# Closing the Substrate Gap in the JEPA Family

A Patent-Disclosed Multi-Actuator Closed-Loop
Testbed for Human-Centered World Models

Georg Joeri Gredig*[1]

[1]Swissrehamed GmbH, Switzerland

28 June 2026 — DRAFT v0.4.8

**Abstract**

*The machine externalizes, stabilizes and trains the human world model.*

The Joint-Embedding Predictive Architecture (JEPA) family of LeCun and collaborators has, over the past four years, matured into a dominant paradigm for self-supervised representation learning, with extensions across images, video, speech, and spatiotemporal physics. We argue that a structurally distinct construct has so far remained outside the JEPA program: a **Human-Centered World Model (HCWM)**, in which the human is not simulated but is instantiated as a physical controller node, either (a) in parallel with the machine controller in real time or (b) as a sequential stage in the controller's program sequence. We introduce the HCWM construct, anchor it in a patent disclosure (PCT/EP2024/076634) that names the human as a controller with sensorimotor state $x$, experiential setpoint $w$, and active output $y = f(x)$, and present one instantiation of the substrate that the HCWM construct requires. The hardware platform is a four-actuator EtherCAT system operating at a 100 Hz hard-real-time cycle in joint-position-velocity-torque mode; above the platform, a coordination layer realizes a configurable phase pattern, an analytic torque-feedforward stack drawn from a patent disclosed in 2009, and an externally instrumentable forward / inverse-model regime selector. The substrate's notation $x / w / y / z$ aligns term-for-term with both the JEPA predictive architecture and the active-inference formalism of Friston. We report empirical substrate-characterization results, including a survey of six configured phase patterns and a diagnostic series in which the substrate identifies its own pendulum-mass profile, attenuates the corresponding $1 \times$ rev acoustic envelope component from 84% to 54% modulation depth via patent-derived analytic feedforward, and discriminates between mechanical and feedback-loop sources of residual perturbation under manned operation. We position the substrate layer as a complement to, not a competitor of, the open JEPA research program and the existing application-layer neurorehabilitation product family, and we close with seven open research questions for the JEPA community, the hardware-platform engineering community, and the cognitive-neuroscience community.

**Keywords:** JEPA, Human-Centered World Models, Active Inference, Predictive Coding, Closed-Loop Control, Embodied AI, Substrate Licensing, Patent-Disclosed Architecture, Real-Time Systems, Joint-Position-Velocity-Torque Control.

## 1 Introduction — The JEPA Family as Theoretical Frame

The Joint-Embedding Predictive Architecture (JEPA) family, articulated by LeCun in the position paper *A Path Towards Autonomous Machine Intelligence* [14], is the most recent point on

*Corresponding author: joeri@gredig.swiss. Independent Researcher, Switzerland. Co-inventor of EP 2 364 686 (filed 2009, granted 2015) and sole inventor of PCT/EP2024/076634 (priority 27.09.2023, filed 23 September 2024 at the European Patent Office, The Hague), both held by Swissrehamed GmbH.

a research trajectory of joint-embedding methods that LeCun has been pursuing since the early 1990s. The lineage runs from the original Siamese-network construction at Bell Labs (Bromley, LeCun and collaborators, around 1993, originally developed for signature verification), through the contrastive-learning generation of the late 2010s, into the non-contrastive joint-embedding methods — in particular Barlow Twins (Zbontar, Deny, LeCun et al., 2021), VicReg, and the DINO family — and into the JEPA family proper. The present paper makes no contribution to this trajectory itself; we cite it here in the form most useful for our argument, and we acknowledge that the predictive-architecture program described in the rest of this section belongs to LeCun and his collaborators, not to the present author.

I-JEPA [2] established that joint-embedding prediction in latent space outperforms pixel-level reconstruction on static images. V-JEPA [3] carried the formulation into video, with temporal predictive coding and without generative reconstruction. By 2026 the family had branched into speech (GMM-Anchored JEPA [12]), high-level world modeling from video [15], and representation learning over spatiotemporal physical systems [17]. The trajectory — image, video, speech, physics — demonstrates that the JEPA formalism is not bound to any single modality.

In parallel, JEPA-style architectures have been extended by other groups, without LeCun's direct involvement, to a broader range of modalities. Representative recent work includes LiDAR point clouds for autonomous-driving perception [22], electroencephalography for brain-signal analysis [11], and longitudinal electronic health records framed as patient-trajectory world models [1]; further extensions to single-cell transcriptomics, surgical video, ultrasound, and multilingual text have been reported in the literature surveys of the past twelve months. The underlying claim across this growing body of work is consistent: a predictive latent representation, trained without reconstruction, captures the structure of the modality in a manner that supports downstream classification, inference, or control tasks.

**The unfilled slot.** What no JEPA-style work to date addresses is a structurally different class of modality: **closed-loop embodied actuator output**, generated under hard real-time constraints, in a physical substrate where a human acts as a parallel controller inside the same loop. All current JEPA extensions are perceptual — they ingest sensor streams. None close the loop through a deterministic actuator path with bidirectional cognitive interaction. AMI Labs (LeCun et al., founded 22 January 2026) is positioned to extend the JEPA program toward Advanced Machine Intelligence; the public technical record to date describes operations on simulated, recorded, or web-scale datasets, rather than on real-time embodied substrates.

**Position of this paper.** We argue that this empty slot is not incidental but structural. Filling it requires a substrate that is simultaneously

(i) a verified, deterministic, multi-actuator closed-loop system;

(ii) instrumented with multiple complementary observable streams — actuator torque and position, time-of-flight environmental perception, acoustic envelope, and bidirectional speech;

(iii) anchored in a formal control-theoretic notation that is structurally isomorphic to both predictive coding and active inference.

We present such a substrate in Section 5. Before describing the substrate itself, we anchor the work in the older cognitive-neuroscience tradition of predictive coding and active inference (Section 2), of which the JEPA family can be read as a recent computational instantiation. Section 4 defines the substrate gap that the present work is intended to fill. Section 6 reports empirical results from a four-actuator EtherCAT testbed that instantiates the substrate. Section 7 positions the substrate against the existing landscape of commercial neurorehabilitation systems and outlines a licensing model. Section 8 closes with seven open research questions for the JEPA community.

## 2 Predictive Coding and Active Inference — A Cognitive Neuroscience Anchor

Long before the JEPA family appeared, predictive coding and active inference had developed into a dominant computational frame within which biological perception and motor control were modeled. Within this framing [9], the brain is understood as continuously maintaining a generative model of the body's interaction with the environment, minimizing some functional of the discrepancy between sensory observation and prediction. Whether the functional is interpreted as variational free energy in the strict Friston sense, as prediction error in the looser predictive-coding sense [7], or as a mismatch term in a paired forward/inverse model architecture [19, 20], the structural commitment is the same: the system holds a forward model that *predicts* the consequences of its own action, compares prediction against measurement, and acts (or updates) so as to drive the residual toward zero.

We adopt this framing here as a working scaffold, not as a claim about neural ground truth. The free-energy principle in particular remains the subject of substantive critique within cognitive science and computational neuroscience, and the present paper takes no position on whether the brain literally implements variational inference. What we borrow is the formal structure: a forward model (the prediction), an observation (the measurement), an error term (the residual), and an action policy (the response). This minimal scaffold is sufficient for our argument, and is the smallest set of constructs that links the JEPA family back into its biological precedents.

**An adjacent practitioner lineage — modelling from exemplars.** A second lineage, structurally adjacent to predictive coding but developed independently in applied behavioural practice, is the modelling-from-exemplars tradition of neuro-linguistic programming (Bandler and Grinder, 1970s onward), in which a learner's behaviour is shaped against the explicit behaviour, physiology and inner-process descriptions of an exemplar judged competent in a target domain. The structural commitment is again the same: a reference state (the exemplar's profile), an actual state (the learner's current profile), a gap, and an action policy that closes the gap through iterated practice. We name this lineage here only to acknowledge that the construct of *learning toward a reference profile* predates both active inference and JEPA in operational practice; the substrate of the present paper is agnostic between these methodological lineages and admits any of them as a source for the reference variable $w$ of the patent disclosure. The Wolpert–Kawato proposal [20] is particularly useful as a bridge between cognitive theory and engineering practice. Their forward model maps motor commands to predicted sensory consequences; their inverse model maps desired sensory consequences to the motor commands required to produce them. In a closed-loop sensorimotor system, the forward model takes the role of an analytic or learned feedforward term in the control law, while the inverse model corresponds to the feedback corrector. The interaction between the two defines whether the system is operating in a forward-dominant ballistic regime (low-latency, low-feedback), an inverse-dominant reactive regime, or a hybrid regime in which both contribute. Section 6 reports empirical results from a substrate that exposes both regimes through explicit run-time parameters and shows that the substrate itself can be used to differentiate between mechanical sources of prediction error and feedback-loop sources.

**The JEPA family read through this lens.** The JEPA architectures are interpretable as computational instantiations of the predictive-coding scaffold: a learned encoder produces a latent representation of the present observation, a learned predictor produces a latent representation of the masked or future observation, and the training objective minimizes the discrepancy between the two in the embedding space. The non-generative character of JEPA — the absence of a pixel-level reconstruction objective — aligns more naturally with active-inference variants in which the system minimizes a measure of expected free energy than with classical reconstruction-based predictive-coding implementations.

**What the framing does not specify.** Neither predictive coding, nor active inference, nor the JEPA family specifies the *substrate* in which the predictive scaffold is realized. They specify the computational form: an encoder, a predictor, an error functional, an action policy. They are silent on whether the substrate is a biological brain, a simulation, a recorded dataset, or a real-time embodied control loop. The remainder of this paper addresses one specific instantiation of the substrate question, and argues that the choice of substrate is not neutral: a real-time multi-actuator closed-loop system with bidirectional human interaction exposes phenomena that neither perceptual JEPA nor disembodied active inference can address.

## 3 The Human-Centered World Model

The construct that the substrate of this paper is intended to realize we will call the *Human-Centered World Model* (HCWM). The term is introduced here as a working noun for a configuration that the patent disclosure (PCT/EP2024/076634, p. 19) already describes informally and that the JEPA family and active inference describe in their respective vocabularies; the construct is not new with this paper, but the term gives it a single-word handle that we can use throughout what follows.

**Definition.** A Human-Centered World Model is a predictive architecture in which the human is not simulated but is instantiated as a physical controller node — either (a) operating in parallel with the machine controller in real time, or (b) integrated as a sequential stage in the controller's program sequence. The architecture is centered on the human in the precise sense that the externalized predictive scaffold — encoder, predictor, error functional, action policy — takes the human's sensorimotor state, expert judgment, or response trajectory as a primary controlled variable rather than as an observed dataset element. Both modes (a) and (b) are admitted explicitly by the patent disclosure; the construct is the same under either configuration.

**Three anchors.** The HCWM construct rests on three anchors that the rest of the paper returns to in turn:

(1) a **patent anchor** (Patent PCT/EP2024/076634, pp. 17–20), in which the human is given the formal status of a controller in parallel or in sequence with the machine, with sensorimotor state $x$, experiential setpoint $w$, and active output $y = f(x)$ — speech included as a first-class output channel of $f$ (p. 19);

(2) an **architectural anchor**, in which the construct is realized in a four-actuator EtherCAT closed-loop testbed with deterministic hard-real-time control on one compute tier and an on-device language model on a separate compute tier (Section 5);

(3) a **theoretical anchor**, at which the construct sits at the intersection of the JEPA predictive architecture (Section 1) and the predictive-coding / active-inference frame (Section 2), with notation that aligns term-for-term with both (Table 2).

**What the HCWM is not.** The HCWM is not a clinical claim; it is not an assertion that any specific therapeutic outcome follows from instantiating a substrate of this kind. It is not a commitment to a single biomechanical or neurological hypothesis: the substrate is built to admit multiple hypotheses through its run-time configurability, beginning with the six configured phase patterns of Section 6.2. And the HCWM is not a competitor to the JEPA family or to active inference; it is the construct that the substrate question of Section 4 is asking about, and the proposal of this paper is that the JEPA program and the HCWM construct are mutually informative rather than competitive.

**Recent literature anchor.** Beyond the patent, architectural, and theoretical anchors named above, the HCWM construct has a recent literature anchor in Riener's editorial framing of the *AI to Eye* collection [18], where it adopts Chen's definition of intelligence as "the capacity of an agent to predict future states of the world and act in ways that realize preferred outcomes" [5], with the additional observation, in the same essay, that "intelligent predictions must be executable in the world by the agent that makes them." The HCWM, as defined above, is precisely such an executable configuration: its substrate provides the channel between prediction and realisation, rather than the prediction alone.

**The Extended Predictive Hypothesis.** [5, Sec. 4.5] develops this position into the Extended Predictive Hypothesis (EPH), in which intelligence is "the ability to predict the future accurately and to benefit from that prediction." The benefit-from-prediction operator that the EPH adds to predictive processing — decomposed by Chen, after [10], into the proposal of afforded actions and an embodied state-evaluation function — is left latent by perceptual JEPA architectures and is the precise operator that the HCWM substrate is designed to expose. We return to this decomposition when discussing the patent foundation in Section 5.

**A practitioner anchor that predates the construct.** The HCWM construct as named here has a practitioner anchor that predates its formal computational expression by approximately seventeen years. The present author's 2008 MBA thesis, *The New CEO in face of the 21st century* (GSBA Zürich / University of Maryland), develops a benchmark-against-reference framework in which a learner (*Joe*) reduces a per-capability gap to an exemplar-aggregated reference profile (*Jack*), itself constructed by Delphi-style aggregation across a panel of fifty-five domain experts. The construct uses arithmetic on capability values (Tables 6–8 of that thesis) and a gap-driven action policy on the resulting deviation. The HCWM construct of the present paper is, in this lineage, the 2026 operationalisation of the 2008 framework: the dimensionality changes from fifty-two capability values to an embedding dimension $n$ of a learned encoder, the domain shifts from strategic management to sensorimotor rehabilitation, and the aggregation method admits additional lineages (see Section 2), but the structural commitment — predict toward a reference $w$, observe the deviation $w - x$, act to reduce it — is preserved. We name this anchor here for completeness and for the seventeen-year practitioner continuity it documents; the remainder of the paper does not depend on it.

## 4 The Substrate Gap

If predictive coding and the JEPA family describe a computational *form* — encoder, predictor, error functional, action policy — the natural follow-up question is what constitutes a *substrate* for that form. We use the term *substrate* in a deliberately narrow sense: the physical and computational system inside which the predictive scaffold runs in real time, with measurable consequences that are not reducible to the mathematics of the scaffold itself. By this definition, a brain is a substrate, a robot under closed-loop control is a substrate, a simulation environment is a substrate, and a recorded dataset is not a substrate. The distinction matters because the JEPA literature, as the literature on classical predictive coding before it, has so far been dominated by the dataset and simulation case.

We make three observations about the published state of the JEPA program as of early 2026, with the caveat that the field is moving quickly and that all three statements should be read as descriptive of what is currently in the public record rather than as exhaustive claims.

**Observation 1 — Perceptual, not actuating.** Every published JEPA-style architecture we are aware of consumes a sensor stream and produces a representation. None close the loop through a deterministic actuator path with hard real-time guarantees, and none place a human controller in parallel with the machine within the same loop. The action policies that appear in active-inference formulations of JEPA are evaluated either offline on recorded trajectories or in simulation; the actuator-side commitment is not made.

**Observation 2 — Datasets and simulators, not closed-loop physical substrates.** The training signal in the JEPA literature is overwhelmingly drawn from large datasets (image, video, speech, point cloud, sensor records) or from physics simulators. Where physical robots appear at all, they appear as deployment targets after offline training, not as the medium in which the predictive structure is exercised in real time. The verification claims that follow from this are accordingly limited: they are claims about representation quality on held-out data, not claims about closed-loop behavior in a substrate that can be externally instrumented and audited under hard real-time constraints.

**Observation 3 — Open architectures, not protected substrates.** The JEPA program, in line with mainstream practice in machine learning, has been published openly and without intellectual-property protection on the architectures themselves. AMI Labs (LeCun et al., founded 22 January 2026) is positioned to extend the JEPA program toward Advanced Machine Intelligence on this same open basis. This is a deliberate choice by the field, and we make no normative claim against it. We note only that an alternative configuration is possible, in which the substrate that grounds a predictive architecture is itself protected by a granted-or-pending patent position, with the substrate licensable to substrate-consumers. This configuration is the one we describe in Sections 5 and 7.

**An external corroboration.** A definitional argument independent of the substrate question reaches a structurally similar conclusion. [5, Sec. 5.1] holds, under the EPH framework, that "AI agents that are now widely used in business cannot be considered intelligent due to their lack of spontaneous predictive ability or gainfulness across the elements of intelligence," and characterises the missing elements as an active action–perception cycle in the sense of [13], continuous (post-training) updating of the reward function, a Bayesian update of the predicted probability distribution, and what [8] calls a "sense of for-me-ness." We adopt this characterisation for the present argument: the substrate described in the remainder of the paper is intended to materialise exactly these elements at the hardware level, where they can be externally instrumented rather than asserted.

**Closing the gap.** The gap we identify is therefore not a gap in the formal scaffold — the JEPA family and its biological precedents have specified the form in considerable detail. The gap is at the substrate level: a real-time, multi-actuator, deterministically controlled, externally instrumentable, patent-disclosed system in which the scaffold can be exercised under conditions that perceptual JEPA and disembodied active inference cannot exercise it. The remainder of the paper describes one such substrate, instantiated in working hardware, and proposes that substrates of this kind are a complement to, not a competitor of, the open JEPA research program.

**Embodiment, after Bender–Koller and Riener.** Riener [18] sharpens the grounding critique of [4] as follows: "Intelligence without embodiment can simulate understanding but cannot fully participate in it." The substrate described in the remainder of the paper is intended to operationalise the word *participation* in that sentence — not as a software-side metaphor, but as a hardware-grounded measurement channel between the actuator and the proprioceptor, in the explicit human-as-controller construction of the patent disclosure (Section 5.1).

# 5 HCWM Architecture — A Patent-Disclosed Multimodal Closed-Loop Substrate

This section describes the substrate that we propose as one instantiation of the gap identified in Section 4. The description proceeds in five layers, working outward from the patent-disclosed core: **(i)** the legal and conceptual foundation in two granted-or-pending patent positions; **(ii)** the actuator hardware platform that exposes a joint-position-velocity-torque control mode; **(iii)** the coordination layer that runs above the platform and constitutes the original contribution of the present author; **(iv)** the four-modality observable surface; **(v)** the two-tier compute architecture that separates hard-real-time control from asynchronous language-based decision support. We close the section by relating the patent's notation to the JEPA family and to active inference through a triple mapping (Table 2).

## 5.1 Patent Foundation

The substrate rests on two patent positions held by Swissrehamed GmbH, the affiliation of the present author. The granted predecessor patent EP 2 364 686, filed 2009, granted in the European jurisdiction in 2015, and in force through 2030, covers the original closed-loop training device. The pending application PCT/EP2024/076634 (priority date 27 September 2023 from EP 23 200 102.4; international PCT filing 23 September 2024) extends the disclosure to the multimodal closed-loop configuration described in this paper, including the human-as-parallel-controller construction and the time-of-flight perception channel. The PCT application has entered international publication; national-phase filings (EP, US, JP) were submitted on 27 March 2026. The position that follows in this section is consistent with what is disclosed in those filings; we use the internal page references pp. 17 through 20 of the disclosure document, which is the form in which the present author refers to the patent text in the engineering record.

The pivotal passage for the present argument is the explicit construction of the human as a controller in parallel with the machine, in the Closed-Loop Learning Mode (p. 19):

> "a human being is also a controller with $x$ (all five senses), $w$ (patterns based on experience . . .), and the becoming active to the four extremities, *to speech* or to thinking etc. as $y = f(x)$." — Patent PCT/EP2024/076634, p. 19

This construction is what makes the human-machine loop bidirectional and what justifies the inclusion of speech as a first-class *actuator-side* modality in Table 1, rather than as an afterthought of the language-model layer.

**Mapping to the EPH benefit operator.** [5, Sec. 4.3] decomposes the benefit-from-prediction operator of the EPH into two abilities — proposing afforded actions in the sense of [10] and assigning a value to the agent's current state. The patent-disclosed substrate maps onto both: the encoder channel materialises perceived affordances at the proprioceptor level (pp. 17–19 of PCT/EP2024/076634), and the closed-loop kinematic readout provides the state-evaluation function in the agent's own embodiment frame — the absence of which Chen identifies as the loss of "the causal basis of intelligence and behavior." The mapping is structural rather than computational; we discuss the computational form in the triple mapping of Table 2.

## 5.2 Hardware Platform: HEJ-90 in JPVT Mode

The actuator side of the substrate is implemented with four maxon HEJ-90 high-efficiency joint actuators on an EtherCAT bus. The bus and the Jetson-class compute tier support hard-real-time cycle rates substantially above 100 Hz; the empirical results reported in Section 6 were collected at the production-default 100 Hz cycle, which is the cycle rate referenced throughout the remainder of this paper unless otherwise stated. The HEJ firmware exposes a joint-position-velocity-torque (JPVT) operating mode that is the intended primary mode for the platform; the legacy cyclic-synchronous modes (CSP, CSV, CST) are supported but are not the recommended mode of use for this hardware family. In JPVT mode, the firmware accepts three independent setpoints — target position, target velocity, and target torque feedforward — and combines them into a torque demand of the form

$$\tau_{\text{demand}} \;=\; k_P \cdot (\theta_{\text{target}} - \theta_{\text{actual}}) \;+\; k_D \cdot (\omega_{\text{target}} - \omega_{\text{actual}}) \;+\; \tau_{\text{ff}}, \tag{1}$$

where $k_P$ is a position gain, $k_D$ is a velocity gain, and $\tau_{\text{ff}}$ is a direct torque feedforward channel.

The structural idea behind Equation 1 — driving a joint-position-controlled system through a torque setpoint via a velocity feedforward transformer — was first described in the open patent literature by [21] (filed 2008, lapsed 2009). The Honda filing has since entered the public domain. The HEJ-90 firmware constitutes one specific commercial implementation of this class of control law, with hardware-specific tuning, sensor fusion between motor-side and joint-side encoders, and a proprietary cogging-compensation layer; we make no claims about those internal implementations and we treat the JPVT mode at the boundary, that is, through the documented PDO interface and the documented gain parameters. All algorithmic novelty in this paper sits in the coordination layer (Section 5.3) above this platform, not in the platform itself.

### 5.3 Coordination Layer: The Round Tread Supervisor

Above the JPVT platform we run a coordination layer, the *Round Tread Supervisor*, which is the original engineering contribution of the present author and the layer in which Patent PCT/EP2024/076634 is realized in software. The supervisor is responsible for

(a) phase coordination across the four actuators, which includes a configurable phase pattern and an online phase-correction stage that compensates for inter-joint drift;

(b) analytic torque feedforward, including a thirty-two-bucket gravity feedforward look-up table per actuator and a $\cos(2\theta)$ deadpoint feedforward term derived from the geometry of the pedal mechanism, both of which originate in the disclosure of EP 2 364 686 (2009) and are reactivated under closed-loop conditions in the present substrate;

(c) run-time selection of the operating regime through explicit command-line parameters that gate the feedforward terms and the velocity-loop derivative gain — this is the substrate's externally instrumentable interface to the forward-dominant / inverse-dominant distinction of Section 2;

(d) observation channels (CSV diagnostic logging, microphone capture for acoustic envelope analysis, optional time-of-flight capture) that emit the data on which the substrate-level verification of Section 6 rests.

The supervisor runs as a single-threaded EtherCAT-driven control loop on the safety-classified compute tier described in Section 5.5.

### 5.4 Four-Modality Observable Surface

The substrate exposes four complementary observable streams, summarized in Table 1. The actuator stream is the primary one and the only one that closes the control loop; the remaining three are passive observations of the loop's behavior, used either as external verification (audio envelope), as exogenous-perturbation inputs (time-of-flight), or as language-mediated decision support (bidirectional speech). The patent column of the table indicates whether the modality is named explicitly in the patent disclosure (pp. 17 through 20) or is compatible with the generic disclosure language.

Table 1: Four-Modality Sensorimotor-Cognition Substrate.

| Modality | Description | Patent | JEPA Connection |
|---|---|---|---|
| Actuator | 4×EtherCAT torque/position output (HEJ-90), 100 Hz hard-RT | p. 17–19 explicit | *empty slot* |
| Time-of-Flight | $z$ = interference variable, $(x, y, z)$ position model | p. 20 explicit | LiDAR-/AD-LiST-JEPA |
| Audio envelope | Hilbert modulation depth, prediction-error in frequency domain | pp. 19/20 generic | GMM-Anchored JEPA [12] |
| Speech (bidir.) | Human-to-LLM and LLM-to-human via on-device language model (Apertus 8B) | p. 19 explicit (human→); pp. 17/20 compatible (system→) | GMM-Anchored + LLM-in-loop |

The empty-slot annotation in the actuator row of Table 1 repeats the central observation of Section 1: across the entire JEPA-style literature we are aware of, no published work consumes a closed-loop actuator stream as its primary training or evaluation signal. The other three rows align the substrate with existing JEPA branches, providing a citation path into LiDAR-style point-cloud predictive learning, into the GMM-anchored speech variant of the family [12], and into the broader LLM-in-the-loop literature.

**Optional complementary modalities not in scope of this paper.** The four-modality set above is the substrate's current observable surface; it is not exhaustive. Optional complementary modalities have been investigated by the present author in earlier sensor-research lines, including time-of-flight underwater perception (in collaboration with academic photonics partners since 2016, the patent's $z$-variable channel) and 3D range-imaging contactless authentication via height-profile geometry (a successfully demonstrated proof-of-concept under an Innosuisse innovation cheque in 2020). A video / camera channel — for gaze tracking, attention monitoring, low-back muscle-activity surface estimation, or geometric authentication — would extend the observable surface without altering the substrate's core actuator-driven closed-loop architecture.

A second class of optional complementary observation channels is patient-side physiology — heart-rate variability, respiratory rate, surface electromyography, vocal envelope, or other directly-on-the-patient measurements — in line with research lines investigated by the present author since 2011 (rocking-chair HRV research) and revisited under the patent's Adapted-Scenarios construction (p. 18). These channels can drive a biofeedback loop in the classical sense (the patient is presented with a representation of their own physiological state and learns to modulate it), in which case the loop closes through the asynchronous tier rather than through actuator authority; *biofeedback* in this usage is the loop architecture, not the channel itself. We use the term *channel* here, rather than *modality*, because no universal standardization exists for the latter in the multi-modal AI literature; the relevant naming references on the patient-data side are clinical-data standards (HL7 FHIR, DICOM where imaging, SNOMED CT, LOINC) and on the sensor-hardware side IEEE 1451. These channels are conceptually distinct from the acoustic-envelope row of Table 1, which observes the machine side of the loop. The substrate's two-tier separation (Section 5.5) accommodates patient-side channels naturally on the asynchronous tier, without contaminating the actuator authority of the hard-real-time tier.

We treat all such modalities — visual, geometric, physiological — as add-ons to the present architecture rather than as the core observation channel. The distinction matters because it differentiates the present substrate, where the actuator-side closed loop is the primary ground-truth source, from application-layer products where camera-based motion capture or sensor-only physiological monitoring is itself the core sensor; the substrate's differential is preserved without these channels and is preserved more strongly with them, but is not constituted by them.

### 5.5 Two-Tier Compute Architecture

The compute side of the substrate separates strictly into two tiers, each running on its own embedded compute unit.

**Tier 1 — Hard-real-time control.** A first compute unit (Jetson-class, embedded Linux with PREEMPT_RT kernel) runs the EtherCAT master and the Round Tread Supervisor at 100 Hz with documented cycle-time jitter on the order of tens of microseconds. This tier is intended to operate within the constraints of an IEC 62304 Class C software item; it carries the safety-relevant control authority over the actuators and is the only tier with write-access to the actuator setpoints. The same tier also owns the time-of-flight perception channel of Table 1, since the interference signal $z$ of the patent (p. 20) is part of the closed control loop and must be processed deterministically in the same hard-real-time cycle as actuator command and sensor fusion. A run-time exergame layer that takes its own perceptual input from the same Tier-1 process executes alongside the supervisor.

**Tier 2 — Asynchronous language-based decision support.** A second compute unit, on a separate physical board, runs the on-device language model (Apertus 8B in our reference configuration) with inference latencies in the 100 ms to 500 ms range. This tier implements the human-to-system and system-to-human speech channels of Table 1 and can produce decision-support output in the sense of patent p. 17 ("computer program product such as a learning algorithm"). It is intended to operate within the constraints of an IEC 62304 Class B or Class A software item, since its outputs reach the patient or therapist only through the human operator and never directly through the actuator authority of Tier 1.

**An intermediate learning tier is anticipated.** The two-tier separation above is the deployed configuration at the time of writing. An intermediate *learning tier* that hosts data-driven forward and inverse models (deep-learning predictors, representation learners, action-conditioned world models) between the hard-real-time control tier and the asynchronous language tier is anticipated and is discussed at the architectural level in the companion predictor paper. We name this here, ahead of that deployment, with a mnemonic that the present author has used in working notes since December 2025: "the real-time-operating-system tier is nerves and reflexes, the deep-learning tier is learned motor control, the language tier is awareness and articulation." The mnemonic is a working convention, not a regulatory claim; the safety and IEC 62304 classification of the learning tier is itself an open research item and is flagged as such in Section 8.

The strict tier separation, with no direct write-path from the language model to the actuator, is what allows the language layer to be present in the substrate without contaminating the hard-real-time guarantees of the control loop. The design rationale for this strict separation draws on historical safety failures in medical software — in particular the Therac-25 incidents of the 1980s, in which user-interface inputs reached therapy-critical actuators without independent validation — and treats every user-facing or language-model-mediated input as a regulatory artifact that must pass through the operator and through the design-control file rather than directly to the actuator-authority tier. The interaction between the two tiers is the subject of one of the open questions in Section 8.

**Action-conditioning and JEPA-style consumers.** Action-conditioned variants of the JEPA family, in particular V-JEPA 2 [3], take as input a perceptual stream together with the action signals sent to a robotic actuator and learn to predict the embedding of the next perceptual state given both. The substrate of this section produces exactly such pairs — deterministic actuator commands synchronized to actuator-side and human-side observation streams — as a natural by-product of its operation, since the supervisor records the full set of setpoints and the firmware returns the corresponding torque-demand and sensor-fusion outputs at the cycle resolution of the control loop. We mention this here, in the architectural section, because it is the most concrete substrate-consumer hypothesis we are aware of for this kind of work: a JEPA-style action-conditioned predictor trained or evaluated on data drawn from a substrate of the kind described here.

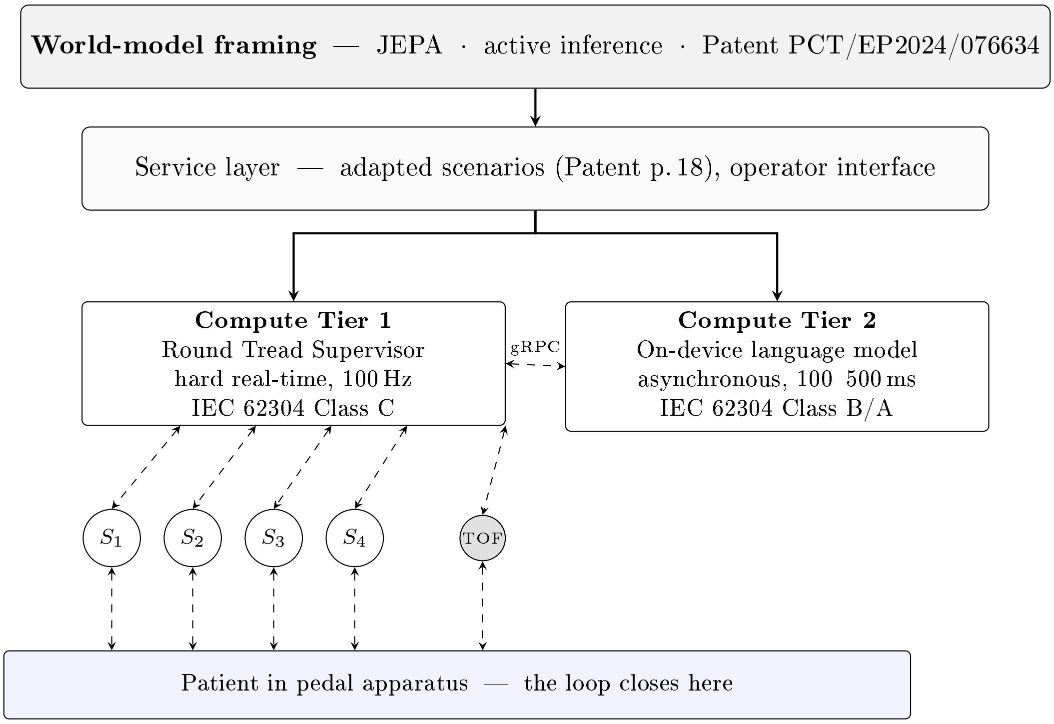


Figure 1: Five-layer stack of the substrate. The world-model framing (top) is the theoretical anchor of Sections 1 and 2. The service layer hosts the adapted-scenarios construction of the patent and the operator interface. Compute Tier 1 runs the Round Tread Supervisor under hard-real-time constraints and is the only tier with actuator-authority; it also owns the time-of-flight perception channel, since the interference signal $z$ of the patent (p. 20) must be processed inside the hard-real-time control loop. Compute Tier 2 runs the on-device language model asynchronously and has no direct write-path to the actuators or to the time-of-flight channel. Four HEJ-90 joint actuators ($S_1$ through $S_4$) and the time-of-flight sensor close the loop on the patient in the pedal apparatus. Solid arrows denote authority flow; dashed double arrows denote bidirectional sensing-and-action.

### 5.6 Patent–JEPA–Active Inference Triple Mapping

The notation in the patent text was adopted at filing time for classical control-theoretic reasons. With the JEPA family and the active-inference framing now established, the same six symbols turn out to align term-for-term with both, as Table 2 shows.

**Sourcing the reference variable $w$.** The reference variable $w$ admits two distinct sourcing paths within the present construct, and the substrate is agnostic between them. $w$ may be *experiential* in the patent's sense (p. 19, "patterns based on experience") and arise inside the patient's own generative model, or it may be *externally provided* as a reference profile constructed from one or more exemplars judged competent in the target task — the practitioner anchor of Section 3. The substrate observes the deviation $w - x$ identically in either case; the architectural slot for external reference profiles is on the same asynchronous tier as the language model (Section 5.5) and is consumed by the same gap-driven action policy. The patent disclosure already

Table 2: Triple Mapping: Patent PCT/EP2024/076634 ↔ JEPA ↔ Active Inference.

| **Patent (pp. 17–20)** | **JEPA Predictive Architecture** | **Active Inference (Friston)** |
|---|---|---|
| $x$ = actual value | latent representation $z_t$ | sensory state |
| $w$ = setpoint | predicted target $\hat{z}_{t+1}$ | preferred state |
| $y = f(x)$ = control signal | policy / action | action |
| $z$ = interference variable (TOF) | exogenous perturbation | environmental perturbation |
| $w - x$ = control deviation | prediction error | free-energy gradient |
| DSS / experts (p. 18) | expert-tuned policy | precision-weighted prior |

anticipates a graphical reproduction of this deviation: the present author's engineering record for the granted predecessor patent EP 2 364 686 contains the formulation that "the deviation, after gap analysis, is rendered in a graphical form modelled on an anatomical reference" — a verbatim instance of the gap-against- reference pattern at the visualisation layer, predating both the JEPA family and the present manuscript.

We do not claim that this alignment is more than structural. The patent and the JEPA family arrived at the notation independently, and the patent makes no commitment about the computational implementation of the predictor. What the alignment does provide is a concrete translation table between three vocabularies that are otherwise disconnected — one from control engineering, one from machine learning, one from cognitive neuroscience — and a basis for reading the empirical results of Section 6 simultaneously in all three.

### 5.7 Phase Patterns as Architectural Choice

A final architectural feature of the supervisor is that the phase pattern across the four actuators is itself a parameter: the system runs a configurable phase pattern, expressed as four phase offsets modulo $2\pi$, spanning the full admissible phase space across the four actuators. We have characterized six configured phase patterns, and we report their empirical characterization in Section 6. The phase-pattern axis matters here, in the architecture description, because it is the most direct demonstration that the substrate is not committed to a single biomechanical hypothesis: the patent permits, and the supervisor implements, a continuous phase-space across which different biomechanical hypotheses can be tested under otherwise identical control conditions.

## 6 Empirical Results — Substrate Diagnostics, Phase Patterns, and Mode Coverage

The empirical content of this section is presented in three layers, each one a different use of the substrate as an instrument: a survey of phase patterns, a diagnostic series in which the substrate is used to identify and discriminate sources of prediction error in its own behavior, and a coverage statement of operating modes that distinguishes what has been demonstrated from what is open empirical work. None of the results below claim a clinical effect or a controlled trial; the present testbed is a substrate-characterization rig, and we report it as such.

### 6.1 Methodology and Instrumentation

Every run is logged through three channels. **(a) Control-loop telemetry.** Per-cycle CSV traces at 100 Hz record target and actual position, target and actual velocity, the torque-feedforward setpoint, and the torque-demand reported by the firmware ($\tau_{\text{demand}}$ from Equation 1). Logging runs at full cycle resolution (`-print-every 1`) for diagnostic runs and at sub-sampled resolution otherwise. **(b) Acoustic capture.** A microphone external to the testbed records 35 s WAV files at 44.1 kHz mono per run, used downstream as an external verification channel through Hilbert-envelope modulation analysis. The microphone is acknowledged as a noisy, indirect observable; we take it as a frequency-domain check on the control-loop telemetry, not as a perceptual ground truth. **(c) Pilot subjective notes.** For manned runs, the pilot records a brief subjective note on the perceived smoothness of the motion. These notes are not a primary endpoint of the present paper but are reported alongside the quantitative measures because acoustic modulation depth and human perception of pulsation do not coincide exactly — the microphone measures the modulation of sound pressure across one revolution, not the spike character that the human ear labels as a pulse.

The control-loop parameters are exposed at the command-line interface of the supervisor, which makes the substrate's operating regime transparent and reproducible: the gravity feedforward gain (`-gravity-scale`), the analytic deadpoint feedforward amplitude (`-deadpoint`), the velocity-loop derivative gain on the phase corrector (`-phase-kD`), and the phase pattern itself are all CLI-selectable parameters. The forward-dominant / inverse-dominant distinction of Section 2 is therefore a literal command-line distinction in this substrate, not a conceptual one.

**Microworld posture.** The testing posture above is consistent with the *microworld* methodology that [5, Sec. 4.5 and 5.1] proposes as the appropriate fair-comparison method for evaluating predictive ability and benefit-from-prediction in both humans and AI agents: a controlled environment in which input conditions and prior knowledge can be aligned across test subjects and in which the substrate's predictions and their measurable consequences can be verified without reliance on benchmark datasets susceptible to leakage. The present substrate functions as such a microworld at the hardware level, with the additional property that no leakage between training data and evaluation environment is possible: the evaluation environment is the real actuator chain itself, instrumented per Section 6.1.

## 6.2 Phase Pattern Survey

Six configured phase patterns have been characterized on the substrate. Table 3 gives each in two forms: the controller-side `--phases` command values that reproduce it, and the resulting anatomical limb phases.

Table 3: Phase pattern survey on the four-actuator substrate. The *controller* column gives the command values (`--phases` $S_1\,S_2\,S_3\,S_4$, degrees) that reproduce each pattern; the *anatomical* column gives the resulting physical limb phases. Actuators $S_3$ (left leg) and $S_4$ (left arm) are reverse-wired (`dir` $= -1$, $\theta_{\text{ctrl}} = -\theta_{\text{act}}$), so on those two channels the physical phase equals the command value $+180°$; the two legs $S_2/S_3$ are consequently held in $180°$ mechanical opposition in every pattern. Limb assignment: $S_1$ right arm, $S_2$ right leg (reference), $S_3$ left leg, $S_4$ left arm. The status column gives the longest characterized steady-state duration at 40 rpm and whether the pattern has been operated manned.

| **Code** | **Pattern** | **Controller** | **Anatomical** $S_1/S_2/S_3/S_4$ | **Status** |
|---|---|---|---|---|
| E0 | Round tread (production reference) | 315 / 0 / 0 / 45 | 315 / 0 / 180 / 225 | $\leq$ 120 s, manned |
| E3 | Tölt | 315 / 0 / 0 / 315 | 315 / 0 / 180 / 135 | 60 s |
| G0 | Walk | 90 / 0 / 0 / 90 | 90 / 0 / 180 / 270 | 60 s |
| G1 | Trot | 180 / 0 / 0 / 180 | 180 / 0 / 180 / 0 | 60 s |
| G2 | Canter | 120 / 0 / 60 / 180 | 120 / 0 / 240 / 0 | 60 s |
| G4 | Pace | 0 / 0 / 0 / 0 | 0 / 0 / 180 / 180 | 60 s |

Two observations follow from the survey. First, the round-tread pattern E0 retains the lowest aggregate phase-tracking residual across the measured patterns (sum-of-RMS $\approx$ 71 counts at 40 rpm, against $\approx$ 93 for pace and $\approx$ 99 for tölt), and remains the production reference for manned operation. Second, because $S_3$ and $S_4$ are reverse-wired, the two legs $(S_2, S_3)$ are held in 180° opposition in every pattern, while the arms and the leg–arm relationship carry the pattern-specific structure: the round tread E0 places each arm 45° from the same-side leg, and the trot G1 reduces to the contralateral diagonal (right-leg + left-arm versus right-arm + left-leg). The substrate operates stably across all six patterns, which establishes that the full phase space — not only the leg-opposed symmetric configurations — is available for further pattern design.

### 6.3 Substrate Diagnostic Series — Discriminating Mechanical from Feedback-Loop Sources of Prediction Error

A short diagnostic series in which the substrate was used as an instrument on its own behavior is reported here, both as a demonstration of what the substrate-characterization channels of Section 6.1 can do and because the series happened to expose a clean separation between two structurally different sources of prediction error in the control loop. The series consisted of four unmanned and nine manned runs at the round-tread pattern E0, recorded as a set of thirteen synchronized audio captures and nine fully resolved CSV traces (May 2026 session, dataset `2026-05-01-knorzen-diagnose`).

**Single-pilot disclosure.** The manned runs in the diagnostic series were performed by a single pilot (the present author). The results that follow are accordingly substrate-characterization results — statements about what the hardware does — and are not statements about between-subject variability. Multi-pilot extension is open empirical work, and the broader question of how a reference profile $w$ aggregated across multiple competent exemplars is to be operationalised on the substrate is one of the open questions raised in Section 8.

**Result 1 — Identification of mechanical pendulum mass.** Per-actuator torque profiles, binned into thirty-two phase buckets per revolution, exhibit a strong sinusoidal $1\times$rev component with peak amplitudes $\pm 2.35$ Nm on $S_1$ and $\pm 3.33$ Nm on $S_4$. The ratio of $1 \times$ rev amplitude to $16\times$rev amplitude exceeds five across all four actuators and reaches sixteen on one of them. This rules out cogging as the dominant source (cogging would produce a high-order signature locked to the motor pole pair, not a $1 \times$ rev signature) and identifies the source as the gravitational and inertial signature of the pedal mechanism's own mass. The substrate, in other words, can identify its own pendulum-mass profile without external instrumentation, by reading back the torque-demand register that the firmware exposes.

**Result 2 — Substrate-level forward-model action.** A thirty-two-bucket gravity-feedforward look-up table, populated directly from the per-actuator profiles of Result 1 and applied as $\tau_{\mathrm{ff}}$ in Equation 1, attenuates the $1 \times$ rev component of the torque-demand. As an external check, the Hilbert envelope of the microphone capture exhibits modulation depth 84% at the $1 \times$ rev peak (0.25 Hz at 15 rpm) without the feedforward, and 54% with the feedforward applied at gain 0.7 over a 60 s steady-state run — a 36 percentage-point reduction. This is reported as one external verification of substrate-level forward-model action; it should not be read as a perceptual claim, since the microphone measures sound-pressure modulation rather than the spike character of perceived pulsation.

**Result 3 — Reactivation of analytic feedforward disclosed in 2009.** The pedal mechanism has a geometric structure that produces a top-and-bottom dead-center signature with a $\cos(2\theta)$ phase profile. This term, and an analytic feedforward to compensate it, was described in the granted patent EP 2 364 686 (filed 2009). The present substrate exposes the analytic dead-point feedforward as a command-line parameter (`-deadpoint <amplitude>`), which makes the corresponding term of the original patent disclosure live again as a tunable feedforward channel.

We do not claim novelty for the analytic form; we claim only that the substrate is the testbed in which the form, sixteen years after its disclosure, can be evaluated under closed-loop conditions and against an external acoustic verification channel.

**Result 4 — Forward-dominant versus inverse-dominant regimes.** Under manned operation at the production sweet-spot (`-gravity-scale 0.7 -deadpoint 5000 -phase-kD 0`, pattern E0, 25 rpm), disabling the velocity-loop derivative gain on the phase corrector eliminates a residual pulsation that the pilot identifies subjectively as a sharp "pulse" character. The underlying interpretation is that, at this sweet-spot, the forward-model contribution (gravity LUT plus analytic deadpoint) is sufficient to track the motion and that the inverse-model contribution (the derivative term of the phase corrector) was itself acting as a secondary pulse source under the residual phase error. This is a direct empirical instance of the forward-dominant / inverse-dominant distinction introduced in Section 2, and it occurs because the substrate exposes the two regimes as independently switchable parameters of the same control law.

### 6.4 Mode Coverage and Open Empirical Items

The substrate's coverage of control modes and patterns at the time of writing is summarized in Table 4.

Table 4: Control-mode and operating-regime coverage at the time of writing.

| **Mode** | **Status** |
|---|---|
| JPVT joint position / velocity / torque (Equation 1) | demonstrated, primary mode |
| Forward-dominant feedforward (gravity LUT + $\cos(2\theta)$ deadpoint) | demonstrated, manned at 25 rpm |
| Inverse-dominant feedback (phase-corrector with active $k_D$ term) | demonstrated, manned |
| Hybrid forward+inverse (continuous in CLI parameters) | demonstrated, manned |
| Pattern E0 round-tread, manned | demonstrated, primary pattern |
| Patterns G0–G4 and E3, manned | open empirical work |
| Pattern G2 at 40 rpm steady-state | open empirical work |
| Patterns G0–G4, 300 s endurance | open empirical work |
| Power-mode (mechanical-power setpoint, $P = \tau \cdot \omega$, sport regime) | open architectural item, see Section 8 |
| Time-of-flight perception channel (Section 1, row 2) | not yet integrated; firmware path designed |
| Bidirectional speech channel (Section 1, row 4) | language-model layer in implementation, asynchronous tier separation per Section 5.5 |

The mode-coverage statement is intentionally explicit. The substrate demonstrates the JPVT-based control law, the forward-/inverse-regime distinction, and the round-tread pattern under manned operation. It does not yet demonstrate the alternative phase patterns under manned operation, pattern G2 at full pedal speed, the long-duration endurance behavior of the alternative patterns, the power-setpoint regime, the time-of-flight perception channel, or the closed bidirectional speech channel. Each of these is a deliberate open item, several of which are taken up as open research questions in Section 8.

## 7 Substrate Licensing and Position in the Commercial Landscape

The substrate-licensing argument introduced in Section 4 deserves a brief expansion, because it is the part of this paper that sits closest to a commercial proposition and because it is easily misread as a competitive claim against existing neurorehabilitation products. It is not. The argument is that the substrate layer and the application layer are different layers, that they are addressed by different products, and that the substrate layer has so far been under-instantiated.

### 7.1 Substrate Layer versus Application Layer

A useful analogy is the relationship between a foundational dataset such as ImageNet or the Common Crawl, and the application built on top of it. The dataset is the substrate; the application is what consumes the substrate. The substrate layer has economic and scientific value in its own right — it is what makes downstream work possible at scale — and it is licensable to consumers of that downstream work. The substrate is not an application. The same separation applies to the present work: the four-actuator closed-loop testbed, the JPVT-based control law, the Round Tread Supervisor, and the patent position around them are a substrate, intended to ground downstream research and downstream products. The substrate is not itself a clinical device, a therapy product, or a consumer offering.

### 7.2 Comparison to Existing Commercial Neurorehabilitation Products

Commercial neurorehabilitation today is dominated by sensor-and-software products that provide patient-facing therapy through a tablet or console interface, with motion-capture sensors and gamified feedback. The most-deployed family, MindMaze's MindMotion GO, Izar, and Physilog [16], has FDA 510(k) Class II clearance, deployment across more than one hundred clinics in Europe, and a United-States stroke-rehabilitation pilot of approximately two hundred ten patients. Table 5 contrasts this product family with the present substrate along the dimensions that matter for the substrate / application separation introduced above.

Table 5: The application layer (left) and the substrate layer (right) address different dimensions of the same problem space. The intent of the table is differentiation, not competition.

| **Dimension** | **MindMaze product family** | **This work (substrate)** |
|---|---|---|
| Layer | application (patient-facing therapy) | substrate (testbed and licensable platform) |
| Hardware character | tablet plus motion-capture sensors and a passive hand controller | four-actuator EtherCAT closed loop, 100 Hz hard-real-time |
| Closed-loop construction | software-and-gamification loop | deterministic torque-loop with patent-disclosed multimodal observables |
| World-model claim | no explicit claim | explicit claim, with structural mapping to JEPA and active inference (Table 2) |
| Actuator authority | none (passive instrumentation only) | full actuator authority on four joints |
| Patent position around closed-loop substrate | not present in the product literature | EP 2 364 686 (granted 2009) and PCT/EP2024/076634 (priority 2024) |
| Regulatory pathway | FDA 510(k) Class II cleared product | not a regulated product; substrate intended to be consumed by downstream products that themselves enter regulatory pathways |

The two layers are complementary rather than competitive. The application-layer products demonstrate that a market exists and that the patient-facing therapy concept is reimbursable; the substrate layer asks what foundational platform downstream applications can be built on top of, particularly for use cases that require closed-loop actuator authority and bidirectional cognitive interaction rather than passive instrumentation.

**Market context.** The application-layer commercial neurorehabilitation segment underwent significant restructuring during 2025 and early 2026, with several global hardware-platform companies entering insolvency, operational suspension, or merger-driven re-capitalization. We do not interpret these events as evidence that the substrate-layer approach is preferable; we note them only as the contemporary market context against which any new neurorehabilitation-adjacent technical proposal is read. The structural distinction between substrate and application is orthogonal to commercial outcomes at the application layer, and the substrate-layer position taken in this paper is intended to stand on its own technical merits regardless of those outcomes.

### 7.3 Licensability and the Patent Position

A natural question is why a foundational substrate would be patent-protected at all, given that the analogous data-substrate foundations of contemporary AI (ImageNet, Common Crawl) are open. The answer is that an actuator-bearing closed-loop substrate is not analogous to a dataset: it is a physical system that interacts with a human under safety-relevant force and torque, and the disclosure required to operate such a system safely is much larger than the disclosure required to use a dataset. A patent position around the closed-loop construction provides the legal scaffolding under which the substrate can be licensed coherently to downstream consumers without forcing the same disclosure to every user. We treat this as one design pattern for substrate licensing, not as a normative claim that all substrates should be patent-protected. The patent positions referenced throughout this paper are held by Swissrehamed GmbH; substrate-licensing arrangements would accordingly be made between Swissrehamed GmbH and the substrate consumer.

**Two categories of substrate consumer.** Two distinct categories of substrate consumer can be distinguished, with correspondingly different access patterns:

- **Commercial consumers with private capital.** This category includes private AI-research organizations working on JEPA-style or world-model architectures (the AMI Labs program of LeCun et al. in 2026 is the most visible recent example, but other private AI-research organizations belong to the same category), robotics foundation-model companies developing action-conditioned predictive systems on real-world robot data, neurorehabilitation product companies building patient-facing therapy products on top of an actuator substrate, and pharmaceutical organizations developing digital-biomarker substrates for clinical-trial validation. For this category, a paid commercial license to the substrate is the appropriate access pattern.
- **Academic and public-funded research consumers.** This category includes university research groups and public-research-center groups working on JEPA-style, predictive-architecture, biomechanics, or cognitive-neuroscience research. For this category, the appropriate access pattern is non-paying research access in exchange for citation recognition and, where mutually beneficial, co-authorship or joint grant participation. This is comparable to the access models established for foundational datasets such as ImageNet and Common Crawl, where commercial users are licensed and academic users operate under research-use terms.

The two-category structure matters because the dominant funding modes of the two categories are structurally different: commercial private-capital organizations operate under return-on-investment expectations and can absorb a license fee as a research-input cost, while public-funded academic organizations operate under grant-driven budgets that do not, in general, support license-acquisition for foundational research substrates.

**Potential substrate-consumer settings within each category.** Within the commercial category, substrate-consumer settings of particular relevance include the AMI Labs program [14], other private AI-research organizations working on JEPA-style architectures, robotics foundation-model companies developing action-conditioned systems, neurorehabilitation product companies operating in the application layer compared in Table 5, and pharmaceutical organizations developing digital-biomarker substrates. Within the academic category, potential research partners include groups working on JEPA-style and predictive-architecture research at MIT, ETH Zurich, SUPSI, and elsewhere, as well as cognitive-neuroscience and biomechanics groups that find the substrate's instrumentation useful for their own research questions. The substrate is open in principle to dialogue across both categories on the access pattern appropriate to each; we make no specific approach to any single party and have no commitment from any party at the time of submission.

# 8 Open Questions

The substrate as described leaves seven open questions that we believe are productive research directions for the JEPA community, the hardware-platform engineering community, and the cognitive-neuroscience community respectively. We state each question in a form that is scoped to the substrate's current capability and that does not depend on any commitment beyond the patent disclosures already on the public record.

**(Q1) How does an asynchronous language layer co-exist with a hard-real-time control layer without contaminating the latter's guarantees?** The two-tier architecture of Section 5.5 separates a hundred-hertz EtherCAT-driven control loop from a language model with hundred-millisecond inference latency. The bidirectional speech channel of Table 1 runs entirely on the asynchronous tier, with no direct write-path to the actuators. The open question is what categories of language-mediated decision support can be admitted into the human-in-the-loop substrate without crossing into the hard-real-time tier, and how the formal verification of the tier separation is to be conducted under the IEC 62304 framework. This question is open both as a control-engineering question and as a JEPA-architecture question, since any LLM-in-the-loop JEPA variant will have to confront the same latency mismatch.

**(Q2) How are explicit power setpoints best accommodated above the JPVT control law?** The JPVT control law of Equation 1 accepts position, velocity, and torque-feedforward setpoints. It does not accept a mechanical-power setpoint directly, since power emerges as the product $P = \tau \cdot \omega$ rather than as a native firmware register. A sport-mode regime that targets a constant mechanical-power output on the patient or athlete therefore requires an additional control construction on top of the JPVT layer. Three configurations are candidate, in increasing complexity: a direct feedforward mapping $\tau_{\text{ff}} = P_{\text{target}}/\hat{\omega}_{\text{actual}}$ with a filtered velocity estimate; an outer power loop running at a slow update rate that adjusts $\tau_{\text{ff}}$ or $\omega_{\text{target}}$ through an integral controller on the power error; and a hybrid configuration in which the supervisor saturates the feedforward when the instantaneous product exits a configured power band. Each configuration has different stability and safety properties; we have not characterized any of them on the substrate at the time of writing. We invite the hardware-platform engineering community, notably the actuator-platform vendor community, to identify the preferred configuration for a JPVT-based substrate operated under a mechanical-power setpoint, since the answer is likely to be informed by detailed knowledge of the firmware's torque-and-velocity sensing chain that is outside the scope of the present paper.

**(Q3) Does substrate-level grounding improve the predictor quality of perception-only JEPA?** A central testable hypothesis of this paper is that closed-loop embodied grounding adds something to the predictor that perception-only training does not provide. The substrate offers a controlled experimental setup to test this: a JEPA-style predictor trained on the multi-modal observable surface of Table 1 can be evaluated against a JEPA-style predictor trained on the perceptual modalities only. We have not yet conducted this experiment. We propose it as a concrete next step that the JEPA community could conduct collaboratively with the substrate, with no commitment to the substrate's commercial position required for the experimental work itself.

**(Q4) How does the substrate accommodate explicit free-energy-based optimization, beyond the structural alignment of Table 2?** The triple mapping of Table 2 aligns the control deviation of the patent's Closed-Loop Learning Mode with the free-energy gradient of active inference, but the alignment is structural rather than computational. A substrate that explicitly implements expected-free-energy minimization in the action policy remains open work. The two-tier compute architecture of Section 5.5 provides one architectural slot for such an implementation, on the asynchronous tier where the language model already runs and where the latency budget admits the inference cost of a Bayesian policy evaluation. We mention this as a research direction, not as a commitment.

**(Q5) What does the substrate add for asymmetric-deficit clinical conditions such as hemineglect?** A class of clinical conditions, of which hemispatial neglect after stroke is the most striking instance, is characterized by a sustained asymmetry in the patient's internal world model. The substrate is, by construction, capable of producing asymmetric force fields across its four actuators, and the patent's Adapted Scenarios construction (PCT/EP2024/076634 p. 18) covers this configuration. A future, distinct line of work would couple this configuration with an asymmetric measurement endpoint to investigate whether closed-loop substrate-level asymmetry can serve as an externalized scaffold for the patient's internal asymmetry. We mention this only as a future research direction, post-dating the present paper, and we note explicitly that the present paper does not claim any clinical efficacy in this or any other indication.

**(Q6) How best to integrate optional visual modalities into the substrate as complementary observation channels?** Earlier sensor-research lines pursued by the present author point at time-of-flight perception (the patent's $z$-variable channel) and 3D range-imaging-based authentication and geometric verification methods as candidate complementary modalities. A video / camera channel — for gaze tracking, attention monitoring, low-back muscle-activity surface estimation, or geometric authentication — would extend the observable surface without altering the substrate's core actuator-driven closed-loop architecture. The open question is what the right integration pattern is: which observations belong on Tier 1 (real-time, actuator-relevant), which on Tier 2 (asynchronous, decision-support), and which off-substrate (pure documentation). The question matters because the answer differentiates the present substrate, where camera-based observation is an add-on, from the existing application-layer products of the neurorehabilitation landscape, where camera-based motion capture is the core sensor.

**(Q7) How should an exemplar-aggregated reference profile $w$ be operationalised on the substrate?** The triple mapping of Table 2 aligns the reference variable $w$ across the patent, the JEPA family and the active-inference frame. The substrate observes the deviation $w - x$ deterministically once $w$ is supplied, but the question of where $w$ comes from is left open. Two sourcing paths are admitted under the patent disclosure (Section 5): an experiential path in which $w$ arises inside the patient's own generative model, and an externally-sourced path in which $w$ is a reference profile constructed from one or more exemplars judged competent in the target task. The open question concerns the externally-sourced path specifically: what minimal schema admits an exemplar-aggregated reference trajectory on the substrate? What is the latency budget within which an externally-sourced $w$ must be available to the control loop? What aggregation mechanism preserves between-exemplar disagreement information that the substrate could subsequently expose? This question is open both as a substrate-engineering question and as a methodological question for the JEPA community, since exemplar-aggregated reference profiles are a structurally distinct class of conditioning signal from the action signals that action-conditioned JEPA variants admit today.

**Closing remark.** This paper is a reflection from a position of engineering practice on a substrate that exists in working hardware. It is not an attempt to extend the predictive-architecture theory itself; that theory belongs to the JEPA program and to its predecessors in joint-embedding, predictive-coding, and active-inference research. The contribution intended here is to describe one substrate that the theory could be exercised on, to make the construct that the substrate realizes explicit (the HCWM), and to invite that exercise.

The intended posture of this paper is enabling rather than proprietary: the substrate, the patent-disclosed architecture, and the open questions of Section 8 are presented in order to open a research space, not to close one. A separate methodological line of work, currently in preparation, will apply a living-case methodology — in the spirit of Stähli's anti-Harvard tradition of real-time, dynamic, longitudinally-accompanied case description — to develop a reference nomenclature for multi-modal closed-loop substrate description. That methodological work will build on existing clinical-data and sensor-hardware standards (DICOM, HL7 FHIR, SNOMED CT, LOINC, IEEE 1451) where applicable and propose complementary terminology where standards are absent. The substrate of the present paper is intended to serve as the living case for that methodological work.

**Differ in kind, not in performance.** We adopt the position, after [5] and [6], and as foregrounded by [18], that human intelligence and the predictive-architecture family "differ in kind, not merely in performance." The substrate-aware route presented here is therefore not an attempt to close that kind-gap; it is an attempt to provide a measurement infrastructure on which alignment — what Riener calls "the standard by which all intelligence is finally judged" — can be operationally negotiated rather than only theoretically asserted.

## Acknowledgments

The author thanks the following individuals and institutions for discussions, technical reflections, or institutional support related to the work reported here. Inclusion in this acknowledgment does not imply that any individual has endorsed, reviewed, or shares authorship of the present manuscript.

This work was developed within the framing of the MIT xPRO Medical Technology Program for Global Leaders (Cohort C1, 2024–2025), under the program leadership of Prof. Michael J. Cima.

- Mario Mauerer and Max Erick Busse-Grawitz (maxon, Sachseln, CH) and Kim Zimmermann (maxon) for the HEJ-90 hardware platform and for technical discussions during onboarding.
- Dr. Thomas Degen (Nanoflex Robotics; formerly Lucerne University of Applied Sciences

and Arts, HSLU) for early support of the craPos drive-unit work — including the InnoCheque that helped spark it — and for ongoing systems-engineering reflection.

- Prof. Dr. Ron Clijsen (SUPSI, Landquart) and Prof. Dr. Roger Abächerli (ETH Zurich and Swiss Academy of Engineering Sciences) as reflectors within the Innosuisse Bridge Discovery consortium framework.
- Dr. Matthias Fenzl, Dr. Beat Villiger, and Dr. Klaus Zahn as longstanding clinical, sports-medicine, and applied-research reflectors on the patient-side and exemplar-side framing of the substrate. Their domain expertise informed the reference-profile sourcing discussion of Sections 3 and 8; the framing taken in this paper is the present author's responsibility.
- CSEM (Centre Suisse d'Electronique et de Microtechnique, Neuchâtel and Alpnach) for engineering-services collaboration on the predecessor Aquabike platform under CTI projects 9680.1 PFNM-NM (2010) and 10554.1 PFLS-LS (2011–2012).
- The pilot participants in the manned and unmanned substrate trials, anonymized by internal pilot protocol.

The substrate's instrumentation philosophy — the $x/w/y/z$ notation of the patent disclosure, the per-cycle telemetry logging convention, the explicit run-time-selectable intervention surface — draws on the building-automation engineering tradition (HLKKS-MSRL) in which the present author holds decades of commissioning experience, alongside biomechanics, rehabilitation medicine, predictive coding and motor-neuroscience as complementary systems-engineering inputs. We acknowledge that multi-disciplinary lineage as one of several methodological sources for the substrate's design, none of which is the substrate's sole ancestor.
The author retains sole responsibility for the content, the empirical results, and any errors in this manuscript.

## References

[1] Irsyad Adam, Zekai Chen, David Laprade, Shaun Porwal, David Laub, Erik Reinertsen, Arda Pekis, and Kevin Brown. The patient is not a moving document: A world model training paradigm for longitudinal EHR. *arXiv preprint*, arXiv:2601.22128, January 2026.

[2] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. *arXiv preprint*, arXiv:2301.08243, 2023.

[3] Adrien Bardes, others, and Yann LeCun. V-jepa: Video joint-embedding predictive architecture. *arXiv preprint*, 2024. CC BY-NC license.

[4] Emily M. Bender and Alexander Koller. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 5185–5198, 2020.

[5] Joseph Chen. On the universal definition of intelligence: A definition for human–AI comparison. *arXiv preprint*, arXiv:2601.07364, 2026. University of Tokyo. Introduces the Extended Predictive Hypothesis (EPH): intelligence as accurate prediction combined with the ability to benefit from that prediction (gainability).

[6] François Chollet. On the measure of intelligence. *arXiv preprint*, arXiv:1911.01547, 2019.

[7] Andy Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

[8] Lorenzo Farina. Artificial intelligence systems, responsibility and agential self-awareness. In *Studies in Applied Philosophy, Epistemology and Rational Ethics*, pages 15–25. Springer International Publishing, 2022.

[9] Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11:127–138, 2010.

[10] James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

[11] Amirabbas Hojjati, Lu Li, Ibrahim Hameed, Anis Yazidi, Pedro G. Lind, and Rabindra Khadka. From video to EEG: Adapting joint embedding predictive architecture to uncover spatiotemporal dynamics in brain signal analysis. *arXiv preprint*, arXiv:2507.03633, July 2025.

[12] Y. Ioannides, others, and Yann LeCun. GMM-Anchored JEPA for speech. *arXiv preprint*, arXiv:2602.09040, February 2026.

[13] Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253, 2017.

[14] Yann LeCun. A path towards autonomous machine intelligence. OpenReview preprint, 2022. Position paper, 27 June 2022.

[15] J. Maes, Yann LeCun, and Randall Balestriero. LeWorldModel: Latent world modeling from video. *arXiv preprint*, arXiv:2603.19312, March 2026.

[16] MindMaze SA. Investor presentation december 2025, 2025. MindMotion GO, Izar, Physilog product family. Vibra US-Stroke pilot, 210 patients.

[17] S. Qu, others, and Yann LeCun. Representation learning for spatiotemporal physical systems. *arXiv preprint*, arXiv:2603.13227, March 2026.

[18] Robert Riener, editor. *AI to Eye — Between Code and Conscience*. vdf Hochschulverlag AG, ETH Zürich, 1 edition, 2026. Essay 14, "After Intelligence: On What Remains Human", is the primary anchor cited in the present paper.

[19] Daniel M. Wolpert and Zoubin Ghahramani. Computational principles of movement neuroscience. *Nature Neuroscience*, 3:1212–1217, 2000.

[20] Daniel M. Wolpert and Mitsuo Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7–8):1317–1329, 1998.

[21] Taizo Yoshikawa. Open-loop torque control on joint position-controlled robots (WO2008131333A1). WIPO PCT application by Honda Motor Co. Ltd., April 2008. Filed 21 April 2008. Patent has lapsed (Ceased, expected expiry 23 October 2009); subject matter is in the public domain. Describes the torque-to-velocity transformer (TVT) underlying joint-position-velocity-torque control modes of the kind exposed by the maxon HEJ firmware as the JPVT mode.

[22] Haoran Zhu, Zhenyuan Dong, Kristi Topollai, Beiyao Sha, and Anna Choromanska. Self-supervised representation learning with joint embedding predictive architecture for automotive LiDAR object detection (AD-L-JEPA). *arXiv preprint*, arXiv:2501.04969, January 2025. Accepted to AAAI 2026.