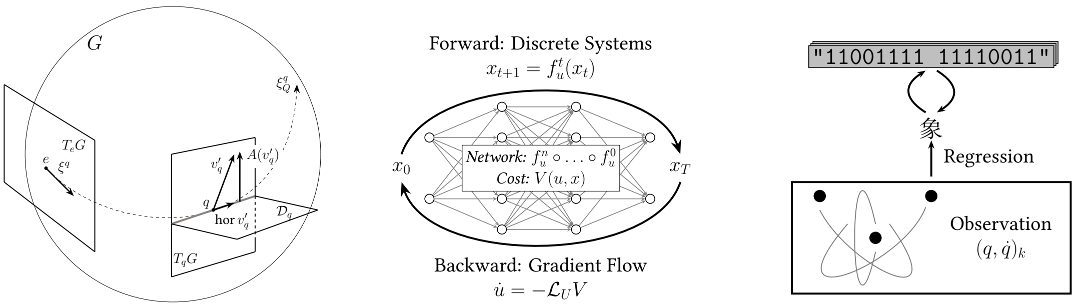
Baiyue Wang baiyuew2@illinois.edu

Research Statement

June 28, 2026

Most of my previous research experience has been in robotics, where I worked on projects involving mechanics, control theory, and dynamical systems. During this time, I also witnessed the rise of AI in almost every aspect enabled by the growth of computational power. Yet many modern AI models remain "non-transparent boxes" where we lack understanding. This motivated my overarching research goal: to bridge mathematical structures with machine learning, so that we can not only leverage the power of AI but also know why it works.

The first direction I am passionate about is *Physically Inspired Machine Learning*, which describes how dynamical systems can be learned from data inspired by theories of differential equations and geometry. This naturally leads to my second area of interest, *Neural Dynamical Systems* where I seek to incorporate geometric structures and dynamical system properties into network design. More generally, I outline additional open topics in the section *Future Research Directions*, where I welcome interdisciplinary collaboration. This include exploring how structure-preserving architectures can enhance engineering systems such as robotics, computer vision and language models, and investigate how logic and symbolic representations arise and interact within machine learning.



(a) Physically Inspired ML (b) Neural Dynamical Systems (c) Future Research Directions

Figure 1: My research interests in three parts.

Physically Inspired Machine Learning

My research background in robotics has provided me with significant insights into mechanics and dynamical systems. From mechanics of material [1] (my undergraduate research where we modeled the mechanics of a series of volumetric soft actuator), to Lagrangian mechanics (my undergraduate thesis in modeling and simulating an overconstrained legged robot based on Lagrangian mechanics), to computational dynamical systems [2]. In this work, I learned about how to represent the evolution of a system by discretizing the dynamics and over-approximate reachable sets with zonotopes, and then optimize and plan trajectories with constraints provided by obstacles. These experiences motivated my shift toward studying mathematical representations of physical systems.

So I took a further step into control systems, differential equations, and geometric mechanics. I developed my interest in understanding systems within the framework of differential geometry [3, 4], which also led to my interest in geometric machine learning. After studying *Nonholonomic Mechanics*, I became fascinated by the challenges that nonholonomic constraints pose in machine learning. Unlike constraints on the configuration space, they do not generate a Lagrangian submanifold. Motivated by this challenge, I conducted independent research on learning nonholonomic systems [5]. This work parameterized the unknown constraint of the system, rearranged the dynamics into Hamel's formalism, and trained the dynamics in a neural ODE fashion. The learned dynamics respects the momentum equation, and preserves symmetry over a chosen subgroup (Figure 1 (a) illustrates a nonholonomic system on Lie group G).

Looking ahead, I am interested in expanding this research, such as learning with discretized systems on the entire vector field [6], and developing systematic approaches to learning Lagrangian and Hamiltonian systems [7, 8], PDEs [9, 10], and systems with dissipation [11].

Neural Dynamical Systems

Another question of interest is, *if a neural network can learn a dynamical system, can a dynamical system represent a neural network?* As models scale up in size, this question becomes critical in interpretability, design and control. In our work [5], we incorporated Lie bracket structures in parameterized Hamel equations to describe systems that evolve on a moving basis, and proved that there is a local minimum for the training problem. I am interested in exploring more connections between the two subjects.

There are already different intersections. For example, if we look at the forward direction of a neural network, each layer can be represented by a flow f, the number of layers can be seen as a discretized time step, and the weights are control inputs to the function [12] (Figure 1 (b)). Early work has shown that stable structures such as symplecticity in *Residual Networks* are discovered by treating the network as a Hamiltonian system. There are also network designs that directly learn conservative systems [13, 14] using symplectic structures. More recent architectures, such as the *Transformer* network, can be interpreted as interactive particle systems [15]. Topological properties of deep neural networks were also found to parallel the differential-geometric *Ricci flow* [16, 17]. I am also curious to find out infinite dimensional geometric structure [18] in operator learning.

From a training perspective, weight updates in backpropagation generally take the form of gradient flow $\dot{u} = -\nabla V_u = -\mathcal{L}_U V$, where u denotes the weight of the network, and V_u is the loss. Therefore, one might ask how to design a network with the steepest gradient descent [19], with safety features [20], or a network that converges stably to its minima [21, 22]. Answering these research questions will help improve network efficiency.

Future Research Directions

Although the previous sections focus on theoretical questions, these ideas have numerous engineering applications. For example, with symmetry in deep learning, one can create an equivariant robot policy [23], a diffusion policy [24], an efficient reinforcement learning policy [25]. What network structures from the previous sections can we use to enhance efficiency and interpretability in *reinforcement learning*, *imitation learning*, or *vision-language models* is an interesting interdisciplinary research topic.

I have also maintained a strong interest in symbolic computation (Figure 1 (c)) (during my graduate research, I developed a symbolic dynamics program to compute robot dynamics for system identification). While working on learning nonholonomic systems [5], I realized that I was solving a symbolic question (finding the right equation) numerically (with neural network output dynamics). Indeed, not only math symbols are symbols, human languages are also symbols–evidence includes that words are usually represented by vectors in language models [26]; some purely symbolic attention models have demonstrated mathematician-like problem solving abilities [27]; some symbolic regression could help rediscovering physical functions [28]. We constructed our ideas through our languages and symbols, and we think in these symbols.

Therefore, I am curious to explore the potential relations between dimensionality reduction and symbolic representation [29, 30]. Furthermore, this research question can help to develop insights in neuro-symbolic schemes [31] in machine learning.

References

[1] **B. Wang**, W. Guo, S. Feng, Y. Hongdong, F. Wan, and C. Song, "Volumetrically enhanced soft actuator with proprioceptive sensing," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5284–5291, 2021.

[2] J. Michaux, P. Holmes, B. Zhang, C. Chen, **B. Wang**, S. Sahgal, T. Zhang, S. Dey, S. Kousik, and R. Vasudevan, "Can not touch this: Real-time, safe motion planning and control for manipulators under uncertainty," *IEEE Transactions on Robotics*, vol. 41, pp. 4719–4740, 2025.

[3] V. I. Arnold, *Mathematical methods of classical mechanics.* Springer-Verlag, 1978.

[4] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, ser. Mathematical Physics Monograph Series. W. A. Benjamin, 1967.

[5] **B. Wang** and A. Bloch, "Learning nonholonomic dynamics with constraint discovery," 2025. [Online]. Available: https://arxiv.org/abs/2410.15201

[6] M. D. Hansen, E. Celledoni, and B. K. Tapley, "Learning mechanical systems from real-world data using discrete forced lagrangian dynamics," 2025. [Online]. Available: https://arxiv.org/abs/2505.20370

[7] C. Eldred, F. Gay-Balmaz, and V. Putkaradze, "Variational neural networks for observable thermodynamics (v-nots)," 2025. [Online]. Available: https://arxiv.org/abs/2509.09899

[8] K. Lee, N. A. Trask, and P. Stinis, "Machine learning structure preserving brackets for forecasting irreversible processes," 2021. [Online]. Available: https://arxiv.org/abs/2106.12619

[9] J. Bruna, B. Peherstorfer, and E. Vanden-Eijnden, "Neural galerkin schemes with active learning for high-dimensional evolution equations," *Journal of Computational Physics*, vol. 496, p. 112588, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021999123006836

[10] A. Goertzen, S. Tang, and N. Azizan, "Eco: Energy-constrained operator learning for chaotic dynamics with boundedness guarantees," 2025. [Online]. Available: https://arxiv.org/abs/2512.01984

[11] A. Bloch, P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu, "The euler-poincaré equations and double bracket dissipation," *Communications in Mathematical Physics*, vol. 175, no. 1, pp. 1–42, Jan 1996. [Online]. Available: https://doi.org/10.1007/BF02101622

[12] A. M. Bloch, P. E. Crouch, and T. S. Ratiu, "Symmetric discrete optimal control and deep learning," 2024. [Online]. Available: https://arxiv.org/abs/2404.06556

[13] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis, "Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems," 2020. [Online]. Available: https://arxiv.org/abs/2001.03750

[14] P. Jin, Y. Tang, and A. Zhu, "Unit triangular factorization of the matrix symplectic group," 2020. [Online]. Available: https://arxiv.org/abs/1912.10926

[15] B. Geshkovski, C. Letrouit, Y. Polyanskiy, and P. Rigollet, "A mathematical perspective on transformers," 2024. [Online]. Available: https://arxiv.org/abs/2312.10794

[16] A. Baptista, A. Barp, T. Chakraborti, C. Harbron, B. D. MacArthur, and C. R. S. Banerji, "Deep learning as ricci flow," 2024. [Online]. Available: https://arxiv.org/abs/2404.14265

[17] M. Hehl, M. von Renesse, and M. Weber, "Neural feature geometry evolves as discrete ricci flow," 2025. [Online]. Available: https://arxiv.org/abs/2509.22362

[18] A. Kriegl and P. W. Michor, *The Convenient Setting of Global Analysis*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 1997, vol. 53.

[19] A. M. Bloch, "Steepest descent, linear programming, and Hamiltonian flows," in *Mathematical developments arising from linear programming (Brunswick, ME, 1988)*, ser. Contemp. Math. Amer. Math. Soc., Providence, RI, 1990, vol. 114, pp. 77–88. [Online]. Available: https://doi.org/10.1090/conm/114/1097866

[20] P. Mestres, A. Marzabal, and J. Cortés, "Anytime safe reinforcement learning," 2025. [Online]. Available: https://arxiv.org/abs/2504.16417

[21] S. G. Gessow and B. T. Lopez, "Analysis of the geometric heat flow equation: Computing geodesics in real-time with convergence guarantees," 2025. [Online]. Available: https://arxiv.org/abs/2510.11692

[22] A. C. B. de Oliveira, L. Cui, and E. D. Sontag, "Remarks on the polyak-lojasiewicz inequality and the convergence of gradient systems," 2025. [Online]. Available: https://arxiv.org/abs/2503.23641

[23] H. Huang, D. Wang, A. Tangri, R. Walters, and R. Platt, "Leveraging symmetries in pick and place," *The International Journal of Robotics Research*, vol. 43, no. 4, pp. 550–571, 2024. [Online]. Available: https://doi.org/10.1177/02783649231225775

[24] D. Wang, S. Hart, D. Surovik, T. Kelestemur, H. Huang, H. Zhao, M. Yeatman, J. Wang, R. Walters, and R. Platt, "Equivariant diffusion policy," 2024. [Online]. Available: https://arxiv.org/abs/2407.01812

[25] J. Welde, N. Rao, P. Kunapuli, D. Jayaraman, and V. Kumar, "Leveraging symmetry to accelerate learning of trajectory tracking controllers for free-flying robotic systems," 2025. [Online]. Available: https://arxiv.org/abs/2409.11238

[26] Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. MIT Press, 2000.

[27] A. Alfarano, F. Charton, and A. Hayat, "Global lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers," 2024. [Online]. Available: https://arxiv.org/abs/2410.08304

[28] P. Lemos, N. Jeffrey, M. Cranmer, S. Ho, and P. Battaglia, "Rediscovering orbital mechanics with machine learning," *Machine Learning: Science and Technology*, vol. 4, no. 4, p. 045002, oct 2023. [Online]. Available: https://dx.doi.org/10.1088/2632-2153/acfa63

[29] M. D. Kvalheim and E. D. Sontag, "Why should autoencoders work?" 2024. [Online]. Available: https://arxiv.org/abs/2310.02250

[30] Y. P. Wotte, P. Buchfink, S. Glas, F. Califano, and S. Stramigioli, "Model order reduction via lie groups," 2025. [Online]. Available: https://arxiv.org/abs/2511.03520

[31] C. Wang, K. Ji, J. Geng, Z. Ren, T. Fu, F. Yang, Y. Guo, H. He, X. Chen, Z. Zhan, Q. Du, S. Su, B. Li, Y. Qiu, Y. Du, Q. Li, Y. Yang, X. Lin, and Z. Zhao, "Imperative learning: A self-supervised neuro-symbolic learning framework for robot autonomy," *The International Journal of Robotics Research*, p. 02783649251353181, 2025. [Online]. Available: https://doi.org/10.1177/02783649251353181