

Reviewed Preprint

v1 • August 17, 2026

Not revised

✉ For correspondence:

bbrunton@uw.edu

* These authors contributed equally to this work

Competing interests: No

competing interests declared

Funding: See page 5

Reviewing editor: Maarten Zwart, University of St Andrews, United Kingdom

© 2026, Brunton et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

The digital sphinx: Can a worm brain control a fly body?

Bingni W Brunton^{1,*} ✉, Elliott TT Abe^{1,*}, Lawrence Jianqiao Hu^{1,2}, John C Tuthill³

¹Dept of Biology, University of Washington, Seattle, United States • ²Graduate Program in Neuroscience, University of Washington, Seattle, United States • ³Dept of Neurobiology & Biophysics, University of Washington, Seattle, United States

eLife Assessment

What can a neural network trained to imitate animal behavior tell us about biology? This **valuable** work uses deep reinforcement learning to train an artificial neural network to transform the dynamics of a recurrent neural network based on the *C. elegans* connectome into an adult *Drosophila* walking program in a physical model of the fly body, demonstrating that achieving plausible output dynamics does not in and of itself imply biologically meaningful simulation. Evidence for this basic claim is **solid**, but more extensive analyses, better methodological description, and a discussion of deeper challenges in the undertaking of biological brain modeling would strengthen the study. This result demands the attention of the practitioners of the growing field of connectome simulation for the purpose of gaining mechanistic understanding of nervous system function.

<https://doi.org/10.7554/eLife.111516.1.sa3>

Abstract

Animal intelligence is not purely a product of abstract computation in the brain, but emerges from dynamic interactions between the nervous system and the body. New connectome datasets and musculoskeletal models now enable integrated, closed-loop simulations of the neural and biomechanical systems of the fruit fly *Drosophila*, an ideal model organism to investigate embodied intelligence. However, many biological parameters of the nervous system and the body, as well as how they interface, remain unknown. To fill such gaps, researchers are turning to deep reinforcement learning (DRL), a data-driven optimization framework, to create virtual animals that imitate the behavior of real animals. Here, we provide a cautionary tale about the interpretation of such models. We constructed a virtual chimera of two phylogenetically distant species: a connectome of the *C. elegans* nematode worm and a biomechanical model of the fly body. The worm connectome receives sensory information from the fly body, and an artificial neural network is trained with DRL to map worm motor neuron activations to the fly's leg actuators. The resulting digital sphinx produces highly realistic fly walking—yet it is biologically meaningless. This exercise teaches us nothing about either animal and exposes a core peril of connectome-body models: behavioral fidelity is achievable without biological fidelity, making such models easy to overinterpret. Done carefully, virtual animals can be powerful partners to biological experiments, but only if their components and interfaces are grounded in biology.

Introduction

In the last few years, researchers in neuroscience and AI have been working towards developing computational models of brains capable of controlling realistic animal bodies in virtual environments (e.g., [1, 2, 3, 4, 5, 6]). These virtual animal models have the potential to enable *in silico* experiments at scale, generating new theories and hypotheses about embodied intelligence to be tested in the lab. They could also serve as community repositories of accumulated knowledge

about the nervous system and the body. Motivated by the potential utility of such models, the field of virtual animals is growing rapidly, including both academic and industry research efforts. Some have referred to these models as brain “emulations” or “uploads,” attracting substantial interest from scientists and the public (see <https://x.com/alexwg/status/2030217301929132323> ) . But such models of virtual animals require cautious interpretation.

Among animals that walk, the integration of brain wiring and body models is perhaps closest to fruition in *Drosophila*, due to the recent completion of multiple complete wiring diagrams (known as connectomes) of the fly nervous system. The fly is the only animal with legs for which nearly comprehensive connectomes of its brain and nerve cord exist [7, 8]. There is also a wealth of knowledge about the neural basis of many fly behaviors, along with established tools for experimentally recording and manipulating genetically identified cell types in the fly nervous system. Two open-source biomechanical models of the whole fly body [9, 3] have been implemented in the physics engine MuJoCo [10]. Despite this progress, closed-loop integration of biomechanical and neural models remains far from straightforward. Key parameters governing how they interface remain incompletely characterized, including how motor neuron commands result in forces exerted by the body and how sensory neurons detect and transduce internal and external stimuli.

Where interfaces between brains and body models are missing or only partially characterized, one approach is to train an artificial neural network (ANN) to approximate these interfaces with deep reinforcement learning (DRL). In building virtual animal models, a motor policy is commonly learned by DRL so that the integrated, closed-loop virtual body successfully mimics the detailed kinematics of real animal behavior. The implicit logic is, if it looks and acts like a cat, it is a decent model of a cat.

Additional realism is added when the motor policy network is constrained by a connectome dataset. However, many biophysical parameters for individual neurons and synapses remain unmeasured. To approximate these, researchers have used neurophysiological recordings, optogenetic perturbation experiments, and other biological observations as additional constraints [11, 12, 13, 14]. These approaches have demonstrated success at predicting some key aspects of neural dynamics, such as sensory receptive fields or motor rhythms, which will be essential for developing meaningful closed-loop models based on connectome datasets.

Methods

Here we develop a virtual chimera with the brain of a worm and the body of a fly. We used the adult hermaphrodite *C. elegans* nematode connectome dataset [15, 16, 5], including the identities of its 302 neurons and their synapses (Fig. 1A ) . We created a graded activation model of the network, consistent with the fact that most *C. elegans* neurons do not fire action potentials. Following the model in [5], synaptic weights were based on the number of synapses from the connectome, signed by the neurotransmitter identity of each presynaptic neuron. To simulate the physics of fly walking, we used a MuJoCo biomechanical model of the fly body [3]. The fly body model had 42 actuators for its 6 legs that were driven by torque commands; proprioceptive feedback from 148 sensors reported leg and body angles, along with their derivatives.

To interface brain and body, we routed proprioceptive signals (normalized and filtered through a tanh nonlinearity) from the fly body to the worm’s 82 sensory neurons by random projection, without training any parameters at the sensory interface (Fig. 1B ) . We did not discriminate among different types of sensory neurons (e.g., proprioceptive, chemosensory, tactile, etc.). To actuate the body model, we then used DRL to train a variational encoder-decoder network with a 16-dimensional latent space to imitate the 3D joint kinematic trajectories of real walking flies [17, 18]. This motor decoder maps worm motor neuron activity into torque commands for the fly’s legs. The policy network was trained in closed loop using PPO as implemented by MIMIC-MJX [19].

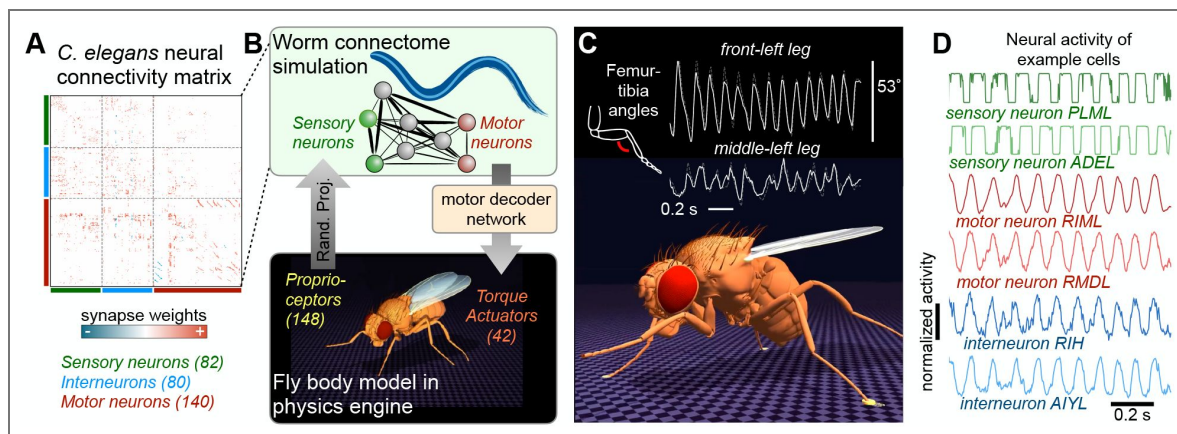


Figure 1. The digital sphinx model had the brain of a worm and the body of a fly, and it is able to reproduce physically realistic walking with DRL training (see walking animation in [Video 1](#)).

A, Connectivity matrix from the adult hermaphrodite *C. elegans* connectome. **B**, A schematic of how the connectome simulation interfaces with the fly body model with proprioceptors and actuators. **C**, Example traces of leg joint kinematics produced by the model (white) trained to imitate target trajectories from real flies (gray dashed). **D**, Example traces of neurons in the simulated connectome model. The traces were highly cyclic because the connectome was driven by proprioceptive sensory signals in closed loop during cyclic stepping. The identities of the neurons are not biologically meaningful in this model.

Results

The digital sphinx model, with a worm brain and a fly body, produced highly realistic walking, including joint angle trajectories and leg coordination patterns (Fig. 1C, Video 1). Yet the model is implausible and scientifically meaningless. Worms move by wiggling; flies walk with articulated legs. The fly brain is 500X larger than the nematode's, and fly legs alone are controlled by more motor neurons than exist in the entire worm nervous system. The model produced neural activity (Fig. 1D), but it is uninterpretable. The model also fails to qualify as brain “emulation,” which would require that the model preserves the biological meaning of its components. The fact that a worm connectome can be trained to control fly walking reveals nothing about either worms or flies.

Discussion

These results carry two lessons. First, users and readers beware—DRL is a capable optimization algorithm. The motion imitation approach is able to fit expressive ANNs to complex target data, even when the imposed constraints are incomplete or outright wrong. Second, even when brain and body models are individually accurate, a biologically unrealistic interface between them renders the whole model meaningless. In our digital sphinx, the worm connectome functions simply as a recurrent neural network (RNN) with rich enough dynamics to support learning of realistic locomotor trajectories. Its role in the movement policy could be fulfilled equally well by a randomly connected RNN, akin to reservoir computing [20], since all the learning happens in the black-box ANN motor decoder. This exercise is an example of how a functioning interface is not necessarily a meaningful one. Here, we did not train synaptic weights or cellular parameters of the connectome simulation; training them would only deepen this problem.

On a positive note, we are optimistic that a virtual animal that captures the coordinated function of neural and biomechanical systems can indeed be achieved, but only through careful design of its modules and interfaces. First, neuromechanical models must be well grounded in biological knowledge. Most critically, the inputs and outputs of the connectome model must be accurately connected to the muscles and sensors in the body model. This mapping has been partially achieved by combining the connectome datasets with other imaging modalities, but much more work is needed to accurately identify the thousands of sensory and motor neurons in the fly connectomes. The body models also lack biological realism in their musculature, sensory neurons, and skeletal mechanics; ongoing efforts are gradually improving these body models.

Second, neuromechanical models should be developed in close collaboration with biological experiments. Models can be powerful even without perfectly realistic components, provided they generate testable predictions about feasible experiments. Looking further ahead, swapping brain and body models of related species may one day yield real insights into how their brains and bodies diverged through evolution. However, far more model development and experimental validation is needed before we can learn anything from such a digital sphinx.

Data availability

All code to run the simulations, results from computational experiments, and scripts to reproduce the analyses are available on GitHub: <https://github.com/Brunton-Lab/DigitalSphinx2026>.

Acknowledgements

We thank Eiman Azim, Carl Bergstrom, Astra Bryant, Michael Dickinson, and Talmo Pereira for comments on the manuscript. We thank Sarah Pugliese for assistance with parsing the *C. elegans* connectome. This work was funded by NIH grant U01NS136507 to B.W.B., R01NS14543 to J.C.T. and B.W.B., and the Richard & Joan Komen University Chair to B.W.B. J.C.T. is a New York Stem Cell Foundation–Robertson Investigator.

Additional information

Author Contributions

B.W.B. and E.T.T.A. conceived of the study and designed the analyses. E.T.T.A. and J.L.H. developed and implemented the model. B.W.B. and J.C.T. wrote the manuscript.

Funding

Funder	Grant reference number	Author
HHS National Institutes of Health (NIH)	U01NS136507	Bingni W Brunton
HHS National Institutes of Health (NIH)	R01NS145438	John C Tuthill Bingni W Brunton

Author ORCID iDs

Bingni W Brunton:  <https://orcid.org/0000-0002-4831-3466>

John C Tuthill:  <https://orcid.org/0000-0002-5689-5806>

Additional files

Video 1  A worm brain controls a fly body and produced realistic fly walking kinematics and coordination patterns. The model also produced neural activity of cells in the worm connectome, shown here overlaid on their locations in a worm body, with a few example traces. The identities of cells in this model are meaningless and their neural activity is not interpretable.

References

- [1] Merel J, Aldarondo D, Marshall J, Tassa Y, Wayne G, Ölveczky B (2019) Deep neuroethology of a virtual rodent. *arXiv* <https://doi.org/10.48550/arXiv.1911.09451>
- [2] Aldarondo D, Merel J, Marshall JD, Hasenclever L, Klibaite U, Gellis A, Tassa Y, Wayne G, Botvinick M, Ölveczky BP (2024) A virtual rodent predicts the structure of neural activity across behaviours. *Nature* **632**:594-602 <https://doi.org/10.1038/s41586-024-07633-4> | [PubMed](#)
- [3] Vaxenburg R, Siwanowicz I, Merel J, Robie AA, Morrow C, Novati G, Stefanidi Z, Both G.-J, Card GM, Reiser MB, *et al.* (2025) Whole-body physics simulation of fruit fly locomotion. *Nature* 1-9 <https://doi.org/10.1038/s41586-025-09029-4> | [PubMed](#)
- [4] Wang-Chen S, Stimpfling VA, Lam TKC, Özdiil PG, Genoud L, Hurtak F, Ramdya P (2024) Neuromechfly v2: simulating embodied sensorimotor control in adult *Drosophila*. *Nature Methods* **21**:2353-2362 <https://doi.org/10.1038/s41592-024-02497-y> | [PubMed](#)
- [5] Zhao M, Wang N, Jiang X, Ma X, Ma H, He G, Du K, Ma L, Huang T (2024) An integrative data-driven model simulating *C. elegans* brain, body and environment interactions. *Nature Computational Science* **4**:978-990 <https://doi.org/10.1038/s43588-024-00738-w> | [PubMed](#)
- [6] Liu X, Loring MD, Zunino L, Fouke KE, Longchamp FA, Bernardino A, Ijspeert AJ, Naumann EA (2025) Artificial embodied circuits uncover neural architectures of vertebrate visuomotor behaviors. *Science Robotics* **10**:eadv4408 <https://doi.org/10.1126/scirobotics.adv4408> | [PubMed](#)
- [7] Bates AS, Phelps JS, Kim M, Yang HH, Matsliah A, Ajabi Z, Perlman E, Delgado KM, Osman MAM, Salmon CK, *et al.* (2025) Distributed control circuits across a brain-and-cord connectome. *bioRxiv* <https://doi.org/10.1101/2025.07.31.667571>
- [8] Berg S, Beckett IR, Costa M, Schlegel P, Januszewski M, Marin EC, Nern A, Preibisch S, Qiu W, Takemura S.-y, *et al.* (2025) Sexual dimorphism in the complete connectome of the *drosophila* male central nervous system. *bioRxiv* <https://doi.org/10.1101/2025.10.09.680999> | [PubMed](#)

- [9] Lobato-Rios V, Ramalingasetty ST, Özdil PG, Arreguit J, Ijspeert AJ, Ramdya P (2022) NeuroMechFly, a neuromechanical model of adult *Drosophila melanogaster*. *Nature Methods* **19**:620-627 <https://doi.org/10.1038/s41592-022-01466-7> | PubMed
- [10] Todorov E, Erez T, Tassa Y (2012) Mujoco: A physics engine for model-based control. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 5026-5033 <https://doi.org/10.1109/iros.2012.6386109>
- [11] Cowley BR, Calhoun AJ, Rangarajan N, Ireland E, Turner MH, Pillow JW, Murthy M (2024) Mapping model units to visual neurons reveals population code for social behaviour. *Nature* **629**:1100-1108 <https://doi.org/10.1038/s41586-024-07451-8> | PubMed
- [12] Lappalainen JK, Tschopp FD, Prakhya S, McGill M, Nern A, Shinomiya K, Takemura S-y, Gruntman E, Macke JH, Turaga SC (2024) Connectome-constrained networks predict neural activity across the fly visual system. *Nature* **634**:1132-1140 <https://doi.org/10.1038/s41586-024-07939-3> | PubMed
- [13] Duan S, Dong LL, Fiete I (2025) From synapses to dynamics: Obtaining function from structure in a connectome constrained model of the head direction circuit. *bioRxiv* <https://doi.org/10.1101/2025.05.26.655406>
- [14] Pugliese SM, Chou GM, Abe ET, Turcu D, Lancaster JK, Tuthill JC, Brunton BW (2025) Connectome simulations identify a central pattern generator circuit for fly walking. *bioRxiv* <https://doi.org/10.1101/2025.09.12.675944> | PubMed
- [15] Szigeti B, Gleeson P, Vella M, Khayrulin S, Palyanov A, Hokanson J, Currie M, Cantarelli M, Idili G, Larson S (2014) OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Frontiers in computational neuroscience* **8**:137 <https://doi.org/10.3389/fncom.2014.00137> | PubMed
- [16] Cook SJ, Jarrell TA, Brittin CA, Wang Y, Bloniarz AE, Yakovlev MA, Nguyen KC, Tang LT, Bayer EA, Duerr JS, et al. (2019) Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**:63-71 <https://doi.org/10.1038/s41586-019-1352-7> | PubMed
- [17] Karashchuk P, Rupp KL, Dickinson ES, Walling-Bell S, Sanders E, Azim E, Brunton BW, Tuthill JC (2021) Anipose: A toolkit for robust markerless 3d pose estimation. *Cell Reports* <https://doi.org/10.1016/j.celrep.2021.109730> | PubMed
- [18] Pratt BG, Lee S.-YJ, Chou GM, Tuthill JC (2024) Miniature linear and split-belt tread-mills reveal mechanisms of adaptive motor control in walking *Drosophila*. *Current Biology* **34**:4368-4381.e5, <https://doi.org/10.1016/j.cub.2024.08.006> | PubMed
- [19] Zhang CY, Yang Y, Sirbu A, Abe ETT, Wärnberg E, Leonardis EJ, Aldarondo DE, Lee A, Prasad A, Foat J, et al. (2025) MIMIC-MJX: Neuromechanical emulation of animal behavior. *arXiv* <https://doi.org/10.48550/arXiv.2511.20532>
- [20] Jaeger H (2001) The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German national research center for information technology. pp. 13

Pratt, Brandon; Lee, Su-Yee; Chou, Grant; Tuthill, John (2024) Miniature linear and split-belt treadmills reveal mechanisms of adaptive motor control in walking *Drosophila*. Dryad Digital Repository. <https://doi.org/10.5061/dryad.mpg4f4r73>

Mengdi, Zhao; Ning, Wang; Xinrui, Jiang; Xiaoyang, Ma; Haixin, Ma; Gan, He; Kai, Du; Lei, Ma; Tiejun, Huang (2024) Data for An Integrative Data-driven Model Simulating *C. elegans* Brain, Body and Environment Interactions. Zenodo. <https://doi.org/10.5281/zenodo.13951773>

Roman Vaxenburg and Igor Siwanowicz and Josh Merel and Alice A Robie and Carmen Morrow and Guido Novati and Zinovia Stefanidi and Gert-Jan Both and Gwyneth M Card and Michael B Reiser and Matthew M Botvinick and Kristin M Branson and Yuval Tassa and Srinivas C Turaga (2025) Whole-body physics simulation of fruit fly locomotion. GitHub. ID TuragaLab/flybody <https://github.com/TuragaLab/flybody>

Peer reviews

Reviewer #1 (Public review):

Summary

The authors build a "digital sphinx" by stitching together two neural network models: (i) a recurrent network with fixed parameters derived from the *C. elegans* connectome and imputed physiological (e.g. neural input/output) functions, and (ii) a feedforward encoder-decoder model with learnable parameters intended to represent a central brain - to - motor interface, then harnessing the combined model to a *Drosophila* biomechanical model situated in a physics simulator, and finally using deep reinforcement learning (DRL) training to optimize the parameters of the encoder-decoder model to reproduce a set of spatiotemporal patterns of jointed limb activations that together produce the overall organismal behavior of walking, within the physics simulator.

The primary intent of this paper is to dispel the recent grandiose claims made in the mainstream press by a private company, Eon Systems, to have achieved a major advance in biologically based brain simulation of the production of a set of ethologically relevant motor behaviors by the fly. Representatives of the company referred to this modeling and training process euphemistically and deceptively as "brain uploading". The authors proceed with a reduction-to-triviality exercise by constructing their own high-parameter dynamical brain-plus-body model situated in a physical simulation that produces, after training by reinforcement learning, satisfying ethological behavioral imitation in the same vein as the private company claim, but based on a clearly absurd and biologically unrealistic set of model assumptions.

Secondarily, the paper provides two overall admonitions that they assert their computational demonstration illustrates: that training high parameter network models to imitate behavior, even if they possess some biological detail, will deliver little or no biological insight, and that models of behavioral generation must be built from detailed biological data and, crucially, developed in a hypothesis generation/falsification loop with experimental validation, in order to be scientifically useful.

Appraisal

The authors are well justified in challenging the non-rigorous claims of "uploading" or even the delivery of a neurobehavioral simulation with potential scientific utility, in unison with the vocal criticisms of many other researchers in the fields of AI and neuroscience, and it is an important message to deliver to the world. However, the authors' own modeling counter-exercise, while clever and vivid in imagery, suffers from its own lack of rigor, both in disclosure of implementation and in scientific case-making. Some sacrifice of clarity and thoroughness in the interest of brevity is inevitable under the brief format of this manuscript; however, we suggest that crucial additions and modifications should be made to avoid falling into a similar trap of non-rigorous sensationalism.

Because the private company claims were not accompanied by a scientific paper, preprint, code repository, or much methodological disclosure of any kind, the authors have the particular challenge of building a refutation case against an undefined target. As a consequence, the authors chose their own task, model structure, and training paradigm.

The authors argue that brain models need to be built from biological data to be useful for yielding biological insight. We agree with the overall principle; however, in practice, this procedure is fraught with epistemological difficulty. Biological modeling suffers from a unique challenge within the larger endeavor of scientific/physical modeling, which is that it is

generally unclear as to precisely what biological quantities should be measured and at what level of detail they should be measured. Additionally, biological data will by necessity be incomplete and noisy, and thus decisions of coarse-graining must be made at the outset of large-scale data collection projects, and some, possibly a substantial, level of data imputation will have to be performed in order to build testable models in our lifetimes. Despite the astonishing success of scaling (in both parameter count and corpus size) in engineered neural networks for certain human-like tasks, it is not at all clear that simply adding more detail to biological models will produce deeper scientific insight, or whether cataloging parameters from snapshot data will yield functional simulations. The failed Blue Brain mega-project should provide a lesson, as well as Marder's longstanding work on parameter variation in neural systems. The coupled, pernicious questions of choosing measurement detail and modeling detail represent a deep, unsolved challenge area for the field, and this context should be raised in the text.

The message about overinterpreting models trained with deep reinforcement learning, while valid and important, should be broadened to be a message about overinterpreting trained high-parameter models in general, in their ability to fit data or reproduce simple behavior. Other parameter optimization/learning procedures for building underdetermined and/or high-parameter models risk the same misinterpretation. The prescription of building models in conjunction with experimental prediction and validation is an important point.

The authors leave out an additional important and underappreciated challenge of brain-model-building, which is that imitating a time segment of behavior is a computational task of unspecified, and possibly low complexity. Successful recapitulation of behavioral time series may simply not be considered cognitively interesting, even if the model is built entirely on biological data. While quantifying task complexity is another open area of computational and neuroscientific research, the authors should, at a minimum, describe their particular task data in explicit mathematical terms and preferentially provide some complexity analysis. In the absence of task complexity analysis, at a minimum, computational controls should be applied to demonstrate the necessity of whatever structure or data is being asserted in the model. This epistemological practice is glaringly absent in much, if not most, of the neurobehavioral modeling literature. This paper would be a good opportunity to set an example of rigor.

Finally, the authors' description of prior work in the field of whole-organism neurobiological simulation feels incomplete and skewed toward work in *Drosophila* versus other model organisms. An internet search reveals many published efforts to build neurobehavioral models at varying levels of detail in *C. elegans*, of which only two are referenced.

We do feel this work constitutes an illustrative scientific exercise and important counterpoint to the sensationalism building around efforts in neurobiological simulation. It should inspire further work in defining a rigorous and scientifically productive epistemological framework for these kinds of brain modeling efforts.

Further Comments

(1) The authors oversell the completeness and quality of connectome datasets and what they lack.

Language such as "complete wiring diagrams," "nearly comprehensive connectomes" neglects the well-appreciated gaps in biological data that most practitioners believe necessary for useful, detailed models to be built. There is a brief mention that biological parameters "remain unknown" and that interfaces are "incompletely characterized", but beyond that, the authors do not explain which parameters are missing, why these parameters might matter, and what still needs to be addressed in order to make any plausible whole-brain emulation claims. This may also inadvertently bolster the sensationalist claims that the manuscript is

trying to deflate by giving the impression that neurobiological and physiological data collection is a near-complete exercise.

(2) Prior work in *C. elegans* neurobehavioral modeling should be more acknowledged, if nothing else, for why it has been largely unsatisfying.

C. elegans is rarely discussed, while *Drosophila* is primarily focused on. The status of *C. elegans* connectomics, physiological mapping, biomechanics, and neurobehavioral modeling is worth more treatment.

(3) Critiques of Eon Systems announcements also, by and large, apply to more detailed and disclosed efforts in neurobehavioral modeling using RL for parameter imputation, and this should be recognized.

By way of reference to a tweet in the first paragraph, the authors are responding to a recent claim made by a startup that they have fully "uploaded" a fly brain, a significant advance vis-à-vis prior work in neurobehavioral modeling in *Drosophila*, such as references [3 and 9], which are mentioned as background in the paper but left out of the methodological critique. But one of the central warnings of the paper is around the challenge of interpretability when using reinforcement learning to optimize model parameters. The authors also should acknowledge that the use of RL has been justified by building neurobehavioral model builders as a proxy for the learning and tuning processes thought to occur during animal development.

(4) Substantiate the reservoir computing explanatory claim with appropriate computational controls.

The reservoir computing idea is the only piece of hypothesizing a necessary function for the central brain component model in the paper. This claim could be substantiated with some basic computational controls rather than just hypothesized. We suggest the following possibilities as additions to the model: (a) replace the connectome with an RRNN, (b) shuffle the connectome, or (c) use other simple dynamical systems in place of the worm brain model.

Specific Manuscript Comments

(1) Abstract

"New connectome datasets and musculoskeletal models now enable integrated, closed-loop simulations of the neural and biomechanical systems of the fruit fly *Drosophila*, an ideal model organism to investigate embodied intelligence."

This sentence could mislead non-specialists into thinking all current simulations are novel because the connectome datasets are new. In fact, FlyWire (2024), NeuroMechFly (2022), and other connectomes have already been available for some years now. We believe that this sentence is a chance to make the opposite point that these resources have existed for a while, and that many simulations have been built before.

"However, many biological parameters of the nervous system and the body, as well as how they interface, remain unknown."

Some examples of specific parameter/physiological data types that are missing and thought to be critical, such as neuronal input/output functions, are warranted. See below for a comment on the confusing construct of "interface" as a distinct entity from the neural network.

(2) Introduction

"Among animals that walk, the integration of brain wiring and body models is perhaps closest to fruition in *Drosophila*, due to the recent completion of multiple complete wiring diagrams

(known as connectomes) of the fly nervous system." ...and... "The fly is the only animal with legs for which nearly comprehensive connectomes of its brain and nerve cord exist."

The walking qualifier allows the authors to skirt around the substantial and decades-long work on connectomes in *C. elegans*, which crawls and does not walk. Yet sinusoidal crawling is a multidimensional, adaptive behavior, so it seems this exclusion was for narrative convenience rather than contextual accuracy.

"Despite this progress, closed-loop integration of biomechanical and neural models remains far from straightforward."

Work (and shortcomings) in *C. elegans* neurobehavioral modeling should also be stated here alongside the fly.

"Where interfaces between brains and body models are missing or only partially characterized, one approach is to train an artificial neural network (ANN) to approximate these interfaces with deep reinforcement learning (DRL)."

The choice of "interface" as a distinct, well-defined neurobiological entity is somewhat confusing and may mislead non-practitioner readers. If neuronal and muscular (and their interactions) physiology are incorporated into a neurobehavioral model, then in principle there is nothing left to call an "interface". It would be clearer to explain that prior neurobehavioral models have often inserted a trainable multilayer feedforward network between sensory and central brain and between the central brain and motor effectors in order to have a substrate for learning, and that this insertion may render the entire biological modeling exercise scientifically pointless, or at a minimum require a set of computational controls.

"In building virtual animal models, a motor policy is commonly learned by DRL so that the integrated, closed-loop virtual body successfully mimics the detailed kinematics of real animal behavior."

The authors could define "motor policy" in simple terms and give a brief example.

"Additional realism is added when the motor policy network is constrained by a connectome dataset. However, many biophysical parameters for individual neurons and synapses remain un-measured."

"motor policy network" is confusing; this is referring to the entire network model here, presumably.

(3) Methods

"We used the adult hermaphrodite *C. elegans* nematode connectome dataset [15, 16, 5], including the identities of its 302 neurons and their synapses (Fig. 1A)."

We believe the authors should specify the dataset type, which is a structural, unsigned connectome lacking grounding in physiological function.

"The policy network was trained in closed loop using PPO as implemented by MIMIC-MJX"

The authors should define "PPO" and "MIMIC-MJX" in simple terms and explain why they were used.

(4) Discussion

"Its role in the movement policy could be fulfilled equally well by a randomly connected RNN, akin to reservoir computing [20], since all the learning happens in the black-box ANN motor decoder."

See above - this computational exercise should actually be performed.

"Looking further ahead, swapping brain and body models of related species may one day yield real insights into how their brains and bodies diverged through evolution. However, far more model development and experimental validation is needed before we can learn anything from such a digital sphinx."

These two sentences about future possible cross-species chimeras feel superfluous and unsubstantiated, and weaken the main argument of the paper about whole-brain emulation.

<https://doi.org/10.7554/eLife.111516.1.sa2>

Reviewer #2 (Public review):

Summary:

The authors use DRL to train a *C. elegans* connectome-based ANN to control stepping in a *D. melanogaster* body model. The resulting system can walk. This shows that one needs further constraints to derive biologically meaningful results from this approach.

Strengths:

The authors perform a very simple experiment with a clear outcome. The interpretation (or lack of interpretation) is a striking cautionary tale.

Weaknesses:

There is little analysis of precisely how robust this result is to parameter variation and network wiring. The worm also undulates in an oscillatory fashion. Thus, it is possible that the network is tapping into biologically meaningful motifs to generate oscillations for walking. As well, it would be useful to examine which heuristics one can use to determine whether modeling efforts are sufficiently constrained (i.e., how much biological data will be necessary to start obtaining fruitful, interpretable outcomes from DRL task optimization). For example, their "solution" using the worm connectome is not sparse (i.e., it uses many neurons). Perhaps a signature of a biologically-meaningful, interpretable result is one that is sparse?

<https://doi.org/10.7554/eLife.111516.1.sa1>

Reviewer #3 (Public review):

Summary:

The authors construct a computational chimera by attaching a *C. elegans* connectome to a *Drosophila* body biomechanical model and use deep reinforcement learning to link neural activity to motor output. The model is able to produce walking, but is considered a priori to be scientifically meaningless, and the work is treated as a cautionary tale in complex interpretation layers unconstrained by experiment or data.

Strengths:

In a period of increasing excitement about linking AI and neuroscience, I respect very much that the authors work through a nontrivial example of nonsense results, rather than just making a theoretical case. It offers a clear and memorable existence proof that matching outputs of complex trained networks does not mean the internal dynamics are themselves emulated.

Weaknesses:

While I understand that the work was a rapidly produced comment on science-by-press-release, the message seems too important to be treated in quite as pithy a manner as it is. In particular, because the computational experiment is so memorable, it is worth getting the message right to avoid a set of readers who take from it that they should dismiss this category of neuroAI wholesale (which the authors absolutely do not imply!).

One part of me reads this work and thinks that by intentionally wiring up the sensory feedback in a particularly nonsense way, the authors have just made a bad model, and sometimes bad models can still generate sensible outputs, especially when expressive models are optimized to fit those sensible outputs. But I think this work is trying to say something more specific than this, and I would like it to be a bit clearer about that. The authors do a fairly good job of sharing a view about what should have been done instead, but this message would benefit from having some more concrete suggestions to avoid a simplistic interpretation. A few thoughts:

(1) It's not entirely obvious to me that the model is "scientifically meaningless." As the authors know extremely well, *Drosophila* walking is thought to be driven by simple central pattern generators coupled to leg-specific implementations. The *C. elegans* neural circuit is clearly capable of producing rhythmic activity as well. A version of the model they ran could have identified biologically valid rhythmic activity in the *C. elegans* circuit and mapped it via the DRL to the right locomotor behavior in the fly. While this would not be a good emulation of the fly, it's not a concept devoid of scientific meaning. Similarly, if the ANN is converting a rhythmic signal to coordinated walking, it's not obvious to me that there aren't useful principles to identify in how it achieves this - it's basically the equivalent of that post-CPG circuitry, no?

(2) Similarly, is this outcome going to be relatively specific to rhythmic behaviors? I suspect that it would be harder to push the *C. elegans* connectome to produce some behaviors than others - for example, adding in visual navigation and other motor patterns, or a ring attractor. Rhythmic circuits arise in many places, and both biology and dynamical systems tell us they can come from numerous configurations of elements and interactions.

(3) Aside from the nonsense formulation of the problem, I would have liked to know more about what the authors should have done to know their model was useless. Put another way, if the authors hadn't known that their model was bad from the beginning (e.g., if they had stuck a fly brain in the middle of it, gotten the sensory feedback right), would there have been some way to figure out if it was meaningful or meaningless based on the results of the trained model itself?

<https://doi.org/10.7554/eLife.111516.1.sa0>