Lifelong Active Inference of Gait Control

Rudolf Szadkowski and Jan Faigl

***Abstract*—Sustaining the robot's longevity becomes challenging in dynamic deployments characterized by new unknown environments and embodiments outside of the prior knowledge. Hence, the knowledge of robot-environment interactions needs to be continually updated for system adaptation. It can be implemented through self-verification as a continual comparison of predictions with observations using the Predictive Coding (PC) principle. The principle has been further extended into the Active Inference Control (AIC) in biomimetic robotics to drive the control, state estimation, and model update. However, continually updating one model leads to catastrophic forgetting in the long term. Therefore, we propose an autonomously expanding self-verifying world model of sensorimotor dynamics utilized in model-based gait control. The model combines PC with the incremental knowledge representation based on the Internal Model (IM) principle. The proposed method is experimentally validated in virtual and real scenarios, where the hexapod walking robot has to recognize and adapt to leg paralysis and then recognize the recovery. The method generates novel behaviors in real-time, improving the performance and outperforming the examined state-of-the-art methods. Furthermore, the robot's decisions and gained knowledge are interpretable and promise further functional scalability.**

***Index Terms*—lifelong learning, locomotion, internal models, predictive coding**

I. Introduction

Lifelong, sustained locomotion requires continual adaptation to the dynamic context given by the robot's surrounding environment and its body state. The physical properties of the body change, especially in novel elastic and sustainable designs using soft [1]–[3] and degradable materials [4]. The body can also get damaged [5], heated [6], or dirtied that together with terramechanical properties of the terrain [7]–[9] affect the robot-environment interaction during terrain traversal. Hence, the robot experiences multiple dynamic contexts [10] that need to be learned and recognized in real time to produce appropriate locomotion behavior. Although current walking robot controllers achieve impressive performance [11], they often lack the ability to lose confidence in own model and adapt to current dynamic context online. For instance, traditional model-based controllers rely on predefined parameters and struggle with unforeseen environmental or physical changes, while learning-based approaches require extensive retraining or large datasets to adapt to new conditions. The proposed method addresses the limitations of existing controllers by providing a computationally efficient and biologically inspired framework for robust locomotion in complex and uncertain environments. Such capabilities are everyday realities in nature, where animals seamlessly adapt to changing conditions through time-tested neural solutions, motivating the biomimetic principles underlying our approach.

In *Predictive Coding* (PC) principle, the agent continually applies Bayesian inference [12] to form a *belief* characterized by prediction and *certainty* [13]. The quantitative theory of PC is posited by *Free Energy Principle* (FEP) where, by minimizing the free energy, the agent minimizes its uncertainty (surprisal) over time [14]. In robotic deployments, the FEP can be implemented as the *Active Inference* of the state estimate [15] and the *Active Inference Control* (AIC) providing robust motor control [16]–[18]. Under the PC, the controller selects actions reducing the agent's uncertainty. However, the optimal actions depend on sensorimotor interaction unique to dynamic context, and during the robot's lifetime, multiple such contexts need to be considered.

The neural management of multiple dynamic contexts is researched in the *Internal Model* (IM) principle [19]. The principle assumes that the agent's hypothesis about sensori-

TABLE I
Symbols and Terms Used

Term	Description
N	Dimension of the motor commands.
M	Dimension of the sensory modalities.
C	Granularity of the sensorimotor embedding.
$\mathbf{u}=(\mathbf{u}_c)_c^C$	Gait as a sequence of motor commands $\mathbf{u}_c \in \mathbb{R}^N$.
$\mathbf{y}=(\mathbf{y}_c)_c^C$	Sequence of sensory values $\mathbf{y}_c \in \mathbb{R}^M$.
$\mathbf{f}=(\mathbf{f}_c)_c^C$	Internal model of embedded sensorimotor dynamics $f_{mc}(\mathbf{u}) = b_{mc} + \mathbf{w}_{mc}(\mathbf{u}-\mathbf{u}')$ predicting the sensory value of the mth modality at the cth gait phase given the gait $\mathbf{u}$; parametrized by the bias b_{mc}, weight $\mathbf{w}_{mc}$, and point $\mathbf{u}'$.
$\mathscr{F}$	Space of all possible internal models.
W	World model $W=\{\mathbf{f}^i\} \subset \mathscr{F}$.
$\bar{\mathbf{u}}$	Efferent copy of the gait command the robot acted out.
D	Sensorimotor data $D=\{(\bar{\mathbf{u}},\mathbf{y})\}$.
$\mathbf{x}$	Robot's state representation about the world.
$\mathbf{x}^{\text{ref}}$	Reference state desired by the robot.
$p(\mathbf{f}\|D,\mathbf{x}^{\text{ref}})$	IM posterior belief.
$p(\mathbf{x}\|\mathbf{f},D,\mathbf{x}^{\text{ref}})$	State posterior belief.
$p(\mathbf{u}\|\mathbf{x},\mathbf{f},D,\mathbf{x}^{\text{ref}})$	Gait posterior belief.
$q(\mathbf{x})$, $q(\mathbf{u})$	Recognition densities approximating posterior beliefs.
$p(\mathbf{x},D\|\mathbf{f},\mathbf{x}^{\text{ref}})$, $p(\mathbf{u},\mathbf{x}\|D,\mathbf{f},\mathbf{x}^{\text{ref}})$	Generative densities representing the robot's assumption on how the evidence is generated.
F	Free energy $F=\int q \log \frac{q}{p}$ quantifying relation between the recognition density q and generative density p.
$\dot{\mathbf{y}},\dot{\bar{\mathbf{u}}}$	Sensorimotor change derived from the data D.
$\dot{\mathbf{f}}$	Sensorimotor change model derived from the internal model $\dot{f}_{mn}(\dot{\bar{\mathbf{u}}}) = \mathbf{w}_{mn}\dot{\bar{\mathbf{u}}}$.
$\dot{\mathbf{f}}^0(\dot{\bar{\mathbf{u}}})=\mathbf{0}$	Zero-model.
$p(\dot{y}_{mc}\|\dot{\bar{\mathbf{u}}},\dot{\mathbf{f}})$	Model likelihood.
L_{mc}^{ij}	Log-odds between the ith and jth models on the mcth sensory modality-phase $L_{mc}^{ij} = \log \frac{p(\dot{y}_{mc}\|\dot{\bar{\mathbf{u}}},\dot{\mathbf{f}}^i)}{p(\dot{y}_{mc}\|\dot{\bar{\mathbf{u}}},\dot{\mathbf{f}}^j)}$.
$\mathbf{f}^*,\mathbf{x}^*,\mathbf{u}^*$	Selected model, state estimate, and acting gait.
$\mathbf{l},\mathbf{l}^*$	Robot location and goal location on $\mathbb{R}^2$.

The presented work has been supported by the Czech Science Foundation (GAČR) under research project No. 21-33041J.

The authors are with the Czech Technical University, Faculty of Electrical Engineering, Dept. of Computer Science, Technicka 2, 166 27, Prague, Czechia. {szadkrud|faiglj}@fel.cvut.cz

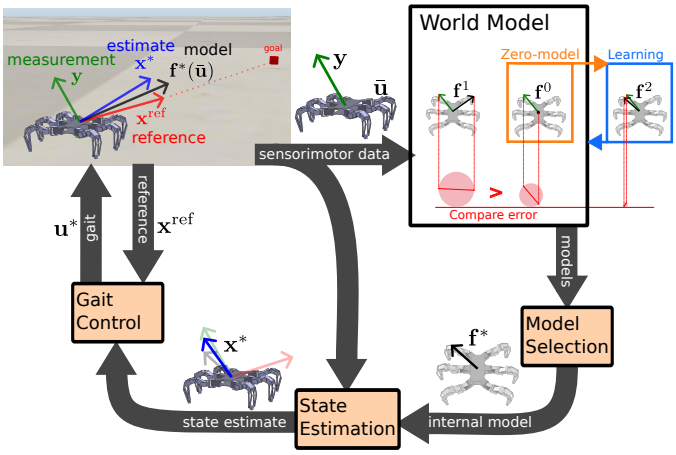


Fig. 1. The proposed continual gait controller architecture for a hexapod walking robot navigating toward a goal location. The robot uses the measured sensorimotor data to update its state estimate and world model. The *world model* is composed of *Internal Models* (IMs) that represent sensorimotor entanglement explanations. If a sensorimotor disentanglement, represented by the zero-model, becomes the best current explanation, a new internal model is learned and added to the world model. The *state estimation* fuses sensory measurements and model prediction into an estimate. The *gait control* is regulated by the difference between the state estimate and reference.

motor interactions is encoded within IM: forward or inverse model [20], [21]. The *Modular Selection and Identification for Control* (MOSAIC) describes *World Model* (WM) as an ensemble of expert IMs, where the control is based on the model that explains the observed data best [22]. The ensemble architecture has been examined in robotic deployments [23], [24] showing flexibility and incremental scalability [25] for new dynamic contexts. The IM ensemble provides a WM architecture that manages multiple contexts, and the PC utilizes the knowledge of the dynamic context. Although both PC and IM principles provide powerful tools for scalable and robust motion control, to the best of the authors' knowledge, there is little research on their relation and combined implementation.

We propose to utilize a model-based controller with a self-improving WM as an expanding IM ensemble. The IM ensemble self-verification and expansion are formulated as a result of Bayesian inference, where an IM represents a hypothesis of sensorimotor interaction with the probability given by sensorimotor evidence. The robot thus selects the most probable hypothesis for motor control. Further, we introduce *zero-model* as a hypothesis of sensorimotor disentanglement representing the robot's uncertainty in control that is resolved by learning a new IM. The proposed method is deployed on the hexapod walking robot, depicted in Fig. 1, in the high-fidelity simulation based on the CoppeliaSim [26], and real environment. The proposed approach is compared with multiple model selection strategies in scenarios requiring learning sensorimotor dynamics, recognizing damage, and utilizing previously gained experience. The contributions of the presented work are considered as follows.

1) We propose a novel combination of an IM ensemble and AIC, enabling real-time lifelong locomotion. It allows the robot to adapt to unforeseen situations using only onboard hardware and collected data.
2) We formalize lifelong locomotion control as active Bayesian inference on linear models, ensuring interpretability [27]. Beyond adaptive behavior, the method provides confidence estimates and generates decisions and insights that are meaningful to human users.
3) We evaluate the lifelong capabilities through comparative and real-world experiments, including scenarios where the hexapod walking robot is partially paralyzed. The comparative study, conducted in high-fidelity simulations, demonstrates that the proposed method produces interpretable results and advances the state of the art. The real-world experiments showcase real-time continual learning, control, and self-verification, all computed using onboard hardware.

The structure of the text is as follows. The next section provides an overview of existing implementations of the AIC and MOSAIC principles. Section III defines the continual learning problem for gait control, formulated as an optimization problem and solved using the proposed method detailed in Section IV. The evaluation setup is outlined in Section V, and the results are presented in Section VI. These results are further discussed in Section VII, followed by concluding remarks in Section VIII. Supplementary material includes Appendix A, which derives the utilized AIC, Appendix B, which details search space reduction and regularization, and Appendix C, which provides additional experimental results. Additionally, the supplementary material features a three-part video documenting (i) hexapod locomotion bootstrapping, (ii) continual adaptation in a simulated environment, and (iii) real-world testing of the proposed method. A summary of the symbols and terms used is provided in Table I.

II. Related Work

The continual motor control is described as an interaction of three processes: (i) model-based estimation, (ii) model-based control, and (iii) model selection. The focus is on the AIC [28] as the solution for model-based estimation and control, while MOSAIC [22] is utilized as a theoretical framework for model selection. The following sections briefly review both approaches and existing results.

A. Locomotion and Lifelong Learning

In a lifelong setup, the robot experiences novel dynamic contexts where the prior knowledge is not enough to sustain the locomotion. The prior knowledge is present in all controllers; the difference is in how general the knowledge is and whether the knowledge can be updated in real time. An example of prior knowledge injection is presented in [11], where authors propose a powerful locomotion controller for a quadruped robot that is trained pre-deployment in simulation using a robot high-fidelity model, designed learning curriculum, and extensive computational resources. The computational resources and model fidelity can be reduced, as shown in [29], by encoding the task and morphology knowledge into biomimetic controller architecture. The compact biomimetic controller can then be tuned during deployment, contrasting the pre-deployment learning in simulation, as shown in [30],

where the robot adapts to rough terrain in real time. While robust and reactive, the aforementioned controllers either do not self-verify or update the prior knowledge without forgetting and thus do not address the lifelong setup.

In lifelong learning and developmental robotics, incremental knowledge scalability can be implemented by a model ensemble. A new model, representing knowledge or skill, is added to the ensemble without losing any previous capabilities, which can make it robust to catastrophic forgetting. The fundamental questions such a method has to solve are: how to (i) combine (or choose) models and (ii) add a new model.

In [31], reinforcement learning is used to train a set of specialized quadruped controllers with respect to (w.r.t.) exteroception that results in robust locomotion in unseen scenarios with externally disrupted walking. Similarly, an exteroception-based combination of reflexes is employed in robust hexapod locomotion [32]. However, the exteroception-based combination requires pre-deployment knowledge of the association between the exteroception and behaviors.

The model combination can be based on tracking the performance of the models. In MOSAIC [22], IMs compete for competency by comparing their prediction errors that are further used for competency predictor training to inform the model selection with the context cues, such as exteroception. Transition probability between the models is introduced in [33] to account for the previous selections in the model selection. MOSAIC was tested on a simulated and real humanoid robot in a sit-to-stand task [24] and heavy object carrying tasks [23], respectively. Hence, MOSAIC shows to be a scalable world model, with interpretable model selection and self-improve selection by exteroceptive cues and previous model transitions.

Besides the model combination, recent studies focus on rules for the addition of a new model. The authors of [25] extend MOSAIC by introducing online IM learning triggered by reference trajectory analysis, where the proposed primitive motion identifier segments reference trajectory and label each either by available IM or new IM.

In [34], the model probability is given by the prediction error and Chinese Restaurant Process (CRP) that distributes prior probability by the model selection frequency, where the learning is triggered when the potential new model is more probable than the previously selected ones. The authors of [35] also utilize the CRP, but instead of using the prediction error, the model likelihood depends on a sensory cluster parameter. Both lifelong learning methods in [34], [35] evaluate the model once per given task; thus, they are *task-driven*. In contrast, a *time-driven* method is presented in [36], where the model learning is triggered when prediction error growth speed crosses a given threshold, which is evaluated every fixed amount of time. The aforementioned approaches focus on learning specific behaviors, such as walking left or right [36] or locomotion at different speed levels [34], [35], where each behavior is associated with a single model. However, the predictive model itself can be used to infer a continuum of behaviors in real-time by using model-based control, such as the AIC.

B. Predictive Coding and Active Inference Control

In the PC paradigm, the nervous system information process is similar to Bayesian infence [12], which is modeled as the free energy minimization [14]. In visual data processing, the PC is implemented as a multilayered predictive model [13] utilized for sensory fusion and prediction. Sound source separation from the mixed waveform is proposed in [37], where the multilayered predictive model isolates the sound component based on a video of a musical instrument being played. In [38], the multilayered predictive model is extended with a recurrent neural network, yielding superior results in predicting future video frames.

In robotics, the PC is implemented as the AIC that continually updates the motor control and state estimation by minimizing the free energy [28]. AIC deployments are reported mainly for robotic manipulators. In [39], the *Model Reference Adaptive Control* (MRAC) is compared with the AIC in torque control of the manipulator with 7 *Degrees of Freedom* (DoF). For both model-based controllers, a model is trained in a simulator that is then used on the real manipulator with additionally modified morphology to make the model imperfect. While the MRAC is unable to cope with imperfect models, the AIC robustly performs given tasks. The update rules can be extended for variance parameters resulting in adaptive dampening reference-error oscillations as observed with the real 7-DoF manipulators [17]. The AIC can be further extended by multi-sensor fusion [16] that outperforms existing approaches based on model predictive control and impedance control in scenarios with damaged 7-DoF manipulator.

Based on the literature review, the AIC provides robust and adaptive control for manipulator robots. Therefore, we propose to expand the AIC to control walking robots. In particular, we focus on gait dynamics that introduce challenges to hysteresis and phase-dependent behavior. The gait dynamics can be modeled by the phase-embedded forward model, where the phase is estimated by the central pattern generator [40]. To the best of the authors' knowledge, the proposed approach is the first model-based controller utilizing a self-improving world model of gait dynamics.

III. Gait Control Problem Specification

Gait is a repetitive interaction of the robot with the environment, which can be adjusted w.r.t. the given reference state and the sensorimotor measurements. The gait adjustment performance is quantified by the *performance error*, as a difference between the reference and true states, which improvement can be derived from the sensorimotor interaction model. However, neither the true state nor sensorimotor interaction model is given, and the robot must infer Bayesian belief about the state and model from observations.

Let denote the *gait* be a sequence of C motor commands $\mathbf{u} = (\mathbf{u}_c)_c^C$ and the posterior $p(\mathbf{u}|D, \mathbf{x}^{\text{ref}})$ represents the robot's belief that a gait $\mathbf{u}$ results in a state close to the reference $\mathbf{x}^{\text{ref}}$ given the sensorimotor data D. From PC and IM principles, we can assume the robot has an internal belief about the state

$\mathbf{x}$ and sensorimotor interactions $\mathbf{f}$, from which the robot infers the gait belief as

$$p(\mathbf{u}|D,\mathbf{x}^{\text{ref}}) = \sum_{\mathbf{f}}\sum_{\mathbf{x}} p(\mathbf{u},\mathbf{x},\mathbf{f}|D,\mathbf{x}^{\text{ref}}), \quad (1)$$

$$= \sum_{\mathbf{f}}\sum_{\mathbf{x}} p(\mathbf{u}|\mathbf{x},\mathbf{f},D,\mathbf{x}^{\text{ref}})p(\mathbf{x}|\mathbf{f},D,\mathbf{x}^{\text{ref}})p(\mathbf{f}|D,\mathbf{x}^{\text{ref}}). \quad (2)$$

The proposed method is to approximate the three posterior beliefs with the minimal statistics computed by the gait control, state estimation, and model selection, respectively, see Fig. 1. The phase-embedded forward model $\mathbf{f}$ used for the IM implementation is briefly introduced in the following section.

A. Forward Model of Gait Dynamics

For repetitive systems, such as the gait dynamics, it is practical to represent the state trajectory segment, which is delimited by the period of repetition, as a point [41]. In [42], the phase-embedding transforms the sensory and motor signals into embedded space, where a linear relation between the motor and sensory embeddings is found. The phase-embedded model encodes the non-linear, phase-dependent, and time-delayed sensorimotor interactions into a set of regressors. Therefore, the model provides an interpretable [27] and computationally efficient forward model of gait dynamics.

The model is a function in phase-embedded sensorimotor space $\mathbf{f} \in \mathscr{F} : \mathbb{R}^{NC} \to \mathbb{R}^{MC}$, with N motors, M sensory modalities, and embedding *granularity* C. The granularity defines a segmentation of the gait period into C segments $[0,1) = \cup_c^C [\frac{c-1}{C}, \frac{c}{C})$, where c is the gait phase index. The *gait* is represented as a sequence of C motor commands $\mathbf{u} = (\mathbf{u}_c)_c^C \in \mathbb{R}^{NC}$. Likewise, the *sensory embedding* is a sensory value sequence $\mathbf{y} = (\mathbf{y}_c)_c^C \in \mathbb{R}^{MC}$. Each sensory modality m at the phase c denoted y_{mc} is predicted by the gait regression $f_{mc}(\mathbf{u})$.

In the present work, the regressor is linear $f_{mc}(\mathbf{u}) = b_{mc} + \mathbf{w}_{mc}(\mathbf{u} - \mathbf{u}')$ with the *bias* b_{mc} and *weight* vector $\mathbf{w}_{mc} \in \mathbb{R}^{MC}$, and *point* $\mathbf{u}'$. The partial derivative $\frac{\partial f_{md}}{\partial u_{nc}}$ is scalar w_{mc}^{nd} describing relation between the nth motor command at the phase d and the mth sensory modality at the phase c. A detailed description of the embedding process follows.

Phase-embedding transforms a trajectory segment of a variable $\alpha \in \mathbb{R}^M$ into a point $\mathbf{a} \in \mathbb{R}^{MC}$ of the phase-space with the granularity C. The embedding takes C equidistant samples from the segment $A(t) = (\alpha(\tau))_t^{t-T}$ of a single period T. The sampling is driven by the gait phase $\phi(t) = 2\pi t T^{-1} \bmod 2\pi$ that, during its evolution, sequentially selects the corresponding segment index $c(t)$ [29]:

$$c(t) = \arg\min_{c'}(|\exp\{i\phi(t)\} - \exp\left\{i\frac{c'+0.5}{C}2\pi\right\}|), \quad (3)$$

$$\dot{\mathbf{a}}_{c'}(t) = \begin{cases} \alpha(t) - \mathbf{a}_{c'}(t) & \texttt{if } c' = c(t), \\ 0 & \texttt{otherwise}, \end{cases} \quad (4)$$

with the imaginary unit i and norm $|\cdot|$.

The embedding update rule (4) is evaluated concurrently with the gait control. The embedding $\mathbf{a}(t) = (\mathbf{a}_c)_c^C$ thus represents the variable trajectory segment $A(t)$. The embedding is applied to motor and sensory signals, resulting in embeddings $\mathbf{u}$ and $\mathbf{y}$, respectively.

IV. Proposed Lifelong Active Inference

The proposed method is based on the premise that for a given reference and sensorimotor data, the robot infers an appropriate gait from continually updated beliefs about the state and IM. The state estimate fuses the sensory measurements with IM predictions. The robot selects IM, a hypothesis about sensorimotor interaction that is supported by the measured sensorimotor data. Following the IM principle [22], the IM is selected from a WM, a finite IM ensemble. However, if the robot encounters an unknown dynamic context, it should select from beyond the WM and thus learn a new IM.

We propose to learn a new IM when the robot believes its actions have no sensory effect and reformulate the gait belief (2) to

$$p(\mathbf{u}|D,\mathbf{x}^{\text{ref}}) = \sum_{\mathbf{f}} p(\mathbf{f}|D,\mathbf{x}^{\text{ref}}) \sum_{\mathbf{x}} p(\mathbf{u},\mathbf{x}|\mathbf{f},D,\mathbf{x}^{\text{ref}}), \quad (5)$$

where the inner sum is approximated by the AIC, and the outer sum is approximated by the MOSAIC approach. The approximations give us the update rules for the *selected model* $\mathbf{f}^*$(22), *state estimate* $\mathbf{x}^*$(12), and *acting gait* $\mathbf{u}^*$(13). The belief update is visualized in Fig. 1. The proposed AIC and model selection are detailed in the rest of the section.

A. Active Inference of the Gait Control

The state and gait are continually inferred from the given sensorimotor data D, IM $\mathbf{f}$, and reference $\mathbf{x}^{\text{ref}}$. The inference is a two-stage optimization process, where the state estimate is inferred from data before inferring the acting gait from the state estimate. The optimization is approximation of posterior beliefs $p(\mathbf{u},\mathbf{x}|\mathbf{f},D,\mathbf{x}^{\text{ref}}) = p(\mathbf{x}|\mathbf{f},D,\mathbf{x}^{\text{ref}})p(\mathbf{u}|\mathbf{x},\mathbf{f},D,\mathbf{x}^{\text{ref}})$ with *recognition densities*

$$q(\mathbf{x};\mathbf{x}^*,\zeta^x) = \mathscr{N}(\mathbf{x};\mathbf{x}^*,\zeta^x), \quad (6)$$

$$q(\mathbf{u};\mathbf{u}^*,\zeta^u) = \mathscr{N}(\mathbf{u};\mathbf{u}^*,\zeta^u), \quad (7)$$

by finding means $\mathbf{x}^*, \mathbf{u}^*$, and variances ζ^x, ζ^u of the respective normal distributions. A relation between recognition density and posterior is quantified by Kullback-Lieber (KL) divergence

$$D_{\text{KL}}[q(\mathbf{x})||p] = \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x}|\mathbf{f},D,\mathbf{x}^{\text{ref}})}\, d\mathbf{x}, \quad (8)$$

$$D_{\text{KL}}[q(\mathbf{u})||p] = \int q(\mathbf{u}) \log \frac{q(\mathbf{u})}{p(\mathbf{u}|\mathbf{x},\mathbf{f},D,\mathbf{x}^{\text{ref}})}\, d\mathbf{u}. \quad (9)$$

Since the numerical optimization of both expressions is computationally intractable, we recall FEP [14] and use the relation between the KL divergence and free energy to express

$$D_{\text{KL}}[q(\mathbf{x})||p] = F^x - \log p(D|\mathbf{f},\mathbf{x}^{\text{ref}}), \quad (10)$$

$$D_{\text{KL}}[q(\mathbf{u})||p] = F^u - \log p(\mathbf{x}|D,\mathbf{f},\mathbf{x}^{\text{ref}}), \quad (11)$$

where $-\log p(D|\mathbf{f},\mathbf{x}^{\text{ref}})$ and $-\log p(\mathbf{x}|D,\mathbf{f},\mathbf{x}^{\text{ref}})$ refer to *surprisal* for which the corresponding free energy F^x and F^u is the upper bound. The minimization of free energy implicitly minimizes the surprisal upper bound and approximates the recognition densities to posterior beliefs.

By using the analysis developed in [28], we can derive differentiable function forms of F^x and F^y. The free energy

is then optimized by finding the state estimate $\mathbf{x}^*$ and acting gait $\mathbf{u}^*$ with the gradient descend

$$\frac{\mathrm{d}x^*_{mc}}{\mathrm{d}t} = \frac{f_{mc}(\bar{\mathbf{u}}) - x^*_{mc}}{\sigma^{f|x}_{mc}} + \frac{y_{mc} - x^*_{mc}}{\sigma^{y|x}_{mc}} - \frac{x^*_{mc} - b_{mc}}{\sigma^{x}_{mc}}, \tag{12}$$

$$\frac{\mathrm{d}u^*_{nc}}{\mathrm{d}t} = \sum_{md}^{MC} \frac{x^{\mathrm{ref}}_{md} - x^*_{md}}{\sigma^{x|u}_{md}} w^{nc}_{md} - \frac{u^*_{nc} - u'_{nc}}{\sigma^{u}_{nc}}, \tag{13}$$

where $\sigma^{f|x}_{mc}$, $\sigma^{y|x}_{mc}$, and σ^{x}_{mc} are standard deviations scaling the respective contribution of the prediction observation and sensory bias in the state estimation. The acting gait update is proportional to the *estimated performance error* $\mathbf{x}^{\mathrm{ref}} - \mathbf{x}^*$ combined by the sensorimotor weight $\mathbf{w}^{md}$ and regularized toward the point $\mathbf{u}'$. The standard deviation $\sigma^{x|u}_{md}$ is a hyperparameter scaling the performance error sensitivity.

The derivations of the update rules (12) and (13) are detailed in Appendix A of the supplementary material to keep the text concise. Further, the gait-space reduction to harmonic motion and regularizing gait toward symmetry in the robot's lateral movement is in Appendix B of the supplementary material.

B. Model Selection

Based on a given sensorimotor evidence, the model that is best supported by the evidence is selected. Thus, the sum (2) is relaxed to the *model selection*

$$\mathbf{f}^* = \arg\max_{\mathbf{f}\in\mathscr{F}} p(\mathbf{f}|D, \mathbf{x}^{\mathrm{ref}}), \tag{14}$$

where only the most plausible model contributes to (2). The probability of the models $\mathbf{f}$ and $\mathbf{f}'$ is compared by log-odds

$$\log \frac{p(\mathbf{f}|D, \mathbf{x}^{\mathrm{ref}})}{p(\mathbf{f}'|D, \mathbf{x}^{\mathrm{ref}})} = \log \frac{p(D, \mathbf{x}^{\mathrm{ref}}|\mathbf{f})p(\mathbf{f})}{p(D, \mathbf{x}^{\mathrm{ref}}|\mathbf{f}')p(\mathbf{f}')}, \tag{15}$$

where the model posterior is expanded by Bayes' rule and factor out $p(D, \mathbf{x}^{\mathrm{ref}})$. The plausibility of $\mathbf{f}$ over $\mathbf{f}'$ is then determined by its likelihood and prior $p(\mathbf{f})$.

Following the IM principle, we assume the prior is much higher for IMs from *world model* $W = \{\mathbf{f}^i\}$, than for *unknown models* $\mathscr{F} - W$. Since prior always prefers the IM, two pairwise comparisons must be resolved by two types of likelihood $p(\mathbf{f}|D, \mathbf{x}^{\mathrm{ref}})$ comparisons. It is a comparison between IMs and a comparison between unknown models, where we refer to the former as *model competition*, and the latter as *model learning*. The robot decides to select the model by learning when it is uncertain about its action.

Let *zero-model* $\mathbf{f}^0 \in W$ represent the hypothesis that the action has no sensory consequence. If the robot believes that an action has no sensory consequences, it implies that the robot has no preference over actions, resulting in a uniform distribution over actions: $p(\mathbf{u}|\mathbf{x}, \mathbf{f}^0, D, \mathbf{x}^{\mathrm{ref}}) = \mathit{uniform}$. Consequently, as the zero-model becomes more plausible, the sum in (2) is increasingly dominated by that uniform distribution over all possible gaits. The dominance can be interpreted as the robot becoming progressively more uncertain about its next action. We assume that action uncertainty can be mitigated by learning a better model of action-sensor entanglement. Therefore, the robot decides to learn a new model when the zero-model becomes more plausible than the IMs from the world model.

We implement the decision process by selecting the IM with the highest likelihood and comparing it with the zero-model. Let $\dot{\mathbf{y}}$ and $\dot{\mathbf{u}}$ represent the last *sensorimotor change* extracted from the data D. Let $\dot{\mathbf{f}} = \{\dot{f}_{mn}\}^{MN}_{mn}; \dot{f}_{mn}(\dot{\mathbf{u}}) = \mathbf{w}_{mn}\dot{\mathbf{u}}$ be the *change model* derived from the model $\mathbf{f}$ and define zero-model as $\dot{\mathbf{f}}^0(\dot{\mathbf{u}}) = \mathbf{0}\dot{\mathbf{u}}$. We reformulate the model likelihood $p(D, \mathbf{x}^{\mathrm{ref}}|\mathbf{f}^i)$ into

$$p(\mathbf{x}^{\mathrm{ref}}|\dot{\mathbf{y}}, \dot{\mathbf{u}}, \dot{\mathbf{f}}^i)p(\dot{\mathbf{y}}|\dot{\mathbf{u}}, \dot{\mathbf{f}}^i)p(\dot{\mathbf{u}}|\dot{\mathbf{f}}^i), \tag{16}$$

showing that the model likelihood depends on how well it fits to reference, sensory data given motor data, and motor data themselves. In the present work, we utilize only the second term, probability of sensory observations, given the IM and gaits. For the other terms, we assume the uniform distribution.

We assume the likelihood of normal distribution and independence between sensory modalities [34]

$$p(\dot{y}_{mc}|\dot{\mathbf{u}}, \dot{\mathbf{f}}^i) = \mathscr{N}(\dot{y}_{mc}; \dot{f}^i_{mc}(\dot{\mathbf{u}}), \sigma^i_{mc}). \tag{17}$$

The ith model is compared to the jth model by log-odds

$$L^{ij}_{mc} = \log \frac{p(\dot{y}_{mc}|\dot{\mathbf{u}}, \dot{\mathbf{f}}^i)}{p(\dot{y}_{mc}|\dot{\mathbf{u}}, \dot{\mathbf{f}}^j)}, \tag{18}$$

where $L^{ij}_{mc} > 0$ holds if the ith model is more plausible explanation than the jth model for the measured mth sensory modality at the phase c. For $i > 0$, we select the most likely IM as

$$i^* = \arg\max_i \sum_j \sum_{mc}^{MC} L^{ij}_{mc} \tag{19}$$

$$= \arg\min_i \sum_{mc}^{MC} \left[\log \sigma^i_{mc} + \left(\frac{\dot{y}_{mc} - \dot{f}^i_{mc}(\dot{\mathbf{u}})}{\sigma^i_{mc}} \right)^2 \right]. \tag{20}$$

The most likely IM i^* is then compared with the zero-model $\dot{\mathbf{f}}^0(\dot{\mathbf{u}})$. If $\sum_{mc} L^{0,i^*}_{mc}$ exceeds a given tolerance *threshold* θ, the robot decides to learn a new model. The model selection (14) is thus implemented as a two-step comparison: (i) model competition, where we select the most likely IM from W, and (ii) zero-model comparison, where the most likely IM is compared with the zero-model using

$$i^* = \arg\min_i \sum_{mc}^{MC} \left[\log \sigma^i_{mc} + \left(\frac{\dot{y}_{mc} - \dot{f}^i_{mc}(\dot{\mathbf{u}})}{\sigma^i_{mc}} \right)^2 \right], \tag{21}$$

$$\mathbf{f}^* = \begin{cases} \arg\max_{\mathbf{f}\in\mathscr{F}} p(D, \mathbf{x}^{\mathrm{ref}}|\mathbf{f}) & \text{if } \sum_{mc} L^{0,i^*}_{mc} > \theta \\ \mathbf{f}^{i^*} & \text{otherwise} \end{cases}. \tag{22}$$

The model learning $\arg\max_{\mathbf{f}\in\mathscr{F}} p(D, \mathbf{x}^{\mathrm{ref}}|\mathbf{f})$ is performed by linear regression over sensorimotor data D gathered during motor babbling.

C. Integrated Algorithm

One time step of the proposed method is visualized in Fig. 2 and summarized in Algorithm 1 that is called every $\Delta_t = 0.01\,\mathrm{s}$ with the sensory reading $\gamma(t)$, efferent (motor) copy $\bar{\nu}(t)$, and reference $\mathbf{x}^{\mathrm{ref}}(t)$. The algorithm output $\nu(t)$ is then sent to the robot's servos and used as the efferent copy in the next step $\bar{\nu}(t + \Delta_t) = \nu(t)$.

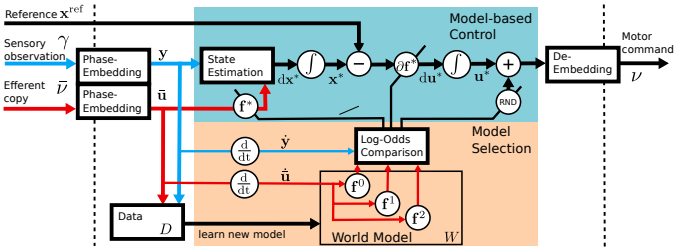


Fig. 2. The data flow of the proposed method. The $\int$ and $\frac{\mathrm{d}}{\mathrm{dt}}$ nodes represent numerical integration and derivation, respectively. The RND node generates the random noise added to the acting gait that is increased during the learning stage, producing the motor babbling.

Algorithm 1 Lifelong Active Inference of Gait Control

1: **Input:** Reference $\mathbf{x}^{\text{ref}}$, observation γ, efferent copy $\bar{\nu}$.
2: **Output:** Motor command ν.
3: $\phi \leftarrow \phi + 2\pi T^{-1}\Delta_t$
4: $c \leftarrow$ by (3)
5: $\mathbf{y}_c^{\text{mem}} \leftarrow \mathbf{y}_c^{\text{mem}} + (\gamma - \mathbf{y}_c^{\text{mem}})\Delta_t$, see (4).
6: $\mathbf{u}_c^{\text{mem}} \leftarrow \mathbf{u}_c^{\text{mem}} + (\bar{\nu} - \mathbf{u}_c^{\text{mem}})\Delta_t$, see (4).
7: **if** $c \neq c'$ **then** ▷ *Phase segment switched.*
8: $c' \leftarrow c$ and $t_{\text{learn}} \leftarrow t_{\text{learn}} - 1$.
9: Add $(\mathbf{y}^{\text{mem}}, \mathbf{u}^{\text{mem}})$ to dataset D.
10: **if** $t_{\text{learn}} < 0$ **then** ▷ *Is in the performing stage.*
11: Compute i^* by (21) from W.
12: Compute $L = \sum_{mc} L_{mc}^{0,i^*}$.
13: **if** $\sum_{mc} L_{mc}^{0,i^*} > \theta$ **then** ▷ *Starts the learning stage.*
14: $t_{\text{learn}} \leftarrow T_{\text{learn}}$ and clear dataset D.
15: $\mathbf{f}^* \leftarrow \mathbf{f}^0$ and $\sigma \leftarrow \sigma_{\text{learning}}$.
16: **else** ▷ *Continues the performing stage.*
17: $\mathbf{f}^* \leftarrow \mathbf{f}^{i^*}$ and $\sigma \leftarrow \sigma_{\text{performing}}$.
18: **end if**
19: **else if** $t_{\text{learn}} = 0$ **then** ▷ *The learning stage ended.*
20: Learn $\mathbf{f}'$ by linear regression over D.
21: $\mathbf{f}^* \leftarrow \mathbf{f}'$ and add $\mathbf{f}'$ to W.
22: **end if**
23: Get $\Delta\mathbf{x}^*$ and $\Delta\mathbf{u}^*$ from (12) and (13).
24: $\mathbf{x}^* \leftarrow \mathbf{x}^* + \Delta\mathbf{x}^*$ and $\mathbf{u}^* \leftarrow \mathbf{u}^* + \Delta\mathbf{u}^*$.
25: $\nu \leftarrow \mathbf{u}_c^* + \mathcal{N}(\mathbf{0}, \sigma)$.
26: **end if**
27: **return** ν

At Lines 3–6, the phase embedding is performed that projects the time to the gait phase c. For each gait phase change, the sensorimotor embedding is added to the dataset D at Line 9, which is implemented as a circular queue, i.e., if the queue is full, the new input overwrites the oldest. Besides IM learning, the dataset is used for statistics extraction and numerical derivation. At Line 10, we can distinguish two lifecycle stages: (i) performing ($t_{\text{learn}} < 0$) and (ii) learning ($t_{\text{learn}} \geq 0$).

In the performing stage, the IMs compete at Lines 11–13; if the best IM is better than the zero-model, the best IM is selected at Line 17. The selected model is then utilized in model-based estimation and control performed at Lines 23 and 24. The gait command $\mathbf{u}^*$ is then de-embedded at Line 25, where it is added random noise drawn from the normal distribution with standard deviation $\sigma = \sigma_{\text{performing}}$. It has been empirically observed that adding the noise $\sigma_{\text{performing}} = 0.01$ improves the IM prediction error evaluation with a nominal effect on the motion properties.

The learning stage is initiated, if the best model is worse than the zero-model, see Line 13. At Line 14, the data collection counter t_{learn} is set to $T_{\text{learn}} = 1200$, which corresponds to five minutes. During the motor babbling, the standard deviation σ is increased to $\sigma_{\text{learning}} = 0.2$ at Line 15 and the zero-model is selected. By definition, the zero-model weights are zero, $\mathbf{w} = \mathbf{0}$, thus the gait command stops updating, $\Delta\mathbf{u}^* = \mathbf{0}$, at Line 24. The only change is induced by random noise at Line 25, which creates sensorimotor observation stored in the dataset at Line 10. The learning stage ends at Lines 19–21, the new model is extracted using linear regression over the dataset, which takes less than one second, and the new model is added to the world model.

V. Evaluation Setup and Performance Measures

The proposed method is implemented[1] and tested on the high-fidelity model of the hexapod walking robot [43] in the CoppeliaSim [26] and demonstrated on a real robot, depicted in Fig. 3, respectively.

(a) Simulated hexapod (b) Real hexapod (c) Leg schema

Fig. 3. (a) The simulated hexapod walking robot with labeled legs. During paralysis, up to two selected legs stop moving. (b) The real hexapod walking robot, HEBI Daisy, with mounted Intel® RealSense™ Tracking Camera T265. (c) Three leg joints, body-coxa, coxa-femur, and femur-tibia, are commanded for the respective angles.

While simulated and real hexapod walking robots have different inertia and kinematics, their sensorimotor modalities are the same in dimensions and semantics. The robot has six legs, each actuated by three servomotors connecting the links on body-coxa, coxa-femur, and femur-tibia joints, see Fig. 3c. The robot learns to locomote its body by controlling the angle of each joint; hence, $N = 18$. The robot's sensory input consists of heading and side velocities y_{head} and y_{side}, change of roll y_{roll}, pitch y_{pitch}, and yaw y_{yaw}, and joint torques. Thus, the overall number of sensory modalities is $M = 23$.

The sensory data are sampled at 10 Hz and 100 Hz for the simulated and real robot, respectively, which is a limiting factor for the gait period and its granularity. A shorter gait period accelerates model learning, as more gait cycles can be aggregated within a given time frame. However, it also reduces the number of sensory samples collected per gait,

[1] The sources are available at https://github.com/comrob/laicg.

making the algorithm more susceptible to outliers in the data. Similarly, a higher granularity results in a smoother motion but divides the sensory samples within the gait cycle. The gait period is set to $T = 1.04\,\mathrm{s}$, which is segmented with the granularity $C = 4$ for the simulated robot, resulting in $10T/C \approx 2$ samples per segment. Despite the low sample rate, it showed adequate performance for small robots operating in simulated environments. In contrast, the real robot uses a granularity of $C = 6$, resulting in $100T/C \approx 17$ samples per segment.

The evaluation scenario is robot navigation with temporarily paralyzed robot leg(s) that is designed to evaluate anomaly detection in locomotion and learning compensation for the paralysis. The robot should also recognize recovery when the paralysis is passed. The robot navigates toward a given goal location by setting heading velocity and yaw change references. The reference signals depend on the robot's position $\mathbf{l}$, roll α, pitch β, yaw γ, and the goal location $\mathbf{l}^*$,

$$x^{\mathrm{ref}}_{\mathrm{head}}(t) = \max\{\|\mathbf{l}^{\Delta}\|\cos\left(\gamma^{\Delta}\right), 0\}, \tag{23}$$

$$x^{\mathrm{ref}}_{\mathrm{yaw}}(t) = \gamma^{\Delta} \mod 2\pi, \tag{24}$$

$$x^{\mathrm{ref}}_{\mathrm{side}}(t) = x^{\mathrm{ref}}_{\mathrm{roll}}(t) = x^{\mathrm{ref}}_{\mathrm{pitch}}(t) = 0, \tag{25}$$

where for $\mathbf{l}^{\Delta} = \mathbf{l}^* - \mathbf{l}$, $\gamma^* = \arctan2(\mathbf{l}^{\Delta}_2, \mathbf{l}^{\Delta}_1)$, and $\gamma^{\Delta} = \gamma^* - \gamma$. Besides, the reference velocity and heading values are limited to be within the intervals $[0\,\mathrm{m\,s^{-1}}, 0.035\,\mathrm{m\,s^{-1}}]$ and $[-0.12\,\mathrm{rad\,s^{-1}}, 0.12\,\mathrm{rad\,s^{-1}}]$, respectively. In simulated scenarios, the goal location is set to $\mathbf{l}^* = (-100\,\mathrm{m}, -100\,\mathrm{m})$.

Every scenario run is initialized by pre-learned WM $W = \mathbf{f}^{\mathrm{l}}$ obtained by repeating the performing and learning stage until the (non-paralyzed) robot is capable of walking forward and turning; see Part 1 of the video in supplementary material. For clarity, we refer to $\mathbf{f}^{\mathrm{l}}$ as *Walking IM*. The scenario run is 3500 simulated seconds long, which corresponds to about one hour. At the 10 % of the scenario progress, $t_1 = 350\,\mathrm{s}$, the robot's legs are paralyzed. The robot recovers at the 70 % progress, $t_2 = 2450\,\mathrm{s}$. Four particular scenarios, each with five trials, are performed with different selections of the paralyzed legs: L2, L1R2, L1R3, and L2R3 according to labels in Fig. 3a.

Reference model selection methods include *Baseline* without any continual model learning and competitive world model learning strategies in [34]–[36], [42]. The methods are compared with the proposed prediction-based model selection (21), (22) with the focus on the model selection. Thus, all the evaluated continual controllers use the same implementation of the gait controller (13) but have different implementations of the model selection (14).

- **Scheduled** [42] adds one model per thousand seconds. The method assesses the bootstrapping approach with the reactive prediction-based model selection methods.
- **CRPpredictive** [34] selects the best model by model likelihood and prior probability based on the Chinese Restaurant Process (CRP). The CRP prior distributes more probability to the models that were selected more often in the past:

$$i^* = \arg\max_i \log \mathcal{N}(\mathbf{y}; \mathbf{f}^i(\mathbf{u}), \sigma^2) + \log K^i(t), \tag{26}$$

where $K^i(t)$ is the number of gaits the model i was selected until t, and constant deviation $\sigma = 0.1$. The best model is then compared with potential new model probability: $p(\mathbf{y}|\mathbf{u}, \mathbf{f}^+)\theta_{\mathrm{CRPpred}}$, where $\theta_{\mathrm{CRPpred}}$ is a concentration parameter. Assuming the new potential model always perfectly fits the data, $p(\mathbf{y}|\mathbf{u}, \mathbf{f}^+) = 1$, the robot decides to learn when the negative log-likelihood (NLL) of the best model is higher than the potential new model:

$$-\log \mathcal{N}(\mathbf{y}; \mathbf{f}^{i^*}(\mathbf{u}), \sigma^2) - \log K^{i^*}(t) > -\log \theta_{\mathrm{CRPpred}}. \tag{27}$$

Moreover, each ith model has initialized $K^i = 300$ corresponding to the number of gait cycles used for training one model.

- **CRPcluster** [35] differs from CRPpredictive by calculating the model likelihood as the distance between observed sensory data and dynamically updated cluster centroids ϑ^i: $i^* = \arg\max_i -\|\mathbf{y} - \vartheta^i\|^2 + \log K^i(t)$. Each selection, the centroids ϑ^i are updated by weighted gradient with the learning rate $\eta = 0.01$ as

$$\Delta\vartheta^i = q(i)2(\mathbf{y} - \vartheta^i)\eta, \tag{28}$$

where the weight $q(i)$ is calculated as the softmax of the model log-likelihoods. The centroids are initialized with the corresponding model bias $\vartheta^i = \mathbf{b}^i$. Similarly to (27), the robot decides to learn when

$$\|\mathbf{y} - \vartheta^{i^*}\|^2 - \log K^{i^*}(t) > -\log \theta_{\mathrm{CRPclus}}. \tag{29}$$

- **L2norm** [36] is an adapted prediction-based model selection used in *Surprise-based Behavioral Modularization into Event-Predictive Structures* (SUBMODES). The model selection tracks the Euclidean norm of the prediction error, $\|\mathbf{y} - \mathbf{f}^i(\mathbf{u})\|$, standardized by the error norm mean μ^i and deviation σ^i measured during the ith model learning and compared to the threshold θ_{L2norm}. When the threshold is exceeded, as defined by:

$$\frac{\|\mathbf{y} - \mathbf{f}^i(\mathbf{u})\| - \mu^i}{\sigma^i} > \theta_{\mathrm{L2norm}}, \tag{30}$$

time-limited competition intervals are initiated among the models. If the best predicting model has its standardized error less than θ_{L2norm}, the model is selected; otherwise, after the competition, a new model is trained. In the preliminary study, a straightforward implementation of the method consistently underperformed. Therefore, we introduced a mechanism to track whether the ground truth values change in the same direction as the predicted values to address the issue. The introduced directionality is quantified using an element-wise transformation:

$$\mathrm{sign}^{\varepsilon}(x) = \begin{cases} -1 & \texttt{if } x < -\varepsilon, \\ 1 & \texttt{if } x > \varepsilon, \\ 0 & \texttt{otherwise}, \end{cases} \tag{31}$$

where $\varepsilon = 0.01$ partitions the real values into intervals and maps them to discrete values of -1, 1, or 0. The

transformation is applied to the sensorimotor *change* errors, $\dot{\mathbf{y}} - \dot{\mathbf{f}}^i(\dot{\mathbf{u}})$, as

$$\frac{\|\text{sign}^\varepsilon(\dot{\mathbf{y}}) - \text{sign}^\varepsilon(\dot{\mathbf{f}}^i(\dot{\mathbf{u}}))\| - \mu^i}{\sigma^i} > \theta_{\text{L2norm}}. \qquad (32)$$

With the modification, the L2norm algorithm achieves performance competitive with the proposed method.

Method sensitivity tuning of the L2norm, CRPpredictive, CRPcluster, and proposed model-selection algorithms have been performed to ensure the fair comparison in the simulated environment. In particular, the threshold parameters θ (22), θ_{CRPpred} (27), θ_{CRPclus} (29), and θ_{L2norm} (30) determine the sensitivity of learning decision. The thresholds were calibrated by the same methodology using three navigation scenarios without paralysis and measuring the decision signal (e.g., zero-model log-odds of the proposed or standardized error of L2norm) and setting threshold as its boundary. The found values are $\theta = -0.022$, $\log\theta_{\text{CRPpred}} = -1.7$, $\log\theta_{\text{CRPclus}} = 5.1$, and $\theta_{\text{L2norm}} = 0.012$. All tested methods trigger the learning after exceeding the threshold for five consequent gait cycles. The calibration ensures that learning is not triggered without paralysis in either of the model selection methods. The threshold values of each algorithm are set to be the same for all scenarios and trials.

Evaluation metrics consist of three metrics derived from the true performance as the *Mean Absolute Error* (MAE) $e(t) = (MC)^{-1}|\mathbf{x}^{\text{ref}} - \mathbf{y}|$, and goal distance. From the performance error, metrics are calculated from the MAE: (i) floating average MAE of the last 25 s: $E_{\text{avg}}(t) = 25^{-1}\sum_{\tau=t-25}^{\tau=t} e(\tau)$ and (ii) cumulative MAE $E_{\text{cum}}(t) = \sum_{\tau=0}^{\tau=t} e(\tau)$. The floating MAE characterizes the robot's immediate performance and is usable in analyzing the robot's progress. The cumulative MAE and goal distance metric measure the consequences of the robot's overall behavior.

VI. Results

In the evaluation scenarios, the proposed and compared methods locomote the robot toward the goal location, detect and compensate for the leg paralysis, and recognize the leg recovery. All six methods have been evaluated in four different paralysis scenarios with five trials. The overall results show that the proposed model has the best average ranking in cumulative MAE and final goal distance, see Table II. Notably, in the goal distance metric shown in Table III, the proposed method obtains the best or second-to-best results. Moreover, in 113 out of 120 experimental runs, the robot sustains the motion toward the goal location after paralysis, thus demonstrating the robustness of the proposed gait controller.

In the L2L3 scenario metrics evolution, depicted in Fig. 4, three scenario acts can be distinguished: (i) Walking $t \in [0, 350)$, (ii) Paralysis $t \in [350, 2450)$, indicated by yellow color in the plots, and (iii) Recovery $t \in [2450, 3500]$. The consequence of using only *Walking IM* can be observed in Baseline performance where, during Paralysis, the MAE reaches the highest values. After Paralysis, the MAE returns to its original level; see Fig. 4. Thus, for Baseline, *Walking IM* controls the dynamic context during Walking and Recovery better than during Paralysis. Different behavior can be observed for other controllers, which implement incremental learning, where the MAE is lower than for Baseline during Paralysis. A scenario run of the proposed method adapting to L2L3 paralysis is shown in Part 2 of the video in the

TABLE II
The ranking of each method averaged over 20 trials. The ranking significance is tested by the Friedman test, where p-value < 0.005 for both metrics.

Method	Cumulative MAE	Goal Distance
Baseline	4.7	4.1
Scheduled [42]	4.3	3.8
CRPpredictive [34]	2.8	5.0
CRPcluster [35]	3.2	3.1
L2norm [36]	3.4	3.3
Proposed	**2.7**	**1.8**

TABLE III
Final distance from the goal location after starting from distance 144 m. The best results are in bold, and the second best are underlined.

Method	**L2** / m	**L1R2** / m	**L1R3** / m	**L2R3** / m
Baseline	54.1(0.2)	107.0(3.3)	41.1(1.0)	80.6(0.7)
Scheduled [42]	54.6(1.6)	**85.0**(**8.2**)	46.0(3.1)	74.1(4.9)
CRPpredictive [34]	60.8(18.1)	106.7(11.3)	55.7(13.0)	104.5(13.3)
CRPcluster [35]	**34.9**(**10.2**)	104.5(15.0)	41.7(1.6)	66.0(14.9)
L2norm [36]	52.0(5.0)	98.9(4.5)	40.5(1.9)	65.8(9.5)
Proposed	41.7(4.3)	90.0(23.6)	**32.7**(**3.5**)	**50.2**(**5.1**)

Average values with standard deviation in brackets among five trials.

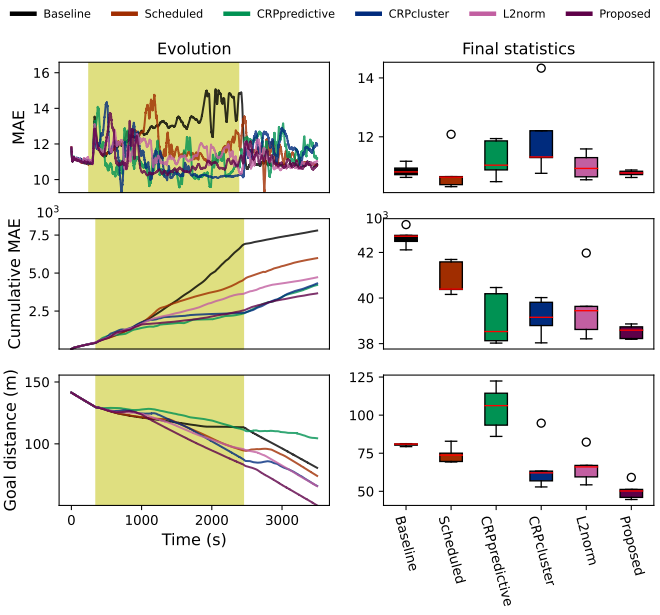


Fig. 4. Performance metrics evolution in the L2L3 scenario with the evaluation metrics averaged over five experimental trials. The rows correspond to three evaluation metrics: (i) the MAE averaged over the last second, (ii) the cumulative sum of the MAE, and (iii) the reached goal distance. The left column shows the evolution of the metrics with paralysis, indicated by the yellow interval. The right column shows the final metrics statistics of five trials for each evaluated method. The cumulative MAE evolution has subtracted trend: $E_{cum}(t) - 10t$, to distinguish the results visually.

supplementary material.[2]

The improved performance of the incrementally learnable controllers is further analyzed by observing their internal behavior. In Section VI-A, it is shown that the performance improvement is the result of the robot's predictive capabilities, where the robot recognizes the dynamic context changes. In Section VI-B, the interpretability is demonstrated by showing why the robot decided to learn a new model and how the damaged legs can be identified from the model parameters. In Section VI-C, it is shown that the decision capabilities translate in the real world, where a large hexapod walking robot recognizes damage (leg paralysis) while traversing through rough terrain.

A. Performance Improvement by IM Management

The performance improvement follows the addition of a new IM during Paralysis. We refer to the first model trained during Paralysis, $\mathbf{f}^2$, as *Paralysis IM*. After 1500 s, all methods, except Baseline, learned a *Paralysis IM* as shown in Fig. 5, and as a result, the Performance MAE decreases under the value of Baseline, see Fig. 4. However, observing the final goal distance of CRPpredictive being higher than that of Baseline, the incremental learning capability by itself is not sufficient.

For every gait, the robot must decide whether to learn or not to learn. The late learning decision of Scheduled, after 1000 s in Fig. 5b, causes the robot to use an IM inappropriate to the context, thus increasing MAE. However, CRPpredictive, which is the fastest to recognize the context change as depicted for the first babbling in Fig. 5c, is also too sensitive, and the robot spends most of the time on motor babbling. More moderate CRPcluster learns two models during Paralysis, see Fig. 5d, resulting in the fastest locomotion during Paralysis as is observable from the steepest slope in the goal distance evolution Fig. 4. During the transition from Paralysis to Recovery, the AIC infers behavior novel to the robot, which is recognized by CRPpredictive and CRPcluster; however, the methods do not recognize that the dynamics of the Recovery are similar to Walking, which is not the case of L2norm and Proposed.

Recalling IM for a similar context saves the robot's time, as IM learning costs five minutes of motor babbling. In the goal distance metric depicted in Fig. 4, the motor babbling is visible as a temporal slowdown, where only the L2norm and Proposed methods sustain the goal approach speed. L2norm learns and selects *Paralysis IM* during Paralysis and selects *Walking IM* during Recovery, see Fig. 5e. Thus, L2norm learns only one IM while being competitive to CRPcluster, which learns three IMs. The distinguishing factor between L2norm and the proposed method is the consistent timeliness of the model selection: over multiple trials, the "using Paralysis IM" (fuchsia triangles) and "return to Walking IM" (orange triangles) events are scattered in Fig. 5e but clustered in Fig. 5f. On average, the L2norm method returned to Walking IM 395 s after recovery, while for the proposed it took only 154 s. The proposed method provides timely model selection and incremental learning that are shown to be crucial characteristics in embodied lifelong learning.

[2]The duration of the video documented experiment is shortened to 1750 simulated seconds.

B. Interpretability of the Decision and Damage Identification

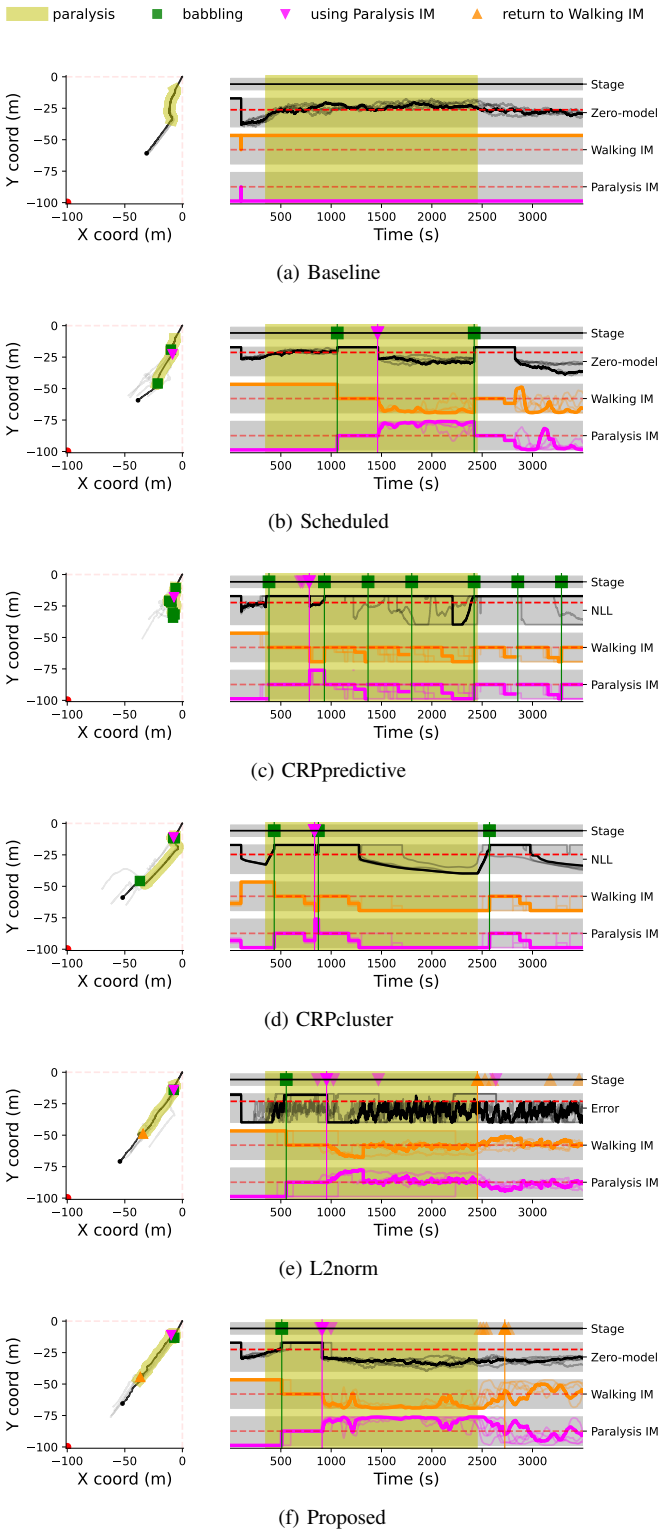
For the L2R3 scenario, we analyzed the cause of the learning decision, how the robot compensates for the paralysis, and how the robot recognizes the recovery. The stages of the experiment are depicted in Fig. 6 with the projected IM selections on both the timeline and traversed path. The robot starts with *Walking IM*, which keeps the zero-model log-odds (black) in negative values, meaning that the zero-model is less likely than *Walking IM* before the paralysis. After the paralysis, the positive zero-model log-odds imply that the zero-model is better than *Walking IM* at explaining the heading velocity, which triggers motor babbling (highlighted by green). The babbling results in *Paralysis IM* that has low values of the L2 and R3 weights $w^{m,c}_{\text{L2},d}$ in Fig. 7, indicating disentanglement between L2 and R3 leg joints and sensory modalities. The low L2 and R3 weights cause the control rule (13) to ignore the paralyzed legs. After the recovery, we can observe that the log-odds between *Paralysis* and *Walking IMs* (fuchsia curve in Fig. 6) are in favor of *Walking IM*; thus, the robot recognizes the recovery.

By analyzing the weights of the phase-embedded model, it is possible to recognize the properties of the dynamic context. Comparing *Paralysis IM* to *Walking IM* in Fig. 7, we can see that the coxa, femur, and tibia of the paralyzed legs L2 and R3 have little effect on the sensory measurement. We can also observe the phase-dependency of the gait dynamics by comparing *Walking IM* with *Resting IM*, which was learned during motor babbling while standing still. The phase relation between heading velocity and L2 coxa, $w^{\text{head},c}_{\text{L2coxa},d}$ of *Resting IM* shows that the positive command at the phase d results in the positive heading velocity at the concurrent phase $c = d$. The symmetry is broken in *Walking IM*, where the command has positive consequences only during the motor phases $d = 2$ and $d = 3$, because during $d = 1$ and $d = 4$, the legs are in swing, and changing the coxa angle during the swing has no immediate consequences to the heading velocity. The delayed sensory consequences are encoded by the values outside the $d = c$ diagonal.

C. Real Robot Deployment

We demonstrate that the robot recognizes and adapts to change in a real environment by studying the behavior of the robot during paralysis and recovery. The proposed method runs onboard computational resources with the Intel® Core™ i7-10710U processor, which is shown to provide sufficient computational power. Similarly to the simulated setup, the hexapod walking robot's *Walking IM* is trained on a smooth office floor, depicted in Fig. 3b, but then deployed outdoors on rough tiled terrain (see Fig. 8a), which perturbs the locomotion with up to 1 cm elevation differences. *Paralysis IM* is trained outdoors with the rear left leg paralyzed.

We compare the following five scenarios where the robot navigates forward: (i) with no paralysis, during rear left leg paralysis, (ii) without and (iv) with *Paralysis IM*, and



(a) Baseline

(b) Scheduled

(c) CRPpredictive

(d) CRPcluster

(e) L2norm

(f) Proposed

Fig. 5. Left: Projected robot trails navigating toward the goal at $(-100,-100)$ with marked events triggered by the continual learning algorithm. Right: Model competition causing the events. Stage row: Projected events—babbling start (green square, each implies learning a new model), first *Paralysis IM* selection (fuchsia triangle), and first *Walking IM* selection after Paralysis (orange triangle). Below: Scaled decision signal (e.g., zero-model log-odds for Proposed) with threshold θ (red dash line) deciding on learning transition. *Walking* and *Paralysis IM* rows: Softmax-normalized log-likelihoods across all models (including non-visualized ones). The 0.5 threshold is indicated by a red dashed line. Baseline and Scheduled events are not log-odds-driven, but signals are shown for comparison. One trajectory per method is shown in full color; translucent colors indicate other runs.

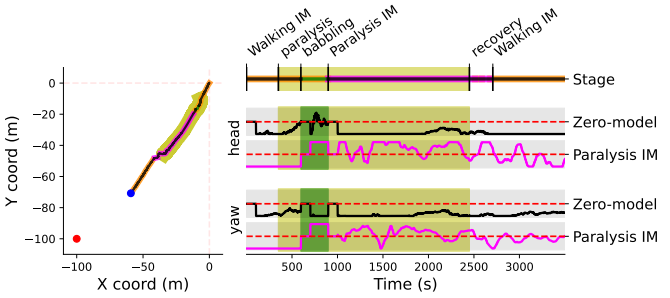


Fig. 6. On the right, log-odds evolution of zero-model:best model (black) where the value lesser than zero (below the red dash line) indicates that the former is less likely than the latter. Similarly, we interpret *Paralysis IM*:*Walking IM* (fuchsia) log-odds evolution. The log-odds are shown for the heading velocity and yaw change and clipped to $[-0.05, 0.05]$. The motor babbling is indicated by green, and the legs are paralyzed by yellow intervals. The experiment stages are projected on the robot's path with the same color coding on the left.

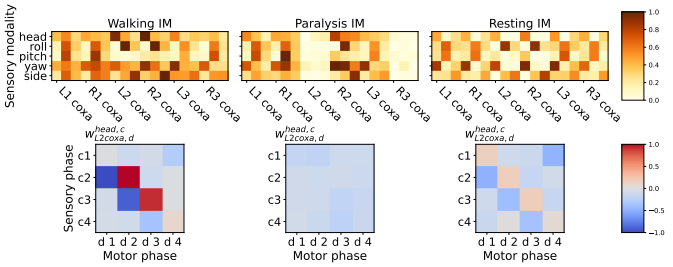
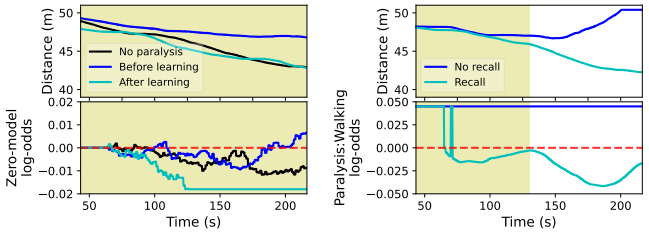


Fig. 7. Detail of the IM parameters for *Walking IM* and *Paralysis IM* used during the scenario depicted in Fig. 6, and *Resting IM* that the robot learns while standing still. The upper row depicts sensorimotor relation strength $s^{\text{sensor}}_{\text{motor}} = \sum |w^{\text{sensor},c}_{\text{motor},d}|/K$ where K is the constant normalizing values to $s \in [0,1]$. After every labeled coxa joint, the two subsequent joints, the femur and the tibia, are followed. The bottom row shows the detail of the phase relation between the paralyzed leg's coxa and heading velocity. The weight at the cth row and dth column encodes the relationship between the motor at the cth and sensor at the dth phase.

during recovery (iv) without and (v) with the capability to recall *Walking IM*. During each 255 s long scenario, the robot navigates toward goal location $\mathbf{l}^* = (50\,\text{m}, 0\,\text{m})$ by tracking heading (23) and yaw (24) speed references limited to $[0\,\text{m s}^{-1}, 0.025\,\text{m s}^{-1}]$ and $[-0.03\,\text{rad s}^{-1}, 0.03\,\text{rad s}^{-1}]$, respectively. Unlike the simulated setup, both IMs generate control signals combined as $\mathbf{u} = (1-\gamma)\mathbf{u}^{\text{Walking IM}} + \gamma\mathbf{u}^{\text{Paralysis IM}}$, where $\gamma \equiv \text{sigm}(50L^{\text{Paralysis IM}})$ is sigmoid function of *Paralysis IM* log-odds. In Figs. 8b and 8c, the goal distance and log-odds evolution averaged from three experimental runs are presented. The real environment scenarios are documented in supplementary material in Part 3 of the video.

During the paralysis, see in Fig. 8b, *Walking IM* performs worse than before the paralysis; however, the performance improves after learning *Paralysis IM*. After the recovery, *Paralysis IM* is unable to control the forward motion, and only after *Walking IM* recall the robot is capable of locomoting toward the goal, see in Fig. 8c. As expected, the robot traverses a larger distance when the model corresponding to the dynamic context is selected.

(a) Real hexapod walking robot on tiled ground



(b) Paralysis (c) Recovery

Fig. 8. (a) Real hexapod walking robot on tiled ground. (b) Comparison of goal distance and log-odds evolutions for different scenarios. On the left, a non-paralyzed robot using only *Walking IM* (black), paralyzed robot before (blue), and after (cyan) learning the *Paralysis IM*. The zero-model log-odds threshold is set to $\theta = 0.0$. On the right, transition from the paralysis to the recovery at 130 s. The behavior with (cyan) and without (blue) ability to recall *Walking IM* is compared.

If the selected model does not correspond to the dynamic context, the robot is capable of recognizing the discrepancy. Before learning *Paralysis IM*, the zero-model log-odds increase above the threshold during the paralysis, see Fig. 8b, thus initiating the learning. Similarly, after the recovery, *Paralysis IM* log-odds rapidly decrease; thus, *Walking IM* is more influential during the motion control, which improves the performance. Therefore, the real robot is capable of recognizing and adapting to the shifting dynamic context.

VII. Discussion

Lifelong deployment in the open world can expose robots to experiences that are unavailable in pre-deployment models. The results show that recognition of the novel experiences and their continual adaptation sustains the locomotion over multiple dynamic contexts. In the presented comparative study, the proposed method provides the best performance due to its timely and consistent decisions on model learning and model selection. It is demonstrated that the robot's decisions using probabilistic reasoning, where the learned models contain interpretable characteristics of the dynamic contexts. In the rest of the discussion section, we provide comments on the made observations.

A. Improving through Testing own Beliefs

A seemingly minor and yet insightful observation is that the world model's ability to falsify itself leads to performance improvement. The observation supports the basic principle of developmental robotics that the system must have the ability to verify itself [44]. The common use of *"verification"* contains both ideas of the confirmation, getting a result consistent with the model, and falsification, trying to find an exception to the model. However, while the controller confirms the model by walking, the controller is not motivated to falsify the said model. The confirmation mechanism itself is not sufficient for the improvement as the controller can select only the actions that are consistent with the model. Only after adding a slight noise to the motor commands did the model selection behave as expected: it switched IM only when the dynamics changed. We conclude that by adding motor noise to the controller, the robot continually tests and eventually falsifies its own beliefs, resulting in world model improvement.

B. Limitation of the Motor Babbling

The data collection determines the quality of the world model and the robot's performance; hence, capturing the right amount of data is crucial. The herein-used motor babbling is an uninformed data collection that consumes a fixed amount of time with the same data sampling methodology regarding the context and current world model. However, the world model has an awareness of which sensory modalities are surprising, as it is reflected in the zero-model log-odds in Fig. 6, where the heading velocity modality has higher log-odds during babbling than yaw change. The informed data collection can decrease the time needed to collect the data, which can be achieved by tracking the observation novelty [45], [46], by searching in the less dimensional goal space [47], or by utilizing predictive models [48]. In future work, we plan to investigate the utilization of the world model in informed data collection to improve the data collection efficiency.

C. Generalizing the Model Selection

The focus of the presented study is on prediction-driven model selection. Thus, the proposed method ignores evidence at the disposal that can anticipate the appropriate IM even before a sensory prediction error is detected. The model likelihood (16) depends on the given reference as it is shown in behavioral [36] or predictive [42] systems. By extending the observed data D with exteroception, the robot learns patterns between exteroception and selected IMs [33], thus gaining the ability to predict (as opposed to detecting) the dynamic context. Beyond the phenomena analysis, introspecting on the history of selected models results in finding transition patterns between IMs [33], [36]. In that context, the prediction-driven model selection is not only a decision-making method but also serves as a bootstrap for learning and utilizing the decision anticipation. In future work, we plan to investigate the integration of the other available evidence into the model selection process.

VIII. Conclusion

Active gait control inference is proposed with an incrementally growing world model. From the Internal Model and Predictive Coding principles, we derive three stages of optimization: selecting the model, estimating the state, and controlling the gait. The proposed method is tested in simulated and real leg paralysis scenarios, where the robot recognizes the paralysis, compensates for the paralyzed leg, and recognizes when the leg recovers in real time. The interpretability of the robot's decisions and knowledge is demonstrated when the robot continually informs of its certainty and represents novel knowledge in linear models, which can be further analyzed.

Moreover, the proposed probabilistic framework for model selection can be easily scaled by other conditionals, improving the growth of the world model. The continual world model learning provides scalability in the robot's reasoning, expanding the robot's longevity and operational domain in real time. In future work, we plan to study the world model hivemind in robot teams, which is essential for establishing a large-scale sustainable ecosystem of evolving robotic deployments.

REFERENCES

[1] D. Drotman, S. Jadhav, D. Sharp, C. Chan, and M. T. Tolley, "Electronics-free pneumatic circuits for controlling soft-legged robots," *Science Robotics*, vol. 6, no. 51, p. eaay2627, 2021. doi: 10.1126/scirobotics.aay2627

[2] J. Z. Ge, A. A. Calderón, L. Chang, and N. O. Pérez-Arancibia, "An earthworm-inspired friction-controlled soft robot capable of bidirectional locomotion," *Bioinspiration & Biomimetics*, vol. 14, no. 3, p. 036004, 2019. doi: 10.1088/1748-3190/aae7bb

[3] M. Cianchetti, M. Calisti, L. Margheri, M. Kuba, and C. Laschi, "Bioinspired locomotion and grasping in water: the soft eight-arm OCTOPUS robot," *Bioinspiration & Biomimetics*, vol. 10, no. 3, p. 035003, 2015. doi: 10.1088/1748-3190/10/3/035003

[4] F. Wiesemuller, C. Winston, A. Poulin, X. Aeby, A. Miriyev, T. Geiger, G. Nystrom, and M. Kovac, "Self-sensing cellulose structures with design-controlled stiffness," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 4017–4024, 2021. doi: 10.1109/LRA.2021.3067243

[5] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, 2015. doi: 10.1038/nature14422

[6] L. Simoni, M. Beschi, G. Legnani, and A. Visioli, "Friction modeling with temperature effects for industrial robot manipulators," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 3524–3529. doi: 10.1109/IROS.2015.7353869

[7] S. Godon, A. Ristolainen, and M. Kruusmaa, "An insight on mud behavior upon stepping," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 039–11 046, 2022. doi: 10.1109/LRA.2022.3194667

[8] H. T. Tramsen, L. Heepe, J. Homchanthanakul, F. Wörgötter, S. N. Gorb, and P. Manoonpong, "Getting grip in changing environments: the effect of friction anisotropy inversion on robot locomotion," *Applied Physics A*, vol. 127, no. 5, p. 389, 2021. doi: 10.1007/s00339-021-04443-7

[9] D. Belter, J. Wietrzykowski, and P. Skrzypczyński, "Employing natural terrain semantics in motion planning for a multi-legged robot," *Journal of Intelligent & Robotic Systems*, vol. 93, no. 3-4, pp. 723–743, 2019. doi: 10.1007/s10846-018-0865-x

[10] D. Valenzo, A. Ciria, G. Schillaci, and B. Lara, "Grounding context in embodied cognitive robotics," *Frontiers Neurorobotics*, vol. 16, 2022. doi: 10.3389/fnbot.2022.843108

[11] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, Oct. 2020. doi: 10.1126/scirobotics.abc5986

[12] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *Trends in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004. doi: 10.1016/j.tins.2004.10.007

[13] R. P. N. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, vol. 2, no. 1, pp. 79–87, 1999. doi: 10.1038/4580

[14] K. J. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology-Paris*, vol. 100, no. 1, pp. 70–87, 2006. doi: 10.1016/j.jphysparis.2006.10.001

[15] P. Lanillos and G. Cheng, "Adaptive robot body learning and estimation through predictive coding," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4083–4090. doi: 10.1109/IROS.2018.8593684

[16] C. Meo, G. Franzese, C. Pezzato, M. Spahn, and P. Lanillos, "Adaptation through prediction: Multisensory active inference torque control," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 1, pp. 32–41, 2023. doi: 10.1109/TCDS.2022.3156664

[17] M. Baioumy, P. Duckworth, B. Lacerda, and N. Hawes, "Active inference for integrated state-estimation, control, and learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4665–4671. doi: 10.1109/ICRA48506.2021.9562009

[18] C. Pezzato, R. Ferrari, and C. H. Corbato, "A novel adaptive controller for robot manipulators based on active inference," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2973–2980, 2020. doi: 10.1109/LRA.2020.2974451

[19] H. Imamizu, T. Kuroda, T. Yoshioka, and M. Kawato, "Functional magnetic resonance imaging examination of two modular architectures for switching multiple internal models," *Journal of Neuroscience*, vol. 24, no. 5, pp. 1173–1181, 2004. doi: 10.1523/JNEUROSCI.4011-03.2004

[20] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion in Neurobiology*, vol. 9, no. 6, pp. 718–727, 1999. doi: 10.1016/S0959-4388(99)00028-8

[21] M. Desmurget, C. Epstein, R. Turner, C. Prablanc, G. E. Alexander, and S. T. Grafton, "Role of the posterior parietal cortex in updating reaching movements to a visual target," *Nature Neuroscience*, vol. 2, no. 6, pp. 564–567, 1999. doi: 10.1038/9219

[22] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Networks*, vol. 11, no. 7, pp. 1317–1329, 1998. doi: 10.1016/S0893-6080(98)00066-5

[23] N. Sugimoto, J. Morimoto, S.-H. Hyon, and M. Kawato, "The emosaic model for humanoid robot control," *Neural Networks*, vol. 29-30, pp. 8–19, 2012. doi: 10.1016/j.neunet.2012.01.002

[24] M. Emadi Andani, F. Bahrami, and P. Jabehdar Maralani, "AMA-MOSAICI: an automatic module assigning hierarchical structure to control human motion based on movement decomposition," *Neurocomputing*, vol. 72, no. 10-12, pp. 2310–2318, 2009. doi: 10.1016/j.neucom.2008.12.016

[25] H. Haghighi, F. Abdollahi, and S. Gharibzadeh, "Brain-inspired self-organizing modular structure to control human-like movements based on primitive motion identification," *Neurocomputing*, vol. 173, no. 3, pp. 1436–1442, 2016. doi: 10.1016/j.neucom.2015.09.017

[26] E. Rohmer, S. P. N. Singh, and M. Freese, "Coppeliasim (formerly V-REP): a versatile and scalable robot simulation framework," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 1321–1326. doi: 10.1109/IROS.2013.6696520

[27] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020. doi: 10.1016/j.inffus.2019.12.012

[28] C. L. Buckley, C. S. Kim, S. McGregor, and A. K. Seth, "The free energy principle for action and perception: A mathematical review," *Journal of Mathematical Psychology*, vol. 81, pp. 55–79, 2017. doi: 10.1016/j.jmp.2017.09.004

[29] M. Thor, T. Kulvicius, and P. Manoonpong, "Generic neural locomotion control framework for legged robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4013–4025, Sep. 2021. doi: 10.1109/TNNLS.2020.3016523

[30] J. Homchanthanakul and P. Manoonpong, "Continuous online adaptation of bioinspired adaptive neuroendocrine control for autonomous walking robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 5, pp. 1833–1845, May 2022. doi: 10.1109/TNNLS.2021.3119127

[31] C. Yang, K. Yuan, Q. Zhu, W. Yu, and Z. Li, "Multi-expert learning of adaptive legged locomotion," *Science Robotics*, vol. 5, no. 49, p. eabb2174, 2020. doi: 10.1126/scirobotics.abb2174

[32] M. Thor and P. Manoonpong, "Versatile modular neural locomotion control with fast learning," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 169–179, 2022. doi: 10.1038/s42256-022-00444-0

[33] M. Haruno, D. M. Wolpert, and M. Kawato, "MOSAIC model for sensorimotor learning and control," *Neural Computation*, vol. 13, no. 10, pp. 2201–2220, 2001. doi: 10.1162/089976601750541778

[34] Z. Wang, C. Chen, and D. Dong, "Lifelong incremental reinforcement learning with online bayesian inference," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 4003–4016, Aug. 2022. doi: 10.1109/TNNLS.2021.3055499

[35] T. Zhang, Z. Lin, Y. Wang, D. Ye, Q. Fu, W. Yang, X. Wang, B. Liang, B. Yuan, and X. Li, "Dynamics-adaptive continual reinforcement learning via progressive contextualization," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2024. doi: 10.1109/TNNLS.2023.3280085

[36] C. Gumbsch, M. V. Butz, and G. Martius, "Autonomous identification and goal-directed invocation of event-predictive behavioral primitives," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 298–311, 2021. doi: 10.1109/TCDS.2019.2925890

[37] Z. Song and Z. Zhang, "Visually guided sound source separation with audio-visual predictive coding," *IEEE Transactions on*

Neural Networks and Learning Systems, pp. 1–15, 2024. doi: 10.1109/TNNLS.2023.3288022

[38] Z. Straka, T. Svoboda, and M. Hoffmann, "Precnet: Next-frame video prediction based on predictive coding," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 10 353–10 367, Aug. 2024. doi: 10.1109/TNNLS.2023.3240857

[39] M. Tarokh, "Hyperstability approach to the synthesis of adaptive controllers for robot manipulators," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 3, 1991, pp. 2154–2159. doi: 10.1109/ROBOT.1991.131947

[40] R. Szadkowski and J. Faigl, "Hexapod gait control through internal model belief update," in *International Symposium on Adaptive Motion of Animals and Machines*, 2023. doi: 10.18910/92290

[41] H. Zhang, R. Chi, and B. Huang, "Data-driven internal model learning control for nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2024. doi: 10.1109/TNNLS.2023.3331367

[42] R. Szadkowski, M. S. Nazeer, M. Cianchetti, E. Falotico, and J. Faigl, "Bootstrapping the dynamic gait controller of the soft robot arm," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 2669–2675. doi: 10.1109/ICRA48891.2023.10160579

[43] M. T. Nguyenová, P. Čížek, and J. Faigl, "Modeling proprioceptive sensing for locomotion control of hexapod crawling robot in robotic simulator," in *2018 Modelling and Simulation for Autonomous Systems (MESAS)*, 2019, pp. 215–225. doi: 10.1007/978-3-030-14984-0_17

[44] A. Stoytchev, "Some basic principles of developmental robotics," *IEEE Trans. Auton. Ment. Dev.*, vol. 1, no. 2, pp. 122–130, 2009. doi: 10.1109/TAMD.2009.2029989

[45] R. Der and G. Martius, "Novel plasticity rule can explain the development of sensorimotor intelligence," *Proceedings of the National Academy of Sciences*, vol. 112, no. 45, pp. E6224–E6232, 2015. doi: 10.1073/pnas.1508400112

[46] L. Pape, C. M. Oddo, M. Controzzi, C. Cipriani, A. Förster, M. C. Carrozza, and J. Schmidhuber, "Learning tactile skills through curious exploration," *Frontiers in Neurorobotics*, vol. 6, 2012. doi: 10.3389/fnbot.2012.00006

[47] A. Baranes and P.-Y. Oudeyer, "Intrinsically motivated goal exploration for active motor learning in robots: A case study," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 1766–1773. doi: 10.1109/IROS.2010.5651385

[48] P. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Trans. Evol. Comput.*, vol. 11, no. 2, pp. 265–286, 2007. doi: 10.1109/TEVC.2006.890271