only task success, the policy identifies a high-effort pattern that diverges from human muscle use: as it learns to type, effort approximately doubles and activations deviate from human templates. Biologically grounded actuators alone are insufficient to produce biologically plausible control.

We study a precision task, touch typing, which is dexterous, temporally structured, and individuated across fingers, and supported by extensive motion-capture data for evaluating human-likeness. We introduce a GPU-scale QWERTY typing benchmark: a floating-base MyoHand on a spring-loaded, touch-sensing keyboard, simulated with MuJoCo-Warp and brax-PPO ($\sim$13k env-steps/s at 8192 environments on an RTX 3090, $\sim 5.6\times$ the MJX-JAX backend). All experiments use the bimanual 78-muscle model.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

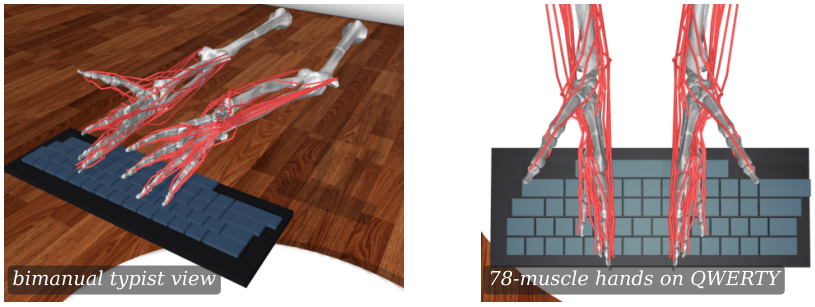

Figure 1: **The benchmark.** A floating-base bimanual MyoHand, an anatomical model with 78 Hill-type muscle-tendon units (39 per hand, red), types on a spring-loaded, touch-sensing QWERTY keyboard, simulated at scale with MuJoCo-Warp. Left: the view the policy controls; right: the key layout, with the current target key marked.

A common approach to achieving human-like control is to modify the action space, confining the policy to a low-dimensional muscle-synergy latent so that its activations must lie in that subspace. For a human-derived basis this makes aggregate metrics human-like by construction, and such synergy spaces are known to benefit RL in sample efficiency and coactivation (Berg et al., 2023). We instead ask whether they yield human-like typing at the level of *individual keystrokes*, and whether the source of the synergy matters. We compare three sources: (i) human-derived synergies (PCA on activations recovered from 28 "How We Type" participants); (ii) dynamics-derived synergies (the moment-arm SVD of joint-space empowerment, JoSE (Heald et al., 2026)); (iii) raw muscles. The human and dynamics bases overlap by only 46%, well below the 75% among individual human bases, so they encode different coordination. Two further variants are evaluated as negatives: a learned predictive (JEPA) basis and generative manifolds.

Each source is scored along four axes: human-likeness, efficiency (effort, learning speed), generalization (transfer across keyboard morphologies), and personalization (fit to individual typists). We measure human-likeness both in aggregate (pooled activation statistics over an episode) and at individual keystrokes (match to the pressed key's human template), and our primary finding concerns the gap between the two: a coordination bottleneck makes muscle use human in aggregate, and sources trade off reliability, efficiency, transfer, and coordination differently, but within *any* action space the policy still learns non-human control at individual keystrokes. What closes the gap is the objective: a behavioral imitation reward recovers per-keystroke human-likeness even with no synergy prior.

Contributions. (i) A GPU-scale musculoskeletal typing benchmark: a floating-base MyoHand on a spring-loaded, touch-sensing QWERTY keyboard, simulated in MuJoCo and at scale with MuJoCo-Warp and brax-PPO, with metrics for typing, effort, coordination, and human-likeness. (ii) A muscle-synergy action space (a 12-D latent per hand, fixed linear decode to each hand's 39 muscles) as a controlled substrate that changes the coordination source while holding policy, task, and training fixed, and a comparison of three sources, human-derived, dynamics-derived, and raw, showing that the source selects *which* human-likeness a policy attains in aggregate and how reliably it types. (iii) Three negatives showing that shaping the action space does not by itself produce human-like behavior at individual keystrokes: personalized bases, a learned (JEPA) basis, and generative manifolds each fail it. (iv) A positive result: a behavioral imitation reward makes control human-like even on raw muscles with no synergy prior, yielding the most coordinated policy in our study. Human-like control is a property of the objective, not the action space. We release the benchmark, synergy bases, and human-data pipeline.

2 Related Work

Muscle synergies in neuroscience. That the nervous system controls movement through a small set of muscle synergies has broad empirical support. d'Avella et al. (2003) decomposed frog motor patterns into a few time-varying synergies, and d'Avella & Bizzi (2005) found both shared and task-specific synergies across behaviors; Ting & McKay (2007) related postural and movement

synergies to biomechanical function, and Tresch et al. (2006) evaluated the matrix-factorization methods (PCA, ICA, NMF) used to extract them. In the hand, Santello et al. (1998) identified low-dimensional postural synergies for grasping, and Overduin et al. (2015) gave neural evidence that synergies are represented in motor cortex. Human dexterous behavior therefore lives on a low-dimensional coordination manifold, and matching it is what we mean by "human-like" control, so we fit a synergy basis on human typing data as the reference for scoring learned control.

Muscle-actuated RL. MyoSuite (Caggiano et al., 2022) and the MyoChallenge series (Caggiano et al., 2023) made over-actuated musculoskeletal control a standard RL benchmark, on OpenSim and MuJoCo muscle models (Delp et al., 2007; Todorov et al., 2012). A recurring lesson is that the action representation and exploration noise matter more than the base RL algorithm. SAR (Berg et al., 2023) fits a fixed ICA+PCA synergy basis and lets the policy act in synergy space, giving sample-efficient, human-like coactivation; DynSyn (He et al., 2024) extends this to state-dependent, dynamics-derived synergies that scale to hundreds of actuators. On exploration, DEP-RL (Schumacher et al., 2023) adds action-correlated noise and LATTICE (Chiappa et al., 2023) latent-time-correlated noise that lowers effort by 20–60%; both won MyoChallenge tracks. Curriculum shaping (Chiappa et al., 2024) and generalist muscle transformers (Chiappa et al., 2025) extend the setting, the latter reporting limited transfer of synergies across tasks, which is why we fit each basis on typing rather than borrowing one; model-based latent control such as MuscleVAE (Feng et al., 2023) targets imitation directly. We differ in two ways: we treat the choice of synergy source as the object of study, and we score against human data on a precision task rather than optimizing task return.

Empowerment and joint-space empowerment. Empowerment is an information-theoretic objective, the channel capacity between an agent's actions and its future (Klyubin et al., 2005; Salge et al., 2014), with variational estimators from Mohamed & Rezende (2015). Joint-space empowerment (JoSE) (Heald et al., 2026) specializes it to the mechanical DoF, measuring how much actions influence next-step joint velocities; under a control-affine Gaussian model of muscle-to-joint dynamics, the empowerment-optimal basis has a closed form, the right singular vectors of the whitened input matrix. JoSE thus produces a task-agnostic, biomechanics-grounded synergy basis with no human data and no reward. Our study is downstream: we take its dynamics-derived basis as one of three coordination sources and ask whether the resulting coordination transfers to a precision task and matches human muscle use, which JoSE does not optimize for. The human and dynamics bases overlap by only 46%, so the two make different predictions.

Typing and keyboard control. Prior typing benchmarks decode recorded forearm signals to keystrokes (Sivakumar et al., 2024) but treat the arm as a black box, mapping signals to text without biomechanics. The "How We Type" dataset (Feit et al., 2016) provides finger-annotated motion capture of everyday typing; we use it both for the empirical finger-to-key map and, via an inverse-dynamics retargeting pipeline, to recover per-participant muscle activations. No prior environment has a muscle-driven hand actually type; our benchmark closes that gap and scores coordination sources against human data on a keyboard task.

3 Benchmark and Method

We describe the benchmark, the synergy action space and its coordination sources, the pipeline recovering human activations from motion capture, and the metrics.

3.1 The Muscle-Driven Typing Benchmark

Body. Our agent (Fig. 1; full pipeline in Fig. 2) is the bimanual MyoSuite MyoHand (Caggiano et al., 2022): two biomechanically accurate hands, each with 23 articulated degrees of freedom actuated by 39 Hill-type muscle-tendon units under MuJoCo's native muscle dynamics (Todorov et al., 2012), 78 muscles in total. MyoSuite ships only a right hand; mirroring it across the sagittal plane gives an anatomically correct left hand, and we combine the two. Each hand sits on a floating base so it can travel over the board (6 base DoF in total).

Keyboard. The keyboard is a parametric 54-key QWERTY layout whose keys are spring-loaded prismatic sliders (4 mm travel) carrying MuJoCo `touch` sensors, so a keypress is a physical event:

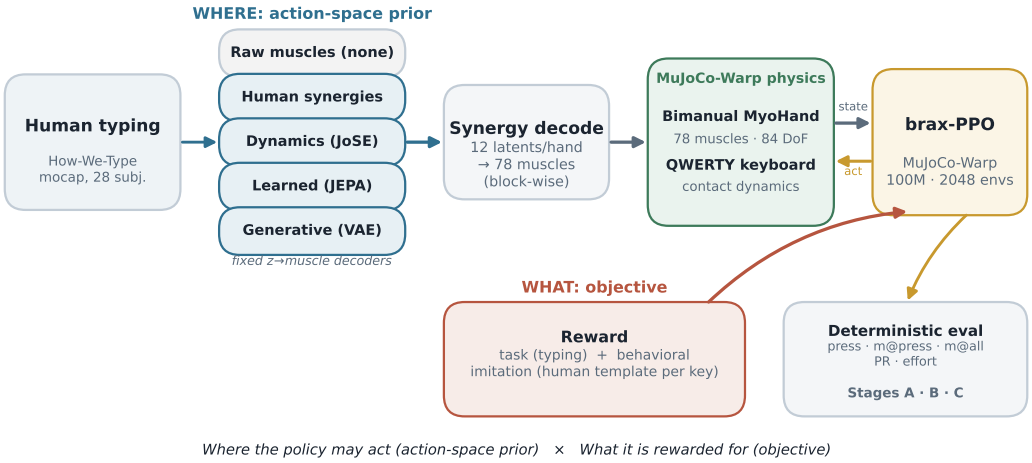

Figure 2: **Benchmark and training pipeline.** Human typing motion capture yields per-key muscle templates and the coordination bases. The policy acts either in raw muscle space or in a 12-per-hand synergy latent that decodes block-wise to 78 muscle activations driving the bimanual MyoHand on a simulated QWERTY keyboard in MuJoCo-Warp. Training is brax-PPO (100M steps, 2048 environments) under a task reward, optionally augmented with a per-key imitation reward. The study varies two axes: *where* the policy may act and *what* it is rewarded for.

the assigned fingertip depressing a key past a threshold. To keep typing plausible we pronate the forearm (palm down, fingers press by flexion), hover the base above the keys, let only fingertip geoms collide with keys, and add a virtual barrier and passive wrist stiffness at the key-top plane so the wrist and palm stay above the board.

Tasks. The benchmark defines three tasks that share the muscle-plus-base action space. In *single-key press*, a target key must be depressed by its assigned fingertip from a randomized start pose. In *touch typing*, each key has a designated finger and off-target fingers are penalized for leaving home, which yields minimal, evenly distributed multi-finger motion. In *word/sequence typing*, keys must be pressed in order, with a release-gated advance so repeated letters register as distinct keystrokes.

GPU-scale simulation. The environment runs in single-instance MuJoCo for development and at scale on GPU through MuJoCo-Warp with brax-PPO (Freeman et al., 2021; Zakka et al., 2025); the synergy decode below is a constant matrix multiply, adding no per-step GPU cost. All bimanual results train on one datacenter GPU (H100- or A100-class) for 100M steps at 2048 environments, the contact-rich 78-muscle model's memory footprint capping the environment count. As a portability check, the single-hand model reaches ∼13,375 env-steps/s at 8,192 environments on a consumer RTX 3090, about 5.6× the MJX-JAX backend's ∼2.4k.

3.2 Synergy Action Space

Instead of acting over M raw muscles (M=39 per hand, 78 across the bimanual model), the policy acts in a K-dimensional latent $z \in \mathbb{R}^K$ (K=12 per hand) that decodes to that hand's muscle activations through a fixed linear map, applied block-wise to each hand (24 synergy latents in total):

$$a \;=\; \sigma_{\text{map}}^{-1}\Big(\text{clip}\big(\bar{a} + C_K^{\top}(z \cdot s),\, 0,\, 1\big)\Big), \tag{1}$$

where $\bar{a} \in \mathbb{R}^M$ is a mean activation, $C_K \in \mathbb{R}^{K\times M}$ the frozen synergy basis, s scales the latent to the human between-key range, and σ_{map}^{-1} inverts MyoSuite's $[-1, 1] \to [0, 1]$ muscle sigmoid so the output is a native command. The policy outputs K synergy coefficients per hand (24 total) plus the 6 base servos, and z=0 reproduces the human mean. Since every commanded activation lies in

$\mathrm{span}(C_K)$, the basis fixes the coordination the policy can express. We compare three ways to obtain it.

(1) Human-derived synergies. We take per-frame muscle activations from human typing (recovered by the inverse-dynamics pipeline below) and keep the top-K principal components. On the 28-participant pool, 12 synergies capture 84% of the activation variance; this global basis is our default C_K^{hum}. Fitting per participant gives personalized bases $C_K^{(p)}$ (used in Stage C).

(2) Dynamics-derived (JoSE-style) synergies. Following joint-space empowerment (Heald et al., 2026), the empowerment-optimal directions are the right singular vectors of the whitened muscle-to-joint map. We compute this from the MyoHand moment-arm matrix $R(q) = \partial\tau/\partial a$, taking its SVD $R = U\Sigma V^\top$ and using the top-K right singular vectors as C_K^{dyn}. The basis needs no human data and no reward. At $K{=}12$ it captures 97% of the moment-arm energy, yet overlaps the human basis by only 0.46 (principal-angle subspace similarity), so the two are different priors.

(3) Predictive human-aligned (JEPA) synergies. Our third basis asks whether a learned source can combine the empowerment of the dynamics prior with the realism of the human prior. We train a joint-embedding predictive architecture (JEPA) (LeCun, 2022; Assran et al., 2023) on raw-policy rollouts, the muscle action entering only through a linear bottleneck $z_t = C_K a_t$; a predictor maps the current state and z_t to a future-state embedding, so the K directions in C_K become those that most control the future, and a VICReg term (Bardes et al., 2022) keeps z informative. A linear decoder reconstructing held-out human activations through the same bottleneck pulls the basis toward the human subspace, and a weight λ interpolates from empowering (0.49 overlap) to human-aligned (0.75) without hurting prediction. We orthonormalize; the reported arm sits at overlap 0.66, between the dynamics (0.46) and human (1.0) bases. Raw muscles (no synergy) are the no-prior baseline.

3.3 Human-Data Pipeline: Pose to Activation

Human muscle activations are not directly observable, so we recover them from motion capture: the "How We Type" dataset (Feit et al., 2016), 52 reflective markers at 240 fps with keypress and finger annotations over 28 participants (Fig. 4). Fingertip markers map to MyoHand tip sites, and because the pressing fingertip lies on the pressed key we fit one similarity transform (Umeyama, 1991) from mocap fingertips-at-press to key positions; the fit is tight (scale $\approx$0.95, median residual $\approx$5 mm at 19 mm key pitch), confirming that our hand and keyboard match the human typing scale. Retargeted joint trajectories then become activations by inverse dynamics: joint torques from MuJoCo's inverse model (`mj_inverse`), then the activations producing them under the moment-arm map via a per-frame minimum-effort quadratic program (Appendix D). This pose-to-activation pipeline yields per-participant activation sequences feeding the human synergy PCA and the per-key human template.

3.4 Metrics

Four families. *Typing*: press-frac (fraction of an episode the target key is held pressed) and word-completion rate. *Efficiency*: effort ($\mathrm{mean}|a|$) and learning speed. *Coordination*: the participation ratio $\mathrm{PR} = (\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$ of the activation covariance spectrum (lower means fewer effective muscle dimensions) and mean muscle-muscle correlation. *Human-likeness*: muscle-match $= \exp(-k \cdot \mathrm{MSE})$ to the per-key human template (1 is closest), at presses (m@press) and over the episode (m@all). We track both because a policy can match human muscle use in aggregate (m@all) yet miss the per-keystroke posture (m@press), and that gap is what separates the methods below. On top of any action space we can add a per-key template reward toward the human pattern.

4 Experiments

We evaluate along four axes: human-likeness, efficiency, generalization, and personalization. **De-risk results** are preliminary measurements that motivated the study; **Stages A–C** are the main GPU (MuJoCo-Warp) runs, evaluated deterministically at the policy mean. The arc is: characterize the synergy action space (Stage A) and its transfer across keyboards (Stage B); test three ways of shaping

action space	overlap to hum.	press ↑	m@press ↑	m@all ↑	effort ↓	PR ↓
raw muscles (no prior)	–	0.77	0.07	0.12	**0.14**	8.64
human-derived (PCA)	1.00	0.74	0.60	0.88	0.24	**5.95**
dynamics (JoSE)	0.46	**0.88**	**0.61**	**0.96**	0.26	8.09
learned (JEPA-hybrid)	0.66	0.73	0.53†	0.84	0.23	4.55
generative VAE (looser)	–	0.92	0.30	0.50	0.21	8.17
generative VAE (tighter)	–	0.92	0.27	0.45	0.20	8.91
human (reference)	1.00	0.74	0.60	0.88	0.24	5.95

Table 1: **Six action spaces, one objective.** Typing success (press-frac), human-likeness at keystrokes (m@press) and over the episode (m@all), efficiency (effort), and coordination (PR; lower is fewer effective muscle dimensions); "overlap" is principal-angle subspace similarity to the human basis. Neither the learned nor the generative action space beats the two plain synergy bases. †m@press is scored only at presses and is inflated for low-typing policies.

where the policy may act (a learned basis, generative manifolds, personalized bases, Stage C); then change *what* it is rewarded for.

4.1 Setup

All Stage runs use the bimanual (78-muscle) MyoHand on the full QWERTY keyboard with brax-PPO on MuJoCo-Warp: 100M steps, 2048 parallel environments, entropy annealed to zero, synergy decode as in Eq. (1) with $K{=}12$ per hand (24 latents plus 6 base DoF, action dimension 84), seed 0. The hyperparameters, the configuration of all 17 trained arms, and extended tables are in Appendices A–E. De-risk results used SB3-PPO (Raffin et al., 2021; Schulman et al., 2017) on CPU on a 14-key single-hand subtask; the main study is bimanual.

4.2 Stage A: Which Coordination Source, and Does Engineering the Action Space Help?

A raw-muscle PPO policy drifts *away* from human coordination as it learns: on a 14-key CPU subtask, typing success rises from 0.16 to 0.43 while mean effort roughly doubles ($0.086 \rightarrow 0.182$) and human-template match falls ($0.131 \rightarrow 0.082$), as uncorrelated per-muscle exploration walks into a high-effort, non-human corner (Appendix G). Stage A therefore trains the bimanual typist under each coordination source, raw muscles, human-derived, and dynamics-derived, on all four axes; we then ask whether *any* engineered action space pushes past the ceiling those arms reach, adding the learned predictive basis (JEPA-hybrid, Method source 3) and two generative decoders. Table 1 collects all six against the human reference. Every arm shares the policy, task, trainer, and budget, so the only difference is where the policy may act.

Finding (source matters, but only in aggregate). No basis wins every axis; the source selects *which* human-likeness the policy attains and how much typing it trades for it. Human-derived and JoSE synergies are comparably human-like at keystrokes (m@press 0.60 and 0.61 vs. 0.07 for raw, roughly $9\times$ closer to the human template); human-derived is the most coordinated but types least reliably, while JoSE types most reliably (press 0.88) and tracks human muscle use best over the episode (m@all 0.96). Raw muscles type moderately well but barely resemble human coordination.

Finding (negative: a learned human-aligned basis). Offline, the JEPA basis interpolates cleanly from JoSE-like (human overlap 0.49) to human-like (0.75) with no loss of predictive quality, but that geometry does not transfer to behavior. At overlap 0.66 it types less reliably than JoSE (press 0.73 vs. 0.88) and is no more human-like over the episode (m@all 0.84 vs. 0.96); its m@press (0.53) tracks seed noise and low press rate rather than alignment. Moving a basis's geometry toward human does not move its behavior.

Finding (negative: generative action manifolds). A VAE decoder reconstructs the human action manifold far better than the linear subspace (held-out L2 0.21–0.33 vs. PCA-12's 0.53), and per-key human actions are multimodal (2–3 modes by BIC), which a linear mean cannot capture. Yet used

policy	press ↑	m@press ↑	m@all ↑	effort ↓	PR ↓
JoSE (no reward)	0.88	**0.61**	**0.96**	0.26	8.09
JoSE + reward	**0.93**	0.58	**0.96**	0.26	5.97
raw + reward (no syn.)	0.87	0.53	0.81	**0.18**	**4.40**
human (reference)	0.74	0.60	0.88	0.24	5.95

Table 2: **A behavioral objective closes the gap.** A per-key human-imitation reward, with and without a synergy action space. On raw muscles with no synergy prior it lifts keystroke human-likeness to the level of the synergy arms, giving the most coordinated, lowest-effort policy while typing stays high.

as action spaces both decoders type well, as reliably as JoSE or better (press 0.92, 0.92 vs. 0.88), while being far less human-like at keystrokes (m@press 0.30, 0.27 vs. 0.61) and over the episode, and tighter reconstruction does not help. Constraining actions to a learned human manifold does not make the policy *choose* human coordination, because nothing rewards it. (A flow-matching decoder collapsed, so both generative arms are VAEs.)

4.3 Stage B: Keyboard Generalization

Stage B drops a reference-board policy, unchanged, onto three keyboard morphologies: tighter key pitch, stiffer springs, and an ortholinear layout (Table 6).

Finding. Every source loses some typing on the harder boards, but the ranking is largely preserved: JoSE types best on the base, stiffer-spring, and ortholinear boards (0.88, 0.82, 0.83), raw edges it only on the tighter-pitch board (0.71 vs. 0.67), and the human-derived basis types least throughout. Per-keystroke human-likeness is essentially morphology-invariant for both synergy arms (human m@press 0.48–0.60, JoSE 0.50–0.61) and stays near 0.07 for raw. The coordination a synergy basis imposes is therefore a property of the basis, not of the keyboard it trained on.

4.4 Stage C: Personalization

Because per-participant human bases are distinct (0.75 mean pairwise similarity across all $\binom{27}{2}$ pairs, 0.52 min), Stage C asks whether a policy trained in participant p's synergy space types in p's style: one policy per synergy space for four participants, each per-key profile scored against every participant's template (Table 7).

Finding (negative). Individual synergy subspaces are distinct, but that structure does not translate into individual typing behavior. A policy trained in participant p's space matches p's own template (0.50 mean) no better than it matches the other participants' (0.50): own $\approx$ other for every participant, and the four policies' profiles differ far less from one another than the human templates they derive from. The likely cause is subspace overlap: participants' bases share most of their structure, so typing recruits that shared core and leaves the individual-specific part unexercised. The global-28 template scores highest for every policy (0.57) because a 28-participant average lies closest to any single profile, an averaging artifact rather than personalization. Reproducing individual style likely needs a behavioral objective, not just p's basis.

4.5 A Behavioral Objective Closes the Gap

The three action-space methods above shape what actions are reachable but leave the policy free to pick non-human coordination, and none improves on a plain synergy basis. We finally add a behavioral objective: a per-key imitation reward $r_{\text{mm}} = \exp(-20\|a - a^{\text{hum}}_{\text{key}}\|^2)$ (weight 2.0) paying the policy for matching the human template of the key it is pressing, tested with a synergy prior (JoSE+reward) and with none at all (raw muscles+reward; Table 2).

Finding (positive). The decisive control is raw muscles+reward with no synergy prior: it lifts keystroke match from the raw baseline 0.07 to 0.53, approaching the human and JoSE synergy arms (0.60, 0.61), while typing stays high (press 0.87), and it is the most coordinated policy in the study

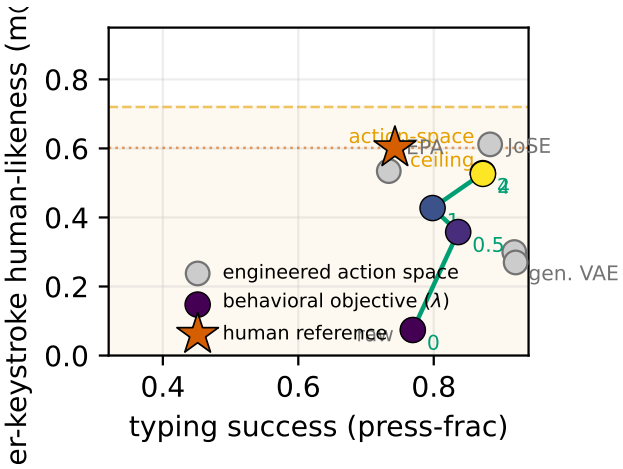

Figure 3: **Where vs. what.** Typing success against per-keystroke human-likeness. Grey points are engineered action spaces (raw, JoSE, JEPA, generative VAE), each fixing where the policy may act; the green curve sweeps the per-key imitation weight λ on raw muscles (no synergy). Raising λ lifts the raw corner toward the human reference (star) while keeping typing high, but plateaus below it.

(PR 4.40) at the lowest effort (0.18). Human-like, coordinated control is therefore a property of the objective, not of any action space, and the low-dimensional bottleneck is neither necessary nor sufficient for it. Adding the reward to the JoSE prior barely changes keystrokes, which are already near their ceiling (m@press 0.61; 0.58 with reward). No arm reaches the human template: bimanual keystroke fidelity plateaus below 0.60.

The full frontier. The reward above is applied at a single weight, one point on a continuous trade-off. Sweeping it on the raw-muscle arm from λ=0 to 4 traces the whole frontier (Fig. 3; per-λ numbers in Table 5). Per-keystroke human-likeness climbs sharply, from 0.07 to 0.53, then plateaus short of the human reference (0.60), while typing stays high throughout (press 0.77 to 0.87), so there is no strong typing-versus-human-likeness trade-off; coordination is most compact at higher weight (PR 4.40 at λ=2, below the human 5.95). No synergy bottleneck is imposed anywhere on the curve, so the coordination is produced by the objective alone.

5 Discussion

The action space constrains; the objective decides. Biologically grounded actuators do not guarantee biologically plausible control: a raw-muscle policy solves typing while staying far off the human manifold (per-keystroke match 0.07 vs. 0.60). It is tempting to fix this with the action space, since a synergy latent bounds which activations are reachable, but our three negatives show that a bounded, even human-shaped, action set does not make the trained policy human: within any such set the optimizer still walks to the typing-optimal, non-human corner, and only a reward on human-likeness moves it (Table 2). The action space constrains what is possible; the objective decides what is chosen. The human and JoSE bases overlap by only 46% yet reach comparable per-keystroke human-likeness, so the specific prior matters less than whether an objective rewards human coordination at all.

Relation to joint-space empowerment. JoSE (Heald et al., 2026) gives a normative, human-data-free account of why low-dimensional coordination emerges: it maximizes joint control authority. Ours is the downstream complement, and because that basis is not fit to human data it is a clean control for whether human-likeness requires human data. It does not: the JoSE basis matches the human basis at keystrokes (m@press 0.61 vs. 0.60) and over the episode (0.96), and types more reliably. Biomechanics buys human-like coordination essentially for free; what no action-space prior buys is per-movement fidelity beyond that shared ceiling.

Limitations. (i) All experiments use the bimanual 78-muscle benchmark, morphology transfer beyond Stage B's boards is future work, and every arm is a single seed, so trade-offs are trends rather than calibrated curves. (ii) Human activations are recovered by inverse dynamics, not measured, so they inherit the inverse-dynamics QP's assumptions (minimum-effort regularization, moment-arm accuracy): a plausible reference, not ground truth. (iii) Our synergies are linear and, for the human bases, static; state-dependent DynSyn/JoSE synergies (He et al., 2024; Heald et al., 2026) may capture posture-dependent coordination a fixed basis cannot. (iv) The finger-to-key assignment is geometric rather than standard fingering, so some keys remain harder than others.

6 Conclusion

On a GPU-scale musculoskeletal typing benchmark, standard PPO over raw muscles types while moving its muscle use off the human manifold, and a 12-per-hand synergy action space restores human-like, coordinated control in aggregate by construction (about $9\times$ closer to human templates, PR 5.95 vs. 8.64) at a cost in reliability; the source of that coordination, not merely its dimensionality, selects which human-likeness the policy attains, and it transfers zero-shot across keyboards. But three negatives, personalized, learned (JEPA), and generative bases, converge on one lesson: constraining *where* actions live never dictates *which* the policy picks. A behavioral objective does: on raw muscles with no prior, a per-key imitation reward lifts keystroke human-likeness from 0.07 to 0.53 while typing stays high, yielding the most coordinated policy we observe: human-like coordination is a property of the objective.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks that require disclosure: to *implement methods*, namely to help write and refactor parts of the simulation environment, the training and evaluation drivers, the synergy-basis fitting code, and the figure-generation scripts; and to *clean and reformat* the derived data, namely to reshape recovered activation arrays and result files into the tables reported here. Every AI-assisted change was read, run, and tested by the authors, and every number we report comes from an executed run rather than from model output.

We have not used generative AI tools for these other tasks that require disclosure: to generate synthetic data sets; to help develop theoretical models or conceptual frameworks; to propose or refine hypotheses; to design or provide feedback on our research methodology or experiments; to interpret results; or to support qualitative and thematic data analysis. The research question, the study design, the choice of coordination sources and metrics, and every interpretation of the results are the authors' own.

The remaining tasks that require disclosure are not applicable to this work: we formulate no mathematical claims and state no theorems, so generative AI tools provided no critical ingredients for proving mathematical claims and gave no assistance in the writing of proofs; and the manuscript was written in English throughout, so no translation was involved.

Additionally, we used generative AI tools for the following tasks with recommended disclosure: to create or edit software code (as above); to create or modify scientific figures, specifically the plotting scripts that render our figures from our own result files; to edit the paper to improve readability, by copy-editing and tightening prose that the authors had drafted and by reformatting the manuscript across submission templates; to format references; and to identify relevant literature, where retrieval and search tools surfaced candidate references that the authors then read and verified against the cited sources. We did not use generative AI tools to draft the scientific content of the paper, and we did not use them to write reviews.

We have reviewed all AI-assisted work and take responsibility for the final content of this work, including its text, claims, figures, code, and released artifacts.

ETHICS STATEMENT

This work raises no human-subjects concerns of its own: all agents are simulated musculoskeletal models, and no new data were collected from people. The human reference data come from the publicly released "How We Type" motion-capture corpus (Feit et al., 2016), which we use under its published terms and in de-identified form; participants appear only as numeric identifiers, and we release derived quantities (per-key muscle templates and synergy bases), not raw marker trajectories that could support re-identification. Our *personalization* experiments (Stage C) fit per-participant coordination bases, so we note explicitly that the negative result there is scientific, not a claim about individuals, and that we do not attempt to infer any attribute of a participant beyond their typing coordination. The broader risk profile of this work is that of musculoskeletal simulation generally: models of this kind are approximations of human physiology, and our human activations are recovered by inverse dynamics rather than measured, so none of our results should be read as clinically validated statements about human muscle recruitment. We highlight this in the Limitations. We see no dual-use concern specific to the benchmark.

REPRODUCIBILITY STATEMENT

We have aimed to make every reported number reproducible. The benchmark (environment, keyboard model, and tasks), the synergy bases, the human-data (pose-to-activation) pipeline, and the training and evaluation drivers are released as open source; an anonymized copy of the full code, the fitted bases, and the per-key human templates is included as supplementary material for review, and we give a permanent public link in the camera-ready version. The main text specifies the body, keyboard, tasks, action spaces, and metrics (Section 3.2–3.4); the appendix gives the details needed to rerun the study. Specifically: Appendix A lists the simulation and episode settings, the backend, the complete PPO and network hyperparameters (Table 3), the action dimensions, and the deterministic eval-at-mean protocol with the exact metric definitions; Appendix B gives the closed-form synergy decode (Eq. 2), the latent scaling, and the construction of each of the four basis sources; Appendix C

enumerates all 17 trained arms with their environment, reward weight, and basis file (Table 4), so each row of each table in the main text maps to a named run; Appendix D documents the dataset, the retargeting fit and its residuals, and the inverse-dynamics quadratic program that recovers human activations; and Appendix E reports the full per-arm sweep, per-board transfer, and personalization tables underlying the main-text summaries. All runs use a single seed (seed 0) at 100M environment steps and 2048 parallel environments; we state this and the resulting single-seed caveat in the Limitations rather than implying variance estimates we did not compute.

REFERENCES

Mahmoud Assran et al. Self-supervised learning from images with a joint-embedding predictive architecture. In *CVPR*, 2023.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR*, 2022.

Cameron Berg, Vittorio Caggiano, and Vikash Kumar. SAR: Generalization of physiological agility and dexterity via synergistic action representation. In *RSS*, 2023.

Nikolai A. Bernstein. *The Co-ordination and Regulation of Movements*. 1967.

Vittorio Caggiano, Huawei Wang, Guillaume Durandau, Massimo Sartori, and Vikash Kumar. MyoSuite: A contact-rich simulation suite for musculoskeletal motor control. In *L4DC*, 2022.

Vittorio Caggiano, Sudeep Dasari, Vikash Kumar, Huawei Wang, Guillaume Durandau, and Massimo Sartori. MyoChallenge 2022: Learning contact-rich manipulation using a musculoskeletal hand. In *NeurIPS Competition Track*, 2023.

Alberto Silvio Chiappa, Alessandro Marin Vargas, Ann Zixiang Huang, and Alexander Mathis. Latent exploration for reinforcement learning. In *NeurIPS*, 2023.

Alberto Silvio Chiappa, Pablo Tano, Nisheet Patel, Abigail Ingster, Alexandre Pouget, and Alexander Mathis. Acquiring musculoskeletal skills with curriculum-based reinforcement learning. *Neuron*, 2024.

Alberto Silvio Chiappa et al. Arnold: A generalist muscle transformer policy. arXiv:2508.18066, 2025.

Andrea d'Avella and Emilio Bizzi. Shared and specific muscle synergies in natural motor behaviors. *PNAS*, 2005.

Andrea d'Avella, Philippe Saltiel, and Emilio Bizzi. Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neurosci.*, 2003.

Scott L. Delp et al. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomed. Eng.*, 2007.

Anna Maria Feit, Daryl Weir, and Antti Oulasvirta. How we type: Movement strategies and performance in everyday typing. In *CHI*, 2016.

Yusen Feng, Xiyan Xu, and Libin Liu. MuscleVAE: Model-based controllers of muscle-actuated characters. In *SIGGRAPH Asia*, 2023.

C. Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax: A differentiable physics engine for large scale rigid body simulation. arXiv:2106.13281, 2021.

Kaibo He, Chenhui Zuo, Chengtian Ma, and Yanan Sui. DynSyn: Dynamical synergistic representation for efficient learning and control in overactuated embodied systems. In *ICML*, 2024.

James B. Heald, Vittorio Caggiano, Vikash Kumar, and Maneesh Sahani. Joint-space empowerment as a theory of dexterous motor coordination. In *ICML*, 2026.

Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *CEC*, 2005.

Yann LeCun. A path towards autonomous machine intelligence. *OpenReview*, 2022.

Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. In *NeurIPS*, 2015.

Simon A. Overduin, Andrea d'Avella, Jinsook Roh, Jose M. Carmena, and Emilio Bizzi. Representation of muscle synergies in the primate brain. *J. Neurosci.*, 2015.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *JMLR*, 2021.

Christoph Salge, Cornelius Glackin, and Daniel Polani. Empowerment: An introduction. In *Guided Self-Organization: Inception*. 2014.

Marco Santello, Martha Flanders, and John F. Soechting. Postural hand synergies for tool use. *J. Neurosci.*, 1998.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv:1707.06347, 2017.

Pierre Schumacher, Daniel Haeufle, Dieter Büchler, Syn Schmitt, and Georg Martius. DEP-RL: Embodied exploration for reinforcement learning in overactuated and musculoskeletal systems. In *ICLR*, 2023.

Viswanath Sivakumar et al. emg2qwerty: A large dataset with baselines for touch typing using surface electromyography. In *NeurIPS Datasets and Benchmarks*, 2024.

Lena H. Ting and J. Lucas McKay. Neuromechanics of muscle synergies for posture and movement. *Curr. Opin. Neurobiol.*, 2007.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *IROS*, 2012.

Matthew C. Tresch, Vincent C. K. Cheung, and Andrea d'Avella. Matrix factorization algorithms for the identification of muscle synergies. *J. Neurophysiol.*, 2006.

Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE TPAMI*, 1991.

Kevin Zakka et al. MuJoCo playground: An open-source framework for gpu-accelerated robot learning. In *RSS*, 2025.

A EXPERIMENTAL AND TRAINING DETAILS

Simulation and episode. All results use the bimanual MyoSuite MyoHand (two anatomical hands: the right hand ships with MyoSuite, the left is a sagittal mirror). Each hand has 23 articulated DoF and 39 Hill-type muscle–tendon units, so the body carries **78 muscles** under MuJoCo native muscle dynamics; each hand sits on a floating base (6 base DoF total). Control runs at $\Delta t_{\text{ctrl}} = 0.02\,\text{s}$ with a physics step $\Delta t_{\text{sim}} = 0.002\,\text{s}$ (10 substeps per control step). Episodes are $T = 100$ control steps ($= 2.0\,\text{s}$). The parametric 54-key QWERTY board uses spring-loaded prismatic key sliders (4 mm travel); a keypress is a physical event (assigned fingertip depresses the key past $0.6\times$ travel, released below $0.2\times$).

Trainer and backend. Policies are trained with **brax PPO** on the **MuJoCo-Warp** backend (`impl=warp`), wrapped through MuJoCo Playground. The synergy decode is a constant matrix multiply applied to the policy output, so it adds zero per-step GPU cost. Table 3 lists the PPO hyperparameters exactly as set in the code (env module `ppo_config`); values not set there use the brax default and are marked as such.

Hyperparameter	Value
Total environment steps	1.0×10^8 (100M)
Parallel environments (`num_envs`)	2048
Learning rate	3×10^{-4}
Optimizer	Adam (brax default)
Discounting γ	0.97
GAE λ	0.95
Entropy cost	1×10^{-3}
PPO clipping ϵ	0.3
Max gradient norm	1.0
Reward scaling	1.0
Action repeat	1
Unroll length	10
Number of minibatches	32
Updates per batch	8
Batch size	256
Observation normalization	on
Number of eval rounds	16
Seed	0
Policy MLP	$(64, 64, 64)$
Value MLP	$(64, 64, 64)$
Activation	swish (brax default)
Action distribution	tanh-Normal
Exploration std	global learned, `init_noise_std`$= 1.0$

Table 3: PPO / network hyperparameters, read from the environment module's `ppo_config` and the training drivers. The 100M step count and 2048 environments are command-line settings for the paper runs (the module default of the bimanual config is 512 envs / 50M steps, overridden per run).

Action dimensions. A raw-muscle arm acts in 78 muscle commands $+ 6$ base servos $= 84$-D. A synergy (or generative) arm acts in $K = 12$ latents per hand $+ 6$ base servos $= 30$-D; setting $z = 0$ reproduces the human mean posture.

Evaluation protocol. All reported metrics come from a **deterministic eval-at-mean** rollout (policy mean action, no exploration noise) over parallel eval environments, scoring `mjData.act` directly. Because deterministic eval on the contact-rich 78-muscle model is memory-heavy, the eval env count is kept small (`--eval_envs` $= 64$; the trainer default of 128/256 OOMs at eval); the policy parameters are pickled before eval so an eval-time OOM never loses a run. Metrics: *press-frac* (fraction of steps the target key is held pressed by the assigned finger), *effort* $=$ mean$|a|$ over the 78 muscles, participation ratio $\text{PR} = (\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$ of the activation covariance spectrum (lower $=$ fewer effective muscle dimensions), and muscle-match $= \exp(-20\,\text{MSE})$ between policy activations and the per-key human template, reported at presses (*m@press*) and over the whole episode

Arm	Environment	mmw	Basis / decoder	Type
`raw`	Bimanual Template	0	– (raw muscle)	baseline
`f05`	Bimanual Template	0.5	– (raw muscle)	obj. sweep
`f1`	Bimanual Template	1	– (raw muscle)	obj. sweep
`f2`	Bimanual Template	2	– (raw muscle)	obj. sweep
`f4`	Bimanual Template	4	– (raw muscle)	obj. sweep
`human`	Bimanual Synergy	0	`synergy_basis_right.npz`	human PCA
`humanRW`	Bimanual Synergy	2	`synergy_basis_right.npz`	human + rew.
`jose`	Bimanual Synergy	0	`jose_basis.npz`	dynamics
`joseRW`	Bimanual Synergy	2	`jose_basis.npz`	dynamics + rew.
`jepa`	Bimanual Synergy	0	`jepa_hybrid_basis.npz`	JEPA-hybrid
`jepaRW`	Bimanual Synergy	2	`jepa_hybrid_basis.npz`	JEPA + rew.
`gvloose`	Bimanual Synergy	0	`decoder_vae_loose.npz`	VAE (looser)
`gvtight`	Bimanual Synergy	0	`decoder_vae_tight.npz`	VAE (tighter)
`sb096115`	Bimanual Synergy	0	`synergy_basis_096115.npz`	per-participant
`sb270221`	Bimanual Synergy	0	`synergy_basis_270221.npz`	per-participant
`sb287951`	Bimanual Synergy	0	`synergy_basis_287951.npz`	per-participant
`sb497021`	Bimanual Synergy	0	`synergy_basis_497021.npz`	per-participant

Table 4: The 17 trained arms. mmw = `reward_config.muscle_match_weight` (the per-key behavioral-imitation reward $r_{mm} = \exp(-20\|a - a^{\text{hum}}_{\text{key}}\|^2)$). All arms: 100M steps, 2048 envs, seed 0, deterministic eval. Synergy arms use $K = 12$ latents per hand.

(*m@all*). m@press is scored only on successful presses and is therefore inflated for low-typing policies; m@all is the press-rate-robust measure and is read alongside it.

B Synergy Action Space

When the synergy action space is enabled, the policy emits a per-hand latent $z \in \mathbb{R}^K$ ($K = 12$) instead of raw muscle commands. Each hand's latent is decoded to 39 muscle activations by a **fixed, block-wise** map (the same 39-muscle basis applies to each hand of the bimanual body). With frozen basis $C_K \in \mathbb{R}^{K\times 39}$, mean activation $\bar{a} \in \mathbb{R}^{39}$, and per-latent scale $s \in \mathbb{R}^K$,

$$a = \text{clip}\big(\bar{a} + C_K^\top (z \odot s),\ \varepsilon,\ 1-\varepsilon\big), \quad \varepsilon = 10^{-3}, \tag{2}$$

$$u_{\text{musc}} = \text{clip}\Big(\tfrac{1}{2} + \tfrac{1}{5}\ \log \tfrac{a}{1-a},\ -1,\ 1\Big). \tag{3}$$

The second line is the *inverse* of the environment's native $[-1, 1] \to [0, 1]$ muscle sigmoid, so the env's own `norm_actions` map recovers exactly a; every commanded activation therefore lies in $\text{span}(C_K)$ and the basis fixes the coordination the policy can express. The **z-scale** is $s = \texttt{span} \cdot \text{std}$ of the human per-key template projected into synergy space ($\text{scores} = (a^{\text{tmpl}}_{\text{valid}} - \bar{a})\, C_K^\top$), with $\texttt{span} = 4.0$, so z near $[-1, 1]$ spans the human key-conditional range. Bases that ship a baked `mean` (human PCA) use it as $\bar{a}$; bases without one (dynamics-derived, per-participant) use the valid-key mean of the human template as $\bar{a}$ so $z = 0$ is a human muscle operating point.

Basis sources. Four sources of C_K are compared:

- **Human-derived (PCA).** PCA on per-frame human activations recovered by the pose-to-activation pipeline (App. D); 12 synergies capture 84% of activation variance. The global 28-participant fit is the default `synergy_basis_right.npz`; per-participant fits are used in Stage C.
- **Dynamics-derived (JoSE-style).** Right singular vectors of the whitened muscle→joint moment-arm map $R(q) = \partial\tau/\partial a$; the top-$K$ right singular vectors form C_K. No human data, no reward. At $K = 12$ it captures 97% of moment-arm energy yet overlaps the human basis by only 0.46 (principal-angle subspace similarity).
- **Predictive human-aligned (JEPA-hybrid).** A JEPA world-model over $[\text{proprio}, \text{qvel}]$ with a *linear* muscle→z bottleneck (the basis) + a future-state predictor + a VICReg

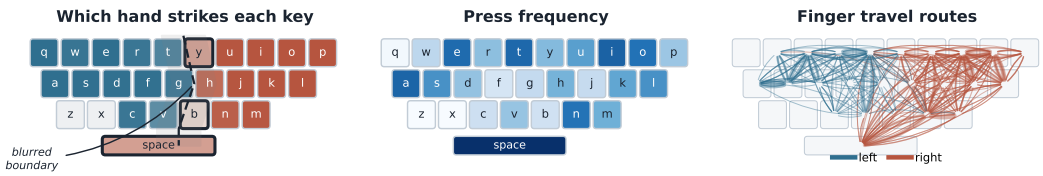

Figure 4: **Human typing on the QWERTY board (How-We-Type; 27 participants, 32.9k keystrokes).** The same 54-key layout under three views. **(A)** Which hand strikes each key, colored by the per-key left-hand fraction. The left/right division is a fuzzy band rather than a clean line: `b`, `y`, and the spacebar are struck by both hands (bold outlines), so the hand boundary is blurred (`z`, `x` absent from the data, shown grey). **(B)** Per-key press frequency; the spacebar and the home-row vowels dominate. **(C)** Per-finger travel routes (key-to-key transitions, line width proportional to frequency), colored by hand. Each hand covers its own side of the board and the index fingers roam widest; the routes overlap near the center, consistent with the blurred hand boundary in panel A.

term keeping z informative + a linear decoder autoencoding held-out human activations. A weight λ interpolates the basis from empowering (overlap 0.49) to human-aligned (0.75); the reported arm is at overlap 0.66. A learned basis is always orthonormalized (QR) before use, otherwise the env decode saturates and diverges (NaN).

- **Generative (VAE).** A nonlinear MLP decoder $g(z)$ replaces the linear basis: the latent is squashed $x = \texttt{z_scale} \cdot 3\tanh(z/3)$, passed through ReLU hidden layers and a final sigmoid to 39 activations in $[0, 1]$, then inverse-sigmoid'd as in Eq. 2. Two decoders at looser vs. tighter human-manifold reconstruction are compared.

Raw muscles (no synergy) are the no-prior baseline.

C Per-Arm Configuration

Table 4 lists the 17 trained arms with their environment, the muscle-match (behavioral-imitation) reward weight (mmw), and the basis or decoder file. Template arms act on raw muscles (84-D); Synergy arms act in the 30-D latent space of App. B.

D Human-Data Pipeline: Pose to Activation, and Dataset

Human muscle activations are not observable, so they are recovered from motion capture.

Dataset. The *How We Type* corpus (Feit et al., 2016): 52 reflective markers at 240 fps with keypress and finger annotations, **28 participants** (27 with usable key data, 32.9k keystrokes total).

Retargeting. Fingertip markers are mapped to the MyoHand tip sites by a single *similarity transform* (Umeyama, 1991) fit from mocap fingertip-at-press positions to the physical key positions. The fit is tight (scale ≈ 0.95, median residual ≈ 5 mm at a 19 mm key pitch), confirming the simulated hand and keyboard match the human typing scale.

Activation recovery. Retargeted joint trajectories are passed through the MuJoCo inverse model (`mj_inverse`) to obtain joint torques, then a **per-frame minimum-effort quadratic program** over the moment-arm map recovers muscle activations (per-hand 39-muscle order). This yields per-participant activation sequences.

Templates and bases. The recovered activations feed (i) the human synergy PCA of App. B and (ii) a per-key *human muscle template*: the mean activation at each key over presses, retained for keys with at least `template_min_count`$= 50$ human presses (others contribute no muscle-match reward). Finger-to-key assignment uses the empirical finger→key map from the dataset annotations; it is geometric rather than standard touch-type fingering, so a few keys remain harder.

E EXTENDED RESULTS

Table 5 gives the full action-space sweep: press-frac, m@press, m@all, effort, and PR for all 13 non-personalization arms. The first block is the behavioral-objective sweep on *raw* muscles (imitation weight $\lambda = 0, 0.5, 1, 2, 4$); the remaining blocks are the engineered action-space priors (human / JoSE / JEPA / VAE), each with and without the same reward. The human reference row is the recovered human template scored in-sim.

Arm (λ / prior)	press	m@press	m@all	effort	PR
raw (λ=0)	0.77	0.07	0.12	0.14	8.64
raw+rew (λ=0.5)	0.84	0.36	0.59	0.16	5.58
raw+rew (λ=1)	0.80	0.43	0.64	0.18	7.46
raw+rew (λ=2)	0.87	0.53	0.81	0.18	4.40
raw+rew (λ=4)	0.87	0.53	0.82	0.18	4.56
human	0.74	0.60	0.88	0.24	5.95
human+rew	0.76	0.64	0.96	0.25	4.74
JoSE	0.88	0.61	0.96	0.26	8.09
JoSE+rew	0.93	0.58	0.96	0.26	5.97
JEPA-hybrid	0.73	0.53	0.84	0.23	4.55
JEPA+rew	0.80	0.65	0.95	0.24	5.70
VAE (looser)	0.92	0.30	0.50	0.21	8.17
VAE (tighter)	0.92	0.27	0.45	0.20	8.91
human (reference)	0.74	0.60	0.88	0.24	5.95

Table 5: Extended per-arm metrics (final 100M, deterministic). m@press is press-rate-confounded (a policy that rarely presses looks artificially human at the few frames it does) and is read alongside m@all. Personalization arms are in Table 7.

Per-board zero-shot transfer. Table 6 drops the reference-board policy unchanged onto three variant boards (tighter pitch, stiffer springs, ortholinear layout) and reports press-frac. The ranking is largely preserved (JoSE best on base, stiffer, and ortho; raw edges it only on the tighter-pitch board; human-derived types least throughout), so coordination transfers zero-shot across keyboard hardware.

Board	raw	human	JoSE
base	0.77	0.74	0.88
tighter pitch	0.71	0.54	0.67
stiffer springs	0.67	0.64	0.82
ortholinear	0.76	0.66	0.83

Table 6: Zero-shot morphology transfer, press-frac. A reference-board policy is evaluated unchanged on each variant board.

Stage C personalization. Table 7 scores each per-participant policy's per-key profile (m@press) against its own template, the mean of three *other* participants' templates, and the global 28-participant template. A policy trained in participant p's synergy space matches p's own template no better than a stranger's (own $\approx$ other for every participant), and the global-28 template scores highest for every policy, an averaging artifact (a 28-participant mean lies closest to any single profile) rather than personalization.

Policy (trained in p's space)	own p	other p'	global-28
sb096115	0.47	0.47	0.54
sb270221	0.54	0.54	0.61
sb287951	0.53	0.53	0.60
sb497021	0.47	0.47	0.55
mean	**0.50**	**0.50**	**0.57**

Table 7: Stage C personalization (m@press). own $\approx$ other for every participant; global-28 highest throughout.

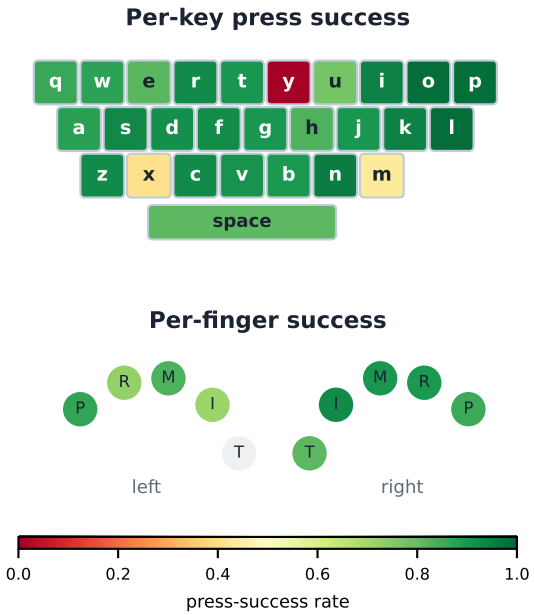

Figure 5: Per-key and per-finger success for the final behavioral model (arm f4: raw muscles with behavioral imitation reward at $\lambda = 4$), from a 512-env deterministic rollout. *Top:* per-key press-fraction over the QWERTY layout. *Bottom:* per-finger success breakdown. Success is key-specific rather than finger-specific: all fingers land in a narrow 0.72–0.93 band, while the failing keys concentrate at the ambiguous left/right-hand boundary.

F Per-key and Per-finger Success Analysis

Figure 5 breaks the model's overall press-fraction of 0.84 down by key and by finger. Most keys are pressed reliably (o, p, l, n, and i all reach 0.94–0.99, and the spacebar reaches 0.81), but success is strongly key-specific: the model fails outright on y (0.00) and struggles on x (0.40) and m (0.43). By contrast, the per-finger breakdown is nearly uniform, with every finger between 0.72 and 0.93, so there is no single weak finger driving the failures. The hardest keys instead cluster at the ambiguous left/right-hand boundary: y is exactly the blurred-boundary key that the main-paper keyboard figure shows humans strike with either hand. The model's errors therefore concentrate where the finger-to-key assignment is itself ambiguous rather than where a finger is weak.

G De-risk Measurements

A raw-muscle PPO policy drifts from human coordination as it learns. On a 14-key CPU subtask, typing success rises from 0.16 to 0.43, but mean effort roughly doubles ($0.086 \rightarrow 0.182$) and human-template match falls ($0.131 \rightarrow 0.082$): uncorrelated per-muscle exploration walks into a high-effort, non-human corner. An early CPU A/B confirmed that a 12-D human-synergy action space makes

activations human-like by construction; whether that survives at scale *as behavior* is what Stage A and the objective study test. The basis geometry framing that comparison is in Section 3.2: both bases are low-dimensional yet overlap by only 0.46, and per-participant human bases are themselves distinct (0.75 mean, 0.52 min), motivating Stage C.