DECISION-AWARE UNCERTAINTY-CALIBRATED DEEP LEARNING FOR ROBUST ENERGY SYSTEM OPERATION

Christopher Yeh, Nicolas Christianson, Steven Low, Adam Wierman & Yisong Yue
Department of Computing and Mathematical Sciences
California Institute of Technology
Pasadena, CA 91125, USA
{cyeh,nchristi,slow,adamw,yyue}@caltech.edu

ABSTRACT

Decision-making under uncertainty is an important problem that arises in many domains. Achieving robustness guarantees requires well-calibrated uncertainties, which can be difficult to achieve in high-capacity prediction models such as deep neural networks. This paper proposes an end-to-end approach for learning uncertainty-calibrated deep learning models that directly optimizes a downstream decision-making objective with provable robustness. We also propose two concrete applications in energy system operations, including a grid scheduling task as well as an energy storage arbitrage task. As renewable wind and solar generation increasingly proliferate and their variability penetrates the energy grid, learning uncertainty-aware predictive models becomes increasingly crucial for maintaining efficient and reliable grid operation.

1 INTRODUCTION

Many real-world energy systems rely on estimated or forecasted quantities for decision-making. In the electric grid, for example, operators schedule generators hours ahead of time to satisfy predicted future electricity demand (Conejo & Sioshansi, 2018). Furthermore, because predictions are imperfect, grid operators benefit from knowing predictive uncertainty, which enables operators to trade off between scheduling costs and the risk of grid instability (or worse, blackouts). Especially as renewable wind and solar energy generation increase the variability of net electricity demand (*i.e.*, total demand minus renewable generation), calibrated uncertainty estimates are vital to reliable grid operation. Likewise, grid-scale energy storage systems rely on predictions of future electricity prices to profit from price arbitrage, storing energy when electricity prices are low, and discharging when prices are high. Since energy storage helps increase utilization of low-carbon generation (de Sisternes et al., 2016), improving storage operation profitability may help reduce carbon emissions.

Deep learning models have made tremendous advances across a variety of tasks, including both electricity demand and price forecasting (Bedi & Toshniwal, 2019; Lago et al., 2018). However, deep learning models are poor at estimating their own uncertainty(Guo et al., 2017; Kuleshov et al., 2018), limiting their utility for downstream decision-making. While recent works such as isotonic regression (Kuleshov et al., 2018) and conformal prediction (Shafer & Vovk, 2008) have made advances in calibrating uncertainty estimates from deep learning models, such methods are generally not differentiable and not part of the training process. Furthermore, predictive models are rarely trained end-to-end with a downstream decision-making objective. Traditional approaches first train a model to predict some quantity (*e.g.*, electricity demand) along with an uncertainty estimate. Then, the predicted quantity and uncertainty (possibly calibrated using a post-hoc method) are used by a downstream controller, but the controller does not provide feedback to the predictive model. Because the downstream objective is often asymmetric with respect to the model's prediction error, the trained model may be suboptimal with respect to the downstream objective.

In this work, we propose training uncertainty-aware deep learning models end-to-end with downstream decision-making objectives. By including approximate calibration layers in our model during training, we remove the need for a post-hoc calibration step. This closes the loop and ensures that feedback from the downstream objective is accounted for in the training process, since not all model

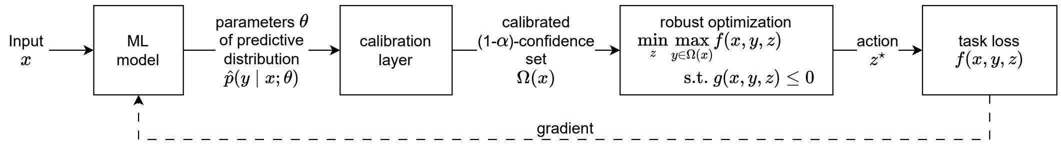

Figure 1: We propose training a machine learning model end-to-end that outputs calibrated uncertainties for solving a robust optimization problem to minimize a task-specific loss.

errors nor uncertainty estimates will result in the same downstream cost. The end-to-end training enables the model to focus its learning capacity on minimizing error and uncertainty on outputs with the largest decision-making cost, with more leeway for outputs that have lower costs.

2 Related Work

Most related to our proposed method is the idea of "task-based end-to-end model learning" introduced by Donti et al. (2017), which trains machine learning models in an end-to-end fashion that directly captures the ultimate task-based objective in which they are used, within the context of stochastic programming. Backpropagating gradients through a stochastic optimization problem is made possible via the implicit function theorem. However, the method proposed by Donti et al. (2017) does not train the model to estimate uncertainty and thereby does not provide any explicit robustness to uncertainty. Our proposed model improves upon this work by outputting calibrated uncertainty sets which can then be used to determine robust decisions.

Various deep learning regression models that provide uncertainty estimates have been proposed, including Bayesian neural networks (Blundell et al., 2015; Gal & Ghahramani, 2016), Gaussian process regression and deep kernel learning (Rasmussen & Williams, 2005; Wilson et al., 2016; Liu et al., 2020), ensembles of models (Lakshminarayanan et al., 2017), and quantile regression (Romano et al., 2019), among other techniques. These methods typically only provide heuristic uncertainty estimates that are not necessarily well-calibrated (Nado et al., 2022).

Post-hoc methods such as isotonic regression (Kuleshov et al., 2018) or conformal prediction (Shafer & Vovk, 2008) may be used to calibrate the uncertainty outputs of deep learning models. These calibration methods generally treat the model as a black box and scale the uncertainty levels such that they are calibrated on a held-out calibration set. Isotonic regression guarantees calibrated outputs in the limit of infinite data, whereas conformal methods provide probabilistic finite-sample calibration guarantees when the calibration set is exchangeable with (*e.g.*, drawn i.i.d. from the same distribution as) test data. These calibration methods are not included in the model training procedure because they involve non-differentiable operators, such as sorting. However, recent works have proposed differentiable losses (Einbinder et al., 2022; Stutz et al., 2022) that approximate the conformal prediction procedure during training and thus allow end-to-end training of models to output more calibrated uncertainty. As approximations, these methods lose the marginal coverage guarantees that true conformal methods provide. However, such guarantees can be recovered at test time by replacing the approximations with true conformal prediction.

There has been significant recent interest in the energy systems community in using techniques such as stochastic, risk-sensitive, chance-constrained, distributionally robust, and robust optimization to schedule grid resources in a manner that is robust to uncertainty (Zheng et al., 2015; Ndrio et al., 2021; Dvorkin, 2020; Zhong et al., 2021; Poolla et al., 2021; Warrington et al., 2012; Bertsimas et al., 2013; Christianson et al., 2022). In these works, the robust optimization methods enable dispatching resources in a manner that is aware of grid uncertainty, *e.g.*, to ensure that sufficient generation will be available to meet demand even on a cloudy day without much solar generation. Typically, however, the construction of uncertainty sets, probability distributions over uncertain demand and renewable generation, or ambiguity sets over distributions takes place offline and is unconnected to the eventual dispatch task. Thus our proposed end-to-end approach will allow for simultaneous calibration of uncertainty sets with optimal decision-making.

3 PROPOSED METHOD

We first describe the abstract problem setting and then give concrete examples for energy systems. Figure 1 graphically depicts how our proposed method works.

Suppose that data $(x \in \mathbb{R}^d,\ y \in \mathbb{R}^k)$ is drawn from an unknown joint distribution p. Upon observing inputs x but not labels y, an agent chooses an action $z \in \mathbb{R}^n$ and incurs *task cost* $f(x, y, z)$, where f is a convex function. The agent's goal is to pick an action z that robustly minimizes the task cost over a $(1-\alpha)$-confidence set $\Omega(x) \subset \mathbb{R}^k$ of possible y values. We additionally allow constraints on z of the form $g(x, y, z) \leq 0$ for some differentiable function $g : \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^h$. Concretely, for each input x, we aim to solve

$$\min_{z \in \mathbb{R}^n} \max_{y \in \Omega(x)} f(x, y, z) \quad \text{s.t. } g(x, y, z) \leq 0. \tag{1}$$

Because we do not know the joint distribution p, we train a machine learning model on samples (x, y) to learn an approximate conditional distribution $\hat{p}(y \mid x)$ representing uncertainty over y for a given input x. From $\hat{p}$, we aim to construct a $(1-\alpha)$-confidence set $\Omega(x)$ of possible y values. To ensure the confidence set is properly calibrated, we propose using conformal prediction methods. Finally, once we have a $(1-\alpha)$-confidence set $\Omega(x)$, we solve the robust convex optimization problem (1) to determine the optimal robust action z, and the agent incurs the task-cost $f(x, y, z)$.

For example, the model may output the mean and covariance parameters $\theta = (\hat{\mu}(x), \hat{\Sigma}(x))$ of a multivariate normal distribution, so that $\hat{p}(y \mid x; \theta) = \mathcal{N}(y \mid \hat{\mu}(x), \hat{\Sigma}(x))$. Then, using conformal methods with a held-out calibration set, we can determine an appropriate factor $\beta(x) \in \mathbb{R}_+$ by which to scale $\hat{\Sigma}(x)$ to output a calibrated $(1-\alpha)$-confidence set

$$\Omega(x) = \{y \in \mathbb{R}^k \mid (y - \hat{\mu}(x))^\top (\beta(x)\hat{\Sigma}(x))^{-1} (y - \hat{\mu}(x)) \leq \chi^2_k(1-\alpha)\} \tag{2}$$

such that $p(y \in \Omega(x) \mid x) \geq 1 - \alpha$. Here, $\chi^2_k(p)$ is the quantile function at probability p for the chi-squared distribution with k degrees of freedom.

Notably, we propose using slightly different procedures between training and testing. Training the model end-to-end to minimize the task cost requires differentiating through both the robust convex optimization problem (which is possible via the implicit function theorem, as shown by Donti et al. (2017)), as well as differentiating through the conformal prediction step (which is possible via the approximation methods of Einbinder et al. (2022)). At test time, to provide a provable robustness guarantee, we can revert to traditional conformal prediction methods instead of the differentiable approximation to obtain a provably calibrated confidence set.

3.1 GRID SCHEDULING

As a concrete example, consider a realistic grid-scheduling task where a power system operator decides electricity generation and storage decisions $z_1, \ldots, z_T$ for each hour in the next T hours based on some (unknown) distribution over net electricity demand. The general robust economic dispatch problem for a single-bus system takes the following form:

$$\min_{z_1, \ldots, z_T \in \mathbb{R}^n} \max_{y \in \Omega(x)} \sum_{t=1}^{T} c_t^\top z_t \quad \text{s.t.} \quad Az \leq b,\ Dz + Ey \leq h$$

where $y \in \Omega(x) \subset \mathbb{R}^T$ is the uncertain vector of net electricity demand for the next T hours. The uncertainty set $\Omega(x)$ may be determined by inputs x such as the current electricity demand, time of day, and weather conditions. The vectors c_t are (possibly time-varying) marginal costs of generation, the constraint $Az \leq b$ includes capacity limits, ramp limits, state of charge constraints, and any other constraints on dispatches that do not involve the uncertain vector y, and the constraint $Dz + Ey \leq h$ includes constraints coupling dispatch decisions with the uncertainty such as supply sufficiency. The min-max form of the problem requires that the optimal dispatch decisions z_t minimize the cost and satisfy all constraints for the worst-case y in the uncertainty set $\Omega(x)$. This semi-infinite program can be tractably reformulated as a convex optimization problem for a variety of choices of uncertainty set $\Omega(x)$ (Ben-Tal et al., 2009).

4 Conclusion and Future Work

Our proposal describes a flexible deep learning method for minimizing task-specific losses with provably robust guarantees. Moreover, our framework can extend beyond downstream robust optimization tasks to stochastic or chance-constrained settings. We will test our method on a grid-scheduling simulation as well as an energy storage arbitrage task (described in Appendix A) based on real historical data from the PJM electric grid (Donti et al., 2017). By increasing the reliability of energy grids under high renewable uncertainty, our method will help facilitate the transition towards 24/7 carbon-free energy.

Acknowledgments

The authors acknowledge support from an NSF Graduate Research Fellowship (DGE-1745301), an Amazon/Caltech AI4Science Fellowship, NSF grants (CNS-2146814, CPS-2136197, CNS-2106403, NGSDI-2105648), Amazon AWS, as well as the Caltech Resnick Sustainability Institute.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0022218. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

We also thank Ivan Jimenez for helpful discussions.

References

Jatin Bedi and Durga Toshniwal. Deep learning framework to forecast electricity demand. *Applied Energy*, 238:1312–1326, March 2019. ISSN 0306-2619. doi: 10.1016/j.apenergy.2019.01.113.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*, volume 28. Princeton University Press, 2009.

Dimitris Bertsimas, Eugene Litvinov, Xu Andy Sun, Jinye Zhao, and Tongxin Zheng. Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem. *IEEE Transactions on Power Systems*, 28(1):52–63, February 2013. ISSN 1558-0679. doi: 10.1109/TPWRS.2012.2205021.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight Uncertainty in Neural Networks. Technical report, May 2015. URL `http://arxiv.org/abs/1505.05424`.

Nicolas Christianson, Lucien Werner, Adam Wierman, and Steven Low. Dispatch-aware planning for feasible power system operation. *Electric Power Systems Research*, 212:108597, November 2022. ISSN 0378-7796. doi: 10.1016/j.epsr.2022.108597.

Antonio J. Conejo and Ramteen Sioshansi. Rethinking restructured electricity market design: Lessons learned and future needs. *International Journal of Electrical Power & Energy Systems*, 98:520–530, June 2018. ISSN 0142-0615. doi: 10.1016/j.ijepes.2017.12.014.

Fernando J. de Sisternes, Jesse D. Jenkins, and Audun Botterud. The value of energy storage in decarbonizing the electricity sector. *Applied Energy*, 175:368–379, August 2016. ISSN 0306-2619. doi: 10.1016/j.apenergy.2016.05.014.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based End-to-end Model Learning in Stochastic Optimization. In *Advances in Neural Information Processing Systems*, volume 30, Long Beach, CA, USA, December 2017. Curran Associates, Inc. doi: 10.48550/arXiv.1703.04529. URL http://arxiv.org/abs/1703.04529.

Yury Dvorkin. A Chance-Constrained Stochastic Electricity Market. *IEEE Transactions on Power Systems*, 35(4):2993–3003, July 2020. ISSN 1558-0679. doi: 10.1109/TPWRS.2019.2961231.

Bat-Sheva Einbinder, Yaniv Romano, Matteo Sesia, and Yanfei Zhou. Training Uncertainty-Aware Classifiers with Conformalized Deep Learning. In *Advances in Neural Information Processing Systems*, New Orleans, LA, USA, November 2022. doi: 10.48550/arXiv.2205.05878. URL http://arxiv.org/abs/2205.05878.

Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *arXiv:1506.02142 [cs, stat]*, October 2016. URL http://arxiv.org/abs/1506.02142.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1321–1330, Sydney, Australia, August 2017. PMLR. URL https://proceedings.mlr.press/v70/guo17a.html.

Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate Uncertainties for Deep Learning Using Calibrated Regression. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 2796–2804. PMLR, July 2018. URL https://proceedings.mlr.press/v80/kuleshov18a.html.

Jesus Lago, Fjo De Ridder, and Bart De Schutter. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386–405, July 2018. ISSN 0306-2619. doi: 10.1016/j.apenergy.2018.02.069.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

Jeremiah Liu, Zi Lin, Shreyas Padhy, Dustin Tran, Tania Bedrax Weiss, and Balaji Lakshminarayanan. Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness. volume 33, pp. 7498–7512. Curran Associates, Inc., 2020.

Zachary Nado, Neil Band, Mark Collier, Josip Djolonga, Michael W. Dusenberry, Sebastian Farquhar, Qixuan Feng, Angelos Filos, Marton Havasi, Rodolphe Jenatton, Ghassen Jerfel, Jeremiah Liu, Zelda Mariet, Jeremy Nixon, Shreyas Padhy, Jie Ren, Tim G. J. Rudner, Faris Sbahi, Yeming Wen, Florian Wenzel, Kevin Murphy, D. Sculley, Balaji Lakshminarayanan, Jasper Snoek, Yarin Gal, and Dustin Tran. Uncertainty Baselines: Benchmarks for Uncertainty & Robustness in Deep Learning. January 2022. doi: 10.48550/arXiv.2106.04015. URL http://arxiv.org/abs/2106.04015.

Mariola Ndrio, Avinash N. Madavan, and Subhonmesh Bose. Pricing Conditional Value at Risk-Sensitive Economic Dispatch. In *2021 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 01–05, July 2021. doi: 10.1109/PESGM46819.2021.9637845.

Bala Kameshwar Poolla, Ashish R. Hota, Saverio Bolognani, Duncan S. Callaway, and Ashish Cherukuri. Wasserstein Distributionally Robust Look-Ahead Economic Dispatch. *IEEE Transactions on Power Systems*, 36(3):2010–2022, May 2021. ISSN 0885-8950, 1558-0679. doi: 10.1109/TPWRS.2020.3034488.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, November 2005. ISBN 978-0-262-25683-4. doi: 10.7551/mitpress/3206.001.0001. URL https://doi.org/10.7551/mitpress/3206.001.0001.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized Quantile Regression. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/5103c3584b063c431bd1268e9b5e76fb-Paper.pdf.

Glenn Shafer and Vladimir Vovk. A Tutorial on Conformal Prediction. *Journal of Machine Learning Research*, 9(12):371–421, 2008. URL http://jmlr.org/papers/v9/shafer08a.html.

David Stutz, Krishnamurthy, Dvijotham, Ali Taylan Cemgil, and Arnaud Doucet. Learning Optimal Conformal Classifiers, May 2022. URL http://arxiv.org/abs/2110.09192.

J. Warrington, P. J. Goulart, S. Mariéthoz, and M. Morari. Robust reserve operation in power systems using affine policies. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 1111–1117, December 2012. doi: 10.1109/CDC.2012.6425913.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P Xing. Deep Kernel Learning. In Arthur Gretton and Christian C. Robert (eds.), *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 370–378, Cadiz, Spain, May 2016. PMLR. URL http://proceedings.mlr.press/v51/wilson16.html.

Qipeng P. Zheng, Jianhui Wang, and Andrew L. Liu. Stochastic Optimization for Unit Commitment—A Review. *IEEE Transactions on Power Systems*, 30(4):1913–1924, July 2015. ISSN 1558-0679. doi: 10.1109/TPWRS.2014.2355204.

Weifeng Zhong, Kan Xie, Yi Liu, Shengli Xie, and Lihua Xie. Chance Constrained Scheduling and Pricing for Multi-Service Battery Energy Storage. *IEEE Transactions on Smart Grid*, 12(6): 5030–5042, November 2021. ISSN 1949-3061. doi: 10.1109/TSG.2021.3109140.

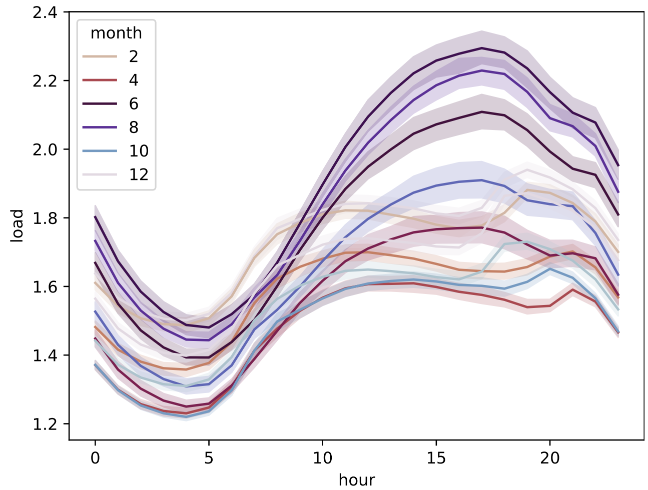

Figure 2: We plot means and bootstrapped 99% confidence intervals for the hourly electricity demand data from the PJM grid across different months from 2008-16.

A Appendix

A.1 Energy Storage Arbitrage

As a second example application of our end-to-end uncertainty-calibrated task-driven framework, we consider a realistic energy storage arbitrage task where a grid-scale battery storage operator wishes to operate their storage resources (*e.g.*, choosing charge/discharge decisions) in order to maximize profit of the electricity bought and sold to the grid. Let $z_t \in \mathbb{R}$ be the amount of energy that the battery charges (purchases) from the grid at each time step t, where $z_t < 0$ indicates the battery discharging (selling) back into the grid. Then, the robust formulation of the storage arbitrage problem is as follows:

$$\min_{z_1,\ldots,z_T \in \mathbb{R}} \max_{y \in \Omega(x)} \sum_{t=1}^{T} y_t z_t + \lambda \sum_{t=1}^{T-1} |z_{t+1} - z_t| \quad \text{s.t.} \quad Az \leq b$$

where $y \in \Omega(x) \subset \mathbb{R}^T$ is the uncertain vector of electricity prices for each of the next T time steps. The uncertainty set $\Omega(x)$ may be determined by inputs x such as the most recent electricity price, current battery state of change, electricity load, time of day, and weather conditions. The first sum in the objective is the negative profit of arbitrage, whereas the second sum penalizes charging or discharging too quickly. The constraint $Az \leq b$ includes all relevant capacity and state of charge constraints on battery operation. As with the grid scheduling application, this semi-infinite program can be tractably reformulated as a convex optimization problem for a variety of choices of uncertainty set $\Omega(x)$ (Ben-Tal et al., 2009).

A.2 Data

We will test our method on real historical electricity demand and price data from the PJM electricity grid, including 7 years of data to train the model, and 1.75 years for testing. We will use the same data[1] as Donti et al. (2017) to ensure a fair comparison to previous work. See Figure 2 for a plot of the historical electricity demand data.

[1] `https://github.com/locuslab/e2e-model-learning/`