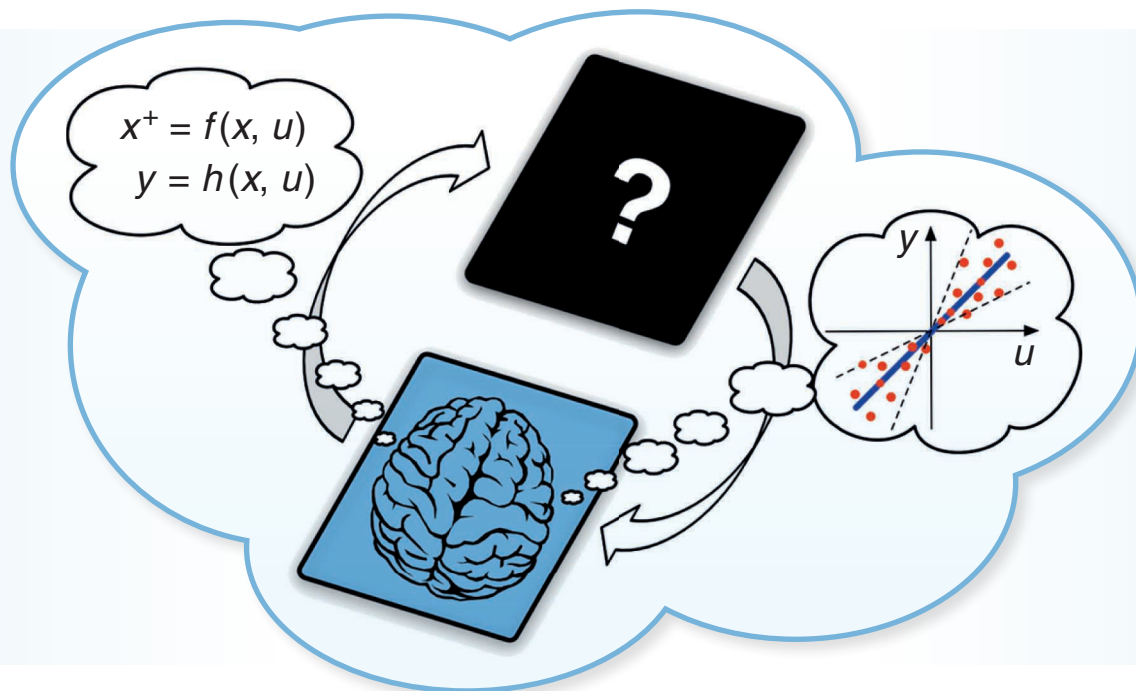


Data-Driven Control: Part Two of Two

HOT TAKE: WHY NOT GO WITH MODELS?



FLORIAN DÖRFLER^{1b}

A recurring question that all authors of this special issue encounter is, “Why not go with models?” Two terms need to be clarified: In this context, a *model* is understood as a parametric system representation often endowed with an interpretable structure, for example, a state-space representation with a readily discernible $F = m \cdot a$ equation. Further, the term *data-driven control*, as we employ it in this special issue, is not just about using data from a black box to inform decision making. Researchers are exploring different paradigms, among others, model-based control design, where the model and uncertainty estimates are learned from data using contemporary system identification and uncertainty quantification techniques. In classical adaptive control terminology [1], [2], this two-stage approach is referred to as *indirect*. In contrast,

direct data-driven control bypasses models in the decision making; see Figure 1 for a graphical illustration of the two paradigms. Hence, the more precise question should be, “When should we embrace direct or indirect data-driven control?” I will delve into the expected “it depends” answer in this “Editorial” column.

The typical answers one encounters often reason with the shortcomings of system identification, take the vantage point of a specific application, or quote the widely recognized success of reinforcement learning in computer and board games.

Let me put good automatic control practice into use, abstract the problem, and see more clearly through the lens of mathematics. Since the minimization of a fitting criterion is the prevalent formulation in system identification, I will take an optimization perspective. Allow me to leave the uncertainty aspect aside for now. In this idealized setup, the indirect paradigm—first identify a model within a

prespecified class, then perform model-based control—is abstracted as a nested optimization problem:

indirect data-driven control	
minimize	control cost (u, y)
subject to	trajectory (u, y) compatible with the model
where	model \in argmin fitting criterion (u^d, y^d) subject to model belongs to a certain class

The indirect approach can be first and foremost described as being *modular* with two well-separated levels. In comparison, the direct approach is more lean and seeks a decision compatible with the data. Abstractly, it is an *end-to-end* monolithic problem:

direct data-driven control	
minimize	control cost (u, y)
subject to	trajectory (u, y) compatible with data (u^d, y^d)

While the direct versus indirect classification is simple and useful, the line between both paradigms is often blurred. Let me now dive into the relative merits of both formulations.

ON MODELS

The distinguishing feature of both approaches is whether to use a model for decision making. Compared with raw data, models are tidied-up representations, that is, compressed, denoised, and typically also approximate. These features are most obvious in subspace system identification, where they are achieved by singular value thresholding of data matrices [3]. Crucially, models are interpretable, often structured, and physically intuitive; for example, think of the class of port-Hamiltonian systems [4]. Further, models are amenable to powerful control design methods (for example, semidefinite programming for optimal control in state space [5]), and they are obvi-

ously useful beyond control, for example, for analysis, simulation, or system design. Hence, there will always be a place for models and indirect design.

However, there are indeed arguments to be made against models. Think about systems with complex physics (for example, soft robotics), complex disturbances (for example, wind farms), or complex sensing modalities (for example, perception-based control) or operating in complex environments (for example, autonomous cars). In such instances, even if first-principle models were available, they might be too complex to be useful for control design, and one may argue for the direct approach. This brings to light that “the main issue in modeling from data is approximation” [6]. To take it a step further, rather than modeling the possibly complex data-generating mechanism, the data *should* be used to directly inform the decision making, as argued in [7], [8], and [9]. Of course, the “should” is debated. More concretely, the ultimate objective is the control policy, and often, it might be easier to learn a control policy than to learn a model. This catchy statement has often been voiced recently [10], and some historic examples include the widely deployed Ziegler–Nichols PID tuning from a single system response [11] or finite-time optimal control design based on a few Markov parameters [12] or step responses [13]. Further, while persistence of excitation is a necessary identifiability condition in the indirect approach, unstable systems can sometimes be directly stabilized but not identified with limited data [14].

I close this point with a famous quote from [9]: “When solving a given problem, try to avoid solving a more general problem as an intermediate step.” I also provide a simple disarming example: Consider asymptotic rejection of a constant disturbance. Indirect approaches require estimates of the model and the disturbance, whereas a direct approach is simply integral control.

LACK OF SEPARATION PRINCIPLE

In general, the indirect approach might be suboptimal since there is no separation principle between identification and control; that is, the best model fitting the data may not be the best model for the ultimate control task. For example, in classical frequency domain control, the dc gain and closed-loop bandwidth are often the two key control specifications, which should again inform at which frequencies to mainly identify the system. This lack of a separation principle stimulated many approaches blending the inner identification and outer control objectives, for example, dual control [15], [16], identification for control [17], [18], or approaches blending the identification and control objectives [19], [20], [21], [22]. However, in

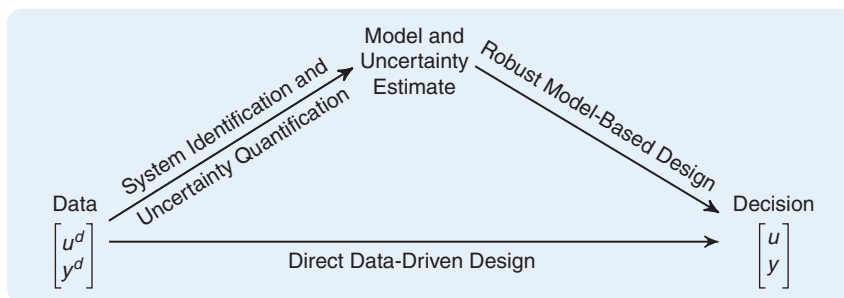


FIGURE 1 Direct and indirect paths from input/output data (u^d, y^d) to deciding upon control and output trajectories u and y . In practice, the design is often iterative (that is, the diagram contains loops) and involves further processing steps, such as filtering of the raw data.

Compared with raw data, models are tidied-up representations, that is, compressed, denoised, and typically also approximate.

some settings, indirect design via system identification is optimal, as discussed next.

ON SYSTEM IDENTIFICATION

The inner system identification problem serves two crucial roles. First, system identification filters the data and reduces variance. If a parameterization of the data generation mechanism is known, an indirect approach based on maximum-likelihood estimation and certainty-equivalence design is optimal in mean square error [18] or regret metrics (for online LQR settings) [23]. Second, one can also interpret system identification as a projection of the data on a specific model class, which is either a priori known or obtained in preprocessing. Indeed, this innermost model selection step is often hard but injects prior knowledge, structure, and any sort of physical bias, such as stability, dissipativity, or positivity (see, for example, approaches on kernel-based system identification [24], [25], [26]).

Model selection is evidently absent in the direct approach, which makes it harder to include side information on the plant. Conversely, no bias error can be incurred due to an erroneously selected model class or inconsistent parameter estimates. While this is not a universal statement (for example, some noniterative direct model reference approaches may also induce an undesired bias), it has been illustrated for specific settings in [27], [28], [29], [30]. In a nutshell, direct approaches with imperfect learning can be more robust. Especially in adaptive control settings, it is argued that “the indirect approach [aiming at model matching] is motivated by optimality and the direct approach [aiming at output matching] is motivated by stability” [31]. However, obviously, no universal conclusions can be drawn for all methods, especially since the direct versus indirect distinction is blurry to begin with. In my opinion, the robustness and bias–variance tradeoffs of both approaches are still to be fully explored and quantified.

On a final note, system identification in practice is often an art and cumbersome and may require readaptation. Further, it is argued in [18] and confirmed by the author’s experience that the engineering cost due to modeling, identification, and (re)commissioning is significant. The industrial report [32] bluntly concludes that “about the only place the cost of dynamic modeling is ever warranted is during MPC implementation.”

ON UNCERTAINTY

Let me quote [33]: “The most outstanding point of [direct] approaches is that the twinborn problem of unmodeled

dynamics and robustness in traditional [model-based] theory do not exist under [the direct] framework.” I illustrate this insight with an example: Consider a batch of noisy data, fed through nonconvex prediction error identification and yielding a nominal model of certain order together with a stochastic parametric uncertainty estimate accounting for both noise as well as unmodeled (for example, higher-order) dynamics. In the indirect approach, these are then later used for optimal \mathcal{H}_∞ -control subject to an unstructured worst-case uncertainty. Observe the *twinborn* problem of the control design having to be robust due to the model being only approximate. Further, in the indirect approach, uncertainty on the data needs to be propagated through the inner system identification problem, and the identification uncertainty estimates may be incompatible with the uncertainty quantification preferred for control design, as in the above example.

In comparison, direct approaches incorporate uncertainty on the data directly in the control design [34], [35], [36], and they do so in a transparent way without propagating it through system identification or approximating uncertainty estimates.

COMPLEXITY AND IMPLEMENTATION

There are many other arguments to be brought up. For instance, there is generally no winning approach in terms of “complexity”; that is, there are instances of both methods that are data-efficient (or wasteful), that are analytically (in)tractable, that have (non)convex problem formulations, that can be carried out partially or entirely offline (online), that are easier (harder) to code or debug, and that require more or less human supervision.

While on the topic of implementation, the indirect approach is obviously modular with well-understood subtasks, which makes it a reliable and interpretable building block in a layered architecture. This is in stark contrast to the monolithic nature of the direct approach. Researchers in academia, industry, and application domains are arguing about the relative merits. For instance, in robotics, this dichotomy is often referred to as *end-to-end* versus a layered *autonomy stack*. Proponents argue in favor of either approach when it comes to implementations, suitability for complex specifications, and all of the previously listed topics. This debate is far from settled; it epitomizes the grand challenge of architecture selection [37], and the infamous *middle road* might be the best path forward (as shown in many robotics implementations blending both paradigms, for example, [38], [39]).

In my opinion, the robustness and bias–variance tradeoffs of both approaches are still to be fully explored and quantified.

THE ARTICLES OF THIS SPECIAL ISSUE

While all articles in this double special issue touch upon the direct versus indirect design questions from different perspectives, they also address numerous further challenges and all have their own unique foci. Specifically, in the present special issue, the articles focus on the informativity of data for different analysis and synthesis tasks and sample complexity estimates for identification and control, as well as automated tuning of hyperparameters in data-driven control design.

The article by van Waarde et al. [A1] puts informativity of the data as the central concept. The authors are interested in the setting when the data are not sufficiently informative to (uniquely) identify a system. Nevertheless, analysis tasks or control design can be carried out for all systems sets that are unfalsified by the data. For noise-free data, the approach leads to a robust theory for affine sets of systems, and for noisy data, the methods draw inspiration from classical robust control.

The article by Tsiamis et al. [A2] surveys recent advances in statistical learning theory relevant to control and system identification. The article focuses on nonasymptotic sample complexity estimates for linear system identification and linear quadratic regulation leveraging modern tools from high-dimensional statistics and learning theory. The tutorial-style and self-contained presentation makes the topic accessible to control engineers. The article also outlines fruitful future directions.

Finally, the article by Breschi and Formentin [A3] addresses the challenges of automatically optimizing hyperparameters, that is, degrees of freedom whose tuning requires significant human expertise. The authors focus on direct data-driven control methods to showcase the potential and the limitations of automatic hyperparameter tuning, namely, virtual reference feedback tuning and data-driven predictive control. They illustrate their strategies with applications from the automotive domain.

SYNOPSIS

I hope to have illuminated the unsatisfying “it depends” answer to my opening question from sufficiently many angles, clarified a few of the paradigms and research gaps, and given food for thought for when to implement either approach. The articles of this special issue each have their own take on this question and develop a rich data-driven control theory.

Let me conclude with a few personal thoughts. Science and engineering based on traditional (that is, first-

principle and parametric) models have brought us far, literally, to the moon and back. However, there are notoriously challenging problems with the traditional approach, and the impact of contemporary data-driven methods will be judged (among others) by whether they will overcome these challenges. For instance, can they take a stab at hard system classes (for example, nonlinear and infinite-dimensional), can they be implemented online with streaming data, and are there applications with a true business case for these methods?

ACKNOWLEDGMENTS

I thank several colleagues that read drafts of this article and provided valuable feedback, among others, all authors of this double special issue and several of my colleagues at ETH Zürich and the NCCR Automation, as well as Andrea Iannelli, Alexandre Sanfelici Bazanella, Claudio De Persis, Anuradha Annaswamy, Ian R. Manchester, Alessandro Chiuso, Håkan Hjalmarsson, and Alireza Karimi. I was hard pressed to include all opinions and different perspectives on this “hot take” topic.

AUTHOR INFORMATION

Florian Dörfler (dorfler@ethz.ch) is an associate professor at the Automatic Control Laboratory at ETH Zürich, 8092 Zürich, Switzerland, and the associate head of the Department of Information Technology and Electrical Engineering.

APPENDIX: RELATED ARTICLES

- [A1] H. J. van Waarde, J. Eising, M. K. Camlibel, and H. L. Trentelman, “The informativity approach: To data-driven analysis and control,” *IEEE Control Syst.*, vol. 43, no. 6, pp. 32–66, Dec. 2023, doi: 10.1109/MCS.2023.3310305.
- [A2] A. Tsiamis, I. Ziemann, N. Matni, and G. J. Pappas, “Statistical learning theory for control: A finite sample perspective,” *IEEE Control Syst.*, vol. 43, no. 6, pp. 67–97, Dec. 2023, doi: 10.1109/MCS.2023.3310345.
- [A3] V. Breschi and S. Formentin, “AutoDDC: Hyperparameter tuning for direct data-driven control,” *IEEE Control Syst.*, vol. 43, no. 6, pp. 98–124, Dec. 2023, doi: 10.1109/MCS.2023.3310368.

REFERENCES

- [1] K. J. Åström and B. Wittenmark, *Adaptive Control*. New York, NY, USA: Dover, 1995.
- [2] A. M. Annaswamy and K. S. Narendra, *Stable Adaptive Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
- [3] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. New York, NY, USA: Springer Science & Business Media, 2012.
- [4] A. Van Der Schaft et al., “Port-Hamiltonian systems theory: An introductory overview,” *Found. Trends® Syst. Control*, vol. 1, no. 2, pp. 173–378, 2014, doi: 10.1561/2600000002.
- [5] C. Scherer and S. Weiland, “Linear matrix inequalities in control,” *Lecture Notes*, vol. 3, no. 2, 2000.

[6] J. C. Willems, "In control, almost from the beginning until the day after tomorrow," *Eur. J. Control*, vol. 13, no. 1, 2007, pp. 71–81, doi: 10.3166/ejc.13.71-81.

[7] L. Breiman, "Statistical modeling: The two cultures," *Statistical Sci.*, vol. 16, no. 3, pp. 199–231, Aug. 2001, doi: 10.1214/ss/1009213726.

[8] M. C. Campi, A. Carè, and S. Garatti, "The scenario approach: A tool at the service of data-driven decision making," *Annu. Rev. Control*, vol. 52, pp. 1–17, 2021, doi: 10.1016/j.arcontrol.2021.10.004.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer Science & Business Media, 1999.

[10] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 2, no. 1, pp. 253–279, 2019, doi: 10.1146/annurev-control-053018-023825.

[11] J. G. Ziegler and N. B. Nichols, "Optimum settings for automatic controllers," *Trans. Amer. Soc. Mech. Eng.*, vol. 64, no. 8, pp. 759–765, Nov. 1942, doi: 10.1115/1.4019264.

[12] G. Shi and R. E. Skelton, "Markov data-based LQG control," *J. Dyn. Syst., Meas., Control*, vol. 122, no. 3, pp. 551–559, Sep. 2000, doi: 10.1115/1.1286868.

[13] C. R. Cutler and B. L. Ramaker, "Dynamic matrix control – A computer control algorithm," in *Proc. Joint Autom. Control Conf.*, 1980, vol. 17, p. 72.

[14] H. J. Van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel, "Data informativity: A new perspective on data-driven analysis and control," *IEEE Trans. Autom. Control*, vol. 65, no. 11, pp. 4753–4768, Nov. 2020, doi: 10.1109/TAC.2020.2966717.

[15] A. Feldbaum, "Dual control theory problems," *IFAC Proc. Volumes*, vol. 1, no. 2, pp. 541–550, Jun. 1963, doi: 10.1016/S1474-6670(17)69687-3.

[16] B. Wittenmark, "Adaptive dual control methods: An overview," *IFAC Proc. Volumes*, vol. 28, no. 13, pp. 67–72, Jun. 1995, doi: 10.1016/S1474-6670(17)69687-3.

[17] M. Gevers, "Identification for control: From the early achievements to the revival of experiment design," *Eur. J. Control*, vol. 11, nos. 4–5, pp. 335–352, 2005, doi: 10.3166/ejc.11.335-352.

[18] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, Mar. 2005, doi: 10.1016/j.automatica.2004.11.021.

[19] M. C. Campi and P. Kumar, "Adaptive linear quadratic Gaussian control: The cost-biased approach revisited," *SIAM J. Control Optim.*, vol. 36, no. 6, pp. 1890–1907, 1998, doi: 10.1137/S0363012997317499.

[20] S. Formentin and A. Chiuso, "Control-oriented regularization for linear system identification," *Automatica*, vol. 127, May 2021, Art. no. 109539, doi: 10.1016/j.automatica.2021.109539.

[21] F. Dörfler, P. Tesi, and C. De Persis, "On the certainty-equivalence approach to direct data-driven LQR design," *IEEE Trans. Autom. Control*, early access, 2023, doi: 10.1109/TAC.2023.3253787.

[22] L. Campestrini, D. Eckhard, A. S. Bazanella, and M. Gevers, "Data-driven model reference control design by prediction error identification," *J. Franklin Inst.*, vol. 354, no. 6, pp. 2628–2647, Apr. 2017, doi: 10.1016/j.jfranklin.2016.08.006.

[23] M. Simchowitz and D. Foster, "Naive exploration is optimal for on-line LQR," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2020, pp. 8937–8948.

[24] M. Khosravi, "Side-information in linear and nonlinear system identification," Ph.D. dissertation, ETH Zurich, Zürich, Switzerland, 2021.

[25] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models From Data*. Cham, Switzerland: Springer Nature, 2022.

[26] H. J. van Waarde and R. Sepulchre, "Kernel-based models for system analysis," *IEEE Trans. Autom. Control*, vol. 68, no. 9, pp. 5317–5332, Sep. 2023, doi: 10.1109/TAC.2022.3218944.

[27] A. Becker, P. Kumar, and C.-Z. Wei, "Adaptive control with the stochastic approximation algorithm: Geometry and convergence," *IEEE Trans. Autom. Control*, vol. 30, no. 4, pp. 330–338, Apr. 1985, doi: 10.1109/TAC.1985.1103963.

[28] V. Krishnan and F. Pasqualetti, "On direct vs indirect data-driven predictive control," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Piscataway, NJ, USA: IEEE Press, 2021, pp. 736–741, doi: 10.1109/CDC45484.2021.9683187.

[29] S. Formentin, K. Van Heusden, and A. Karimi, "A comparison of model-based and data-driven controller tuning," *Int. J. Adaptive Control Signal Process.*, vol. 28, no. 10, pp. 882–897, Oct. 2014, doi: 10.1002/acs.2415.

[30] F. Dörfler, J. Coulson, and I. Markovskiy, "Bridging direct and indirect data-driven control formulations via regularizations and relaxations," *IEEE Trans. Autom. Control*, vol. 68, no. 2, pp. 883–897, Feb. 2023, doi: 10.1109/TAC.2022.3148374.

[31] A. M. Annaswamy, "Adaptive control and intersections with reinforcement learning," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 6, no. 1, pp. 65–93, 2023, doi: 10.1146/annurev-control-062922-090153.

[32] L. Desborough and R. Miller, "Increasing customer value of industrial control performance monitoring-Honeywell's experience," in *Proc. AIChE Symp. Ser.*, New York, NY, USA: American Institute of Chemical Engineers, 1998, vol. 326, pp. 169–189.

[33] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Inf. Sci.*, vol. 235, pp. 3–35, Jun. 2013, doi: 10.1016/j.ins.2012.07.014.

[34] A. Bisoffi, C. De Persis, and P. Tesi, "Data-driven control via Petersen's lemma," *Automatica*, vol. 145, Nov. 2022, Art. no. 110537, doi: 10.1016/j.automatica.2022.110537.

[35] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: Nonconservative design via a matrix S-lemma," *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 162–175, Jan. 2020, doi: 10.1109/TAC.2020.3047577.

[36] J. Berberich, C. W. Scherer, and F. Allgöwer, "Combining prior knowledge and data for robust controller design," *IEEE Trans. Autom. Control*, vol. 68, no. 8, pp. 4618–4633, Aug. 2023, doi: 10.1109/TAC.2022.3209342.

[37] A. M. Annaswamy, K. H. Johansson, and G. J. Pappas, Eds., *Control for Societal-Scale Challenges: Road Map 2030*. Piscataway, NJ, USA: IEEE Press, 2023. [Online]. Available: <https://ieeecs.org/control-societal-scale-challenges-roadmap-2030>

[38] M. P. Poverini, S. Formentin, L. Merzagora, and P. Rocco, "Mixed data-driven and model-based robot implicit force control: A hierarchical approach," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 4, pp. 1258–1271, Jul. 2020, doi: 10.1109/TCST.2019.2908899.

[39] R. T. Fawcett, K. Afsari, A. D. Ames, and K. A. Hamed, "Toward a data-driven template model for quadrupedal locomotion," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 7636–7643, Jul. 2022, doi: 10.1109/LRA.2022.3184007.



IEEE connects you to a universe of information!

As the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity, the IEEE and its Members inspire a global community through its highly cited publications, conferences, technology standards, and professional and educational activities.

Visit www.ieee.org.

IEEE

Publications / IEEE Xplore® / Standards / Membership / Conferences / Education

IMAGE LICENSED BY INGRAM PUBLISHING