



xPRO



APPLICATIONS FOR MACHINE LEARNING IN ENGINEERING AND THE PHYSICAL SCIENCES

HOW TO EVALUATE WHERE MACHINE LEARNING
WILL CONTRIBUTE TO OPTIMAL RESULTS



M

achine learning offers important new capabilities for solving today's complex problems, but it's not a panacea. To get beyond the hype, engineers and scientists must discern how and where machine learning tools are the best option – and where they are not.

Once the province of experimentation, machine learning is hitting its stride now that computational modeling capacity can handle the massive amounts of data and computing power available. We see the application of machine learning across varied industries including aerospace, construction, defense, transportation, energy, semiconductors, pharmaceuticals, materials, climate modeling, seismology, and more.

DEMYSTIFYING MACHINE LEARNING FOR ENGINEERING AND PHYSICAL SCIENCES

Machine learning provides a powerful set of tools that enable engineers and scientists to use data to solve meaningful and complex problems, make predictions, and optimize the systems and products they discover and design. The latest developments in data-driven modeling, applied in conjunction with first-principles modeling techniques, can deliver more rapid results and improve the reliability of predictions.

And yet one of the biggest hurdles with machine learning is seeing opportunities to solve problems with the technology in the first place.

In this paper, we describe different applications of machine learning in engineering and scientific problem-solving. We aim to demystify how machine learning can help solve some difficult problems and describe the risks and costs of using it in your work.

MACHINE LEARNING BUILDS ON THE SAME MATH AS PHYSICS-BASED MODELING

The vast volumes of data and powerful computational processes now available open new avenues for exploration and analysis. This has led some organizations to take the leap too quickly, indiscriminately using machine learning in applications where it may not be efficient or even appropriate.

Considering engineering and the physical sciences, the competitive advantage of machine learning will come through weighing the different tools available and selecting the best ones for each particular job to be done. You need to identify precisely where and how machine learning will create the biggest “lift.” For example, we are familiar with the gaps in traditional model-based computation and simulation. Machine learning can be applied in combination with more traditional methods to close some of these gaps.

Machine learning gives us the ability to fuse theory-based models with data. According to Youssef Marzouk, professor of aeronautics and astronautics at MIT and co-director of the MIT Center for Computational Science and Engineering (CCSE) in the new MIT Schwarzman College of Computing, “There are many situations where theory-based models are not enough. They might be missing some crucial but complicated interactions. They might be too expensive to simulate. They might not capture the phenomena that we're really interested in. We have powerful ways of building models from data and making predictions that we simply could not make before.”



WHAT'S POSSIBLE?

Machine learning is enabling exciting progress in the engineering and scientific fields. For example, machine learning is helping:

Improve the aerodynamics of aircraft.

The aerospace industry has used computational fluid dynamics (CFD) for decades. While the models are highly reliable and sophisticated, the modeling frameworks have acknowledged gaps or limitations. They might dismiss some key physics or the cost of sufficient computational resolution may be prohibitive. Machine learning can take libraries of data from wind tunnel testing and other experiments and fuse them into fluid simulations to come up with better predictive models for aerospace design.

Predict battery life.

Predicting the useful life of a battery is a notoriously difficult problem because of the variability of battery lifetime even off the same manufacturing line. After a machine learning model was trained with a few hundred million data points of batteries charging and discharging, the algorithm was able to predict the useful life of lithium-ion batteries before their capacities started to wane. This research has reduced the extent of battery testing—one of the most time-consuming and expensive steps of battery design—by an order of magnitude.

Develop new catalysts for chemical reactions.

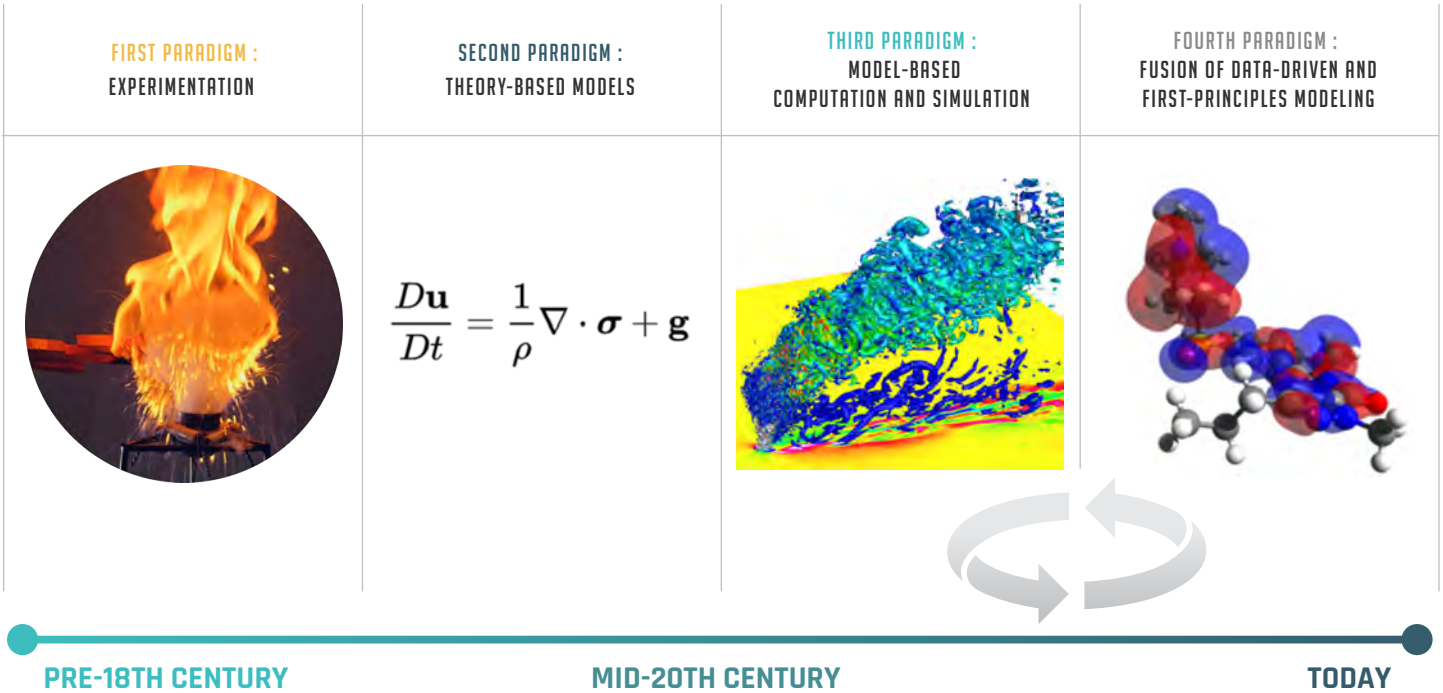
In materials science, if you are trying to design a new catalyst, you may not have data relevant to the new chemical process you care about. Computational simulation can work in concert with machine learning tools by creating a database to make better predictions around how molecules will respond to catalysts even before physical experimentation.

MACHINE LEARNING THROUGH A PHYSICS-INFORMED LENS

Machine learning is bridging the computational paradigms and building on hundreds of years of modeling history and progress. This heritage means engineers less familiar with modern data science or computational methods can still gain advantage in applying machine learning to their problem-solving when appropriate. The machine learning that engineers and scientists will find most useful layers on top of first-principles modeling, simulation, and prediction tools for greater power and results.



COMPUTATIONAL SCIENCE + ENGINEERING PARADIGMS



Before the advent of practical computing from the middle of the 20th century, engineers and scientists started to apply first-principles physical laws built from observation (called here the first paradigm) and



theory-based models (the second paradigm) to describe the physical world with a computer. Eventually, as computers became more powerful and accessible in the third paradigm, computational modeling and simulation were applied to theory-based models providing new levels of prediction. This approach has been an essential tool that underpins everything from the design of new airplanes to the optimization of manufacturing processes.

We now talk about a fourth paradigm, which yields extraordinary capabilities at a mere fraction of the cost. This computational transformation fuses theory-based models with data using high-performance computing, huge datasets, and novel, powerful algorithms including machine learning. We've supercharged the numerical methods and optimizations in the third paradigm to form the backbone of modern engineering and scientific problem-solving.

WHERE WILL MACHINE LEARNING HAVE THE BIGGEST IMPACT ON ENGINEERING AND SCIENCE?

Big data's unwieldiness, inhomogeneity, and nonstop growth make it increasingly difficult to manage. Machine learning can be harnessed in ways to respond to burgeoning data by figuring out the rules to follow and identifying what's most important and what's just fluff.

MACHINE LEARNING SOLVES COMMON CHALLENGES

Engineers and scientists aim to use machine learning to:

1. **Accelerate processing and increase efficiency.** Once you choose a model, machine learning can use it to learn the patterns in your data and then further tune and refine the model to more quickly predict outcomes, especially while inputting new conditions and observations.
2. **Quantify and manage risk.** Machine learning can be used to model the probability of different outcomes in a process that cannot easily be predicted due to randomness or noise. This is especially invaluable for situations where reliability and safety are paramount.
3. **Compensate for missing data.** Accurate learning, inference, and prediction are severely limited in the presence of missing data. Models trained by machine learning improve with more relevant data. Yet as long as it's used correctly, machine learning can help synthesize missing data that round out incomplete datasets.
4. **Make more accurate predictions or conclusions from your data.** By tuning the settings of how your machine learning model's parameters will be updated and learned during training you can streamline your data-to-prediction pipeline. Building better

models of your data will also improve the accuracy of subsequent predictions.

5. **Solve complex classification and prediction problems.** Predicting how an organism's genome will be expressed or what the climate will be like in fifty years are examples of highly complex problems. Many modern machine learning problems take thousands or even millions (or far more) of data samples across many dimensions to build expressive and powerful predictors, often pushing far beyond traditional statistical methods.
6. **Create new designs.** There is often a disconnect between what designers envision and how products are made. It's costly and time-consuming to simulate every variation of a long list of design variables. Machine learning can take the input of key variables to generate good options and help designers identify which best fits their requirements.
7. **Increase yields.** Manufacturers aim to overcome inconsistency in equipment performance and predict maintenance by applying machine learning to flag defects and quality issues before products ship to customers, improve efficiency on the production line, and increase yields by optimizing use of the manufacturing resources.

“Rather than rely on trial and error, machine learning is a powerful tool to accelerate the discovery process and give us shortcuts to solving design problems and finding design rules.”

- MIT xPRO course material from Heather J. Kulik, Associate Professor of Chemical Engineering, MIT

APPLICATIONS OF MACHINE LEARNING IN INDUSTRY

How are engineers and scientists gaining advantage with these tools? Here are just a few specific applications for machine learning we see across the engineering and the physical sciences.

AEROSPACE

Understand the behavior and aerodynamic properties of airfoils in order to make designs with reduced noise, which is critically important for both efficiency and reduced environmental impact.

FLUID MECHANICS

Capture essential flow mechanisms at a fraction of the cost through new avenues for dimensionality reduction and reduced-order modeling by providing a concise framework that complements and extends existing CFD methodologies.

OIL AND GAS EXPLORATION

Predict how formations will react to certain drilling techniques to pinpoint the best route through a rock formation and dig virtually even before drilling equipment arrives onsite.

MATERIALS SCIENCE

Understand the complex relationship between a powder's metallurgical parameters, the printing process, and the microstructure and mechanical properties of additive manufacturing parts to make impact-resistant materials.

MECHANICAL ENGINEERING

Lengthen the remaining useful life of equipment through predictive maintenance of machinery, maximizing asset lifetime, operational efficiency, or uptime.

GEOPHYSICS + SEISMOLOGY

Reconstruct seismic data from under-sampled or missing traces. Enable intelligent interpolation between data sets with similar geomorphological structures, which can significantly reduce costs in engineering applications.

CLIMATE SCIENCE

Determine the changing behavior of extreme weather such as the frequency and ferocity of tropical storms, the intensity and geometry of atmospheric currents, and their relationship with fluctuating ocean temperatures.

BIOMEDICINE

Use gene expression data to classify patients in different clinical groups and to identify new disease groups, while genetic code allows prediction of the protein secondary structure.

CIVIL ENGINEERING

Address complex problems such as river flow forecasting, modeling evaporation, modeling compressive strength of concrete, and groundwater level forecasting.

THE PAYOFF OF MACHINE LEARNING

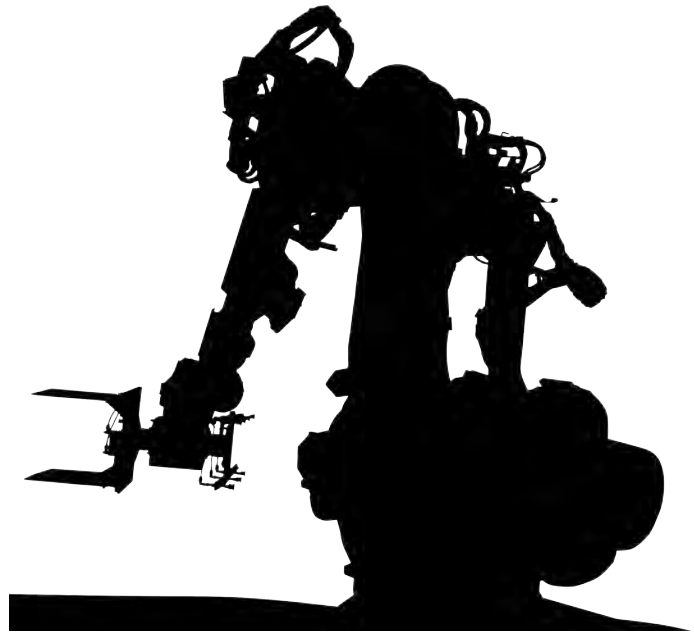
Applying machine learning most typically yields financial impact by reducing risk, increasing speed to delivery, and/or decreasing costs.

For example, a top European manufacturer of chemicals set out to improve production process yield. They used machine learning techniques to examine sensor data for features including carbon dioxide flow, coolant pressures, quantities, and temperatures, and compared these features to determine which was most important according to their model. Carbon dioxide flow rates proved to be the most impactful factor. With a moderate change in parameters, the waste of raw materials was lowered by 20%, energy costs decreased by 15%, and process yield was considerably improved.

Beyond cost savings and increased yields, processing and analyzing data amassed through machine learning often reveal previously unseen behaviors which, in turn, may lead to new opportunities for improvement.

Machine learning improves product quality up to **35%** in discrete manufacturing industries

- Deloitte, 2017



SOME CONSIDERATIONS WHEN APPLYING MACHINE LEARNING

UNDERSTANDING THE LIMITATIONS AND TRADEOFFS

One of the key considerations when choosing a machine learning model is to make a series of choices between trade-offs. The right answers for these problems depend on your priorities as well as the nature of the problem or process being studied. Here are a set of questions that will help you explore and weigh various approaches.

QUESTIONS ABOUT THE DATA

- How much data do we have?
- How diverse is that data?
- Do you have distinct features in the data set?
- How repeatable, reliable, or deterministic is it?
- How much does it cost to obtain that data in terms of time or human effort?

QUESTIONS WHEN CHOOSING THE MODEL



- How interpretable do you want the model to be?
- Do you have enough data to train the model appropriately?
- How much expert knowledge do you have in model training?
- Do you need a measure of model uncertainty?
- Is the model overdetermined or underdetermined?

How do you weigh the tradeoffs and identify the optimal approaches? Learning the capabilities and limits of various models and approaches is a good start. It will deepen your understanding of how regularization methods can be used to help your models provide a better predictive fit.

IT'S USUALLY NOT A SIMPLE YES OR NO

Most often, engineers and scientists will take a hybrid approach that includes physics-based modeling and machine learning.

Physics-based modeling helps whittle down the computational complexity of training your model. Since training tends to be one of the more expensive parts of applying machine learning, an understanding of physics-based modeling can save time and money.

For example, finite element method (FEM) is commonly used to solve engineering and mathematical physics problems. Physics-based machine learning can increase the accuracy of predictions and reduce the cost of training by combining training data from conventional FEM simulations with data from experiments and other variables.

DEFINING THE FEATURE SET IS A CRUCIAL PART OF THE PROCESS

Feature engineering is a critical consideration within the machine learning process. A feature is a transfor-

mation of the raw data in a way that is useful for the modeling task. In machine learning, features might be very abstract, made up of a set of numbers that combine many different quantities. Setting up the problem requires selecting the most important features and knowing their contribution to the end result.

“You don’t have to become a machine learning expert to apply these new tools effectively. Engineers and scientists who more fully understand where machine learning can help – and where it can’t – can achieve real gains from these tools.”

- Youssef Marzouk, Professor of Aeronautics and Astronautics, MIT, and Co-Director, MIT Center for Computational Science and Engineering

For engineers and scientists, context is crucial for AI and machine learning to achieve the desired results. Context determines the techniques that are better or worse and the level of confidence that’s acceptable or unacceptable in a given situation. This is where the domain expertise they bring can make machine learning most powerful.

MACHINE LEARNING CAN WORK FOR INVERSE PROBLEMS, TOO

Machine learning can work for inverse problems as well, when you are estimating the parameters of your problem. You can use machine learning to derive the parameters from the data you have and better understand your assumptions from the data you have.

“Machine learning can also help you work back from the results you have, so you can identify the features that are most important. Using machine learning to solve these inverse problems is a powerful tool, and not one that many are fully accessing,” according to Marzouk.

DEBUNKING PRECONCEPTIONS ABOUT MACHINE LEARNING

PRECONCEPTIONS

- You can only use machine learning with huge data sets.
- All data is useful and more is better.
- Machine learning is only for data scientists.
- Machine learning is not transparent enough. We don't want a “black box.”
- You have to find that one perfect model.
- Finding machine learning talent is tough.

REALITY

- Machine learning can often be applied to help round out missing data.
- Not all data are relevant. It's important to consider the risk of overfitting the model.
- A lot of machine learning is based on principles and simple mathematical techniques that many engineers and scientists already know.
- There's a lot of research in improving transparency. Just because it's not transparent doesn't mean it's not valuable.
- It's usually a combination of approaches that works best. Knowing how to assess the accuracy of models will always be helpful.
- Top engineers and scientists are already looking to learn machine learning techniques in the context of their disciplines.

REFINING YOUR APPROACH TO MACHINE LEARNING

Machine learning is the area of computational science that offers new ways to tackle real-world engineering and scientific problems.

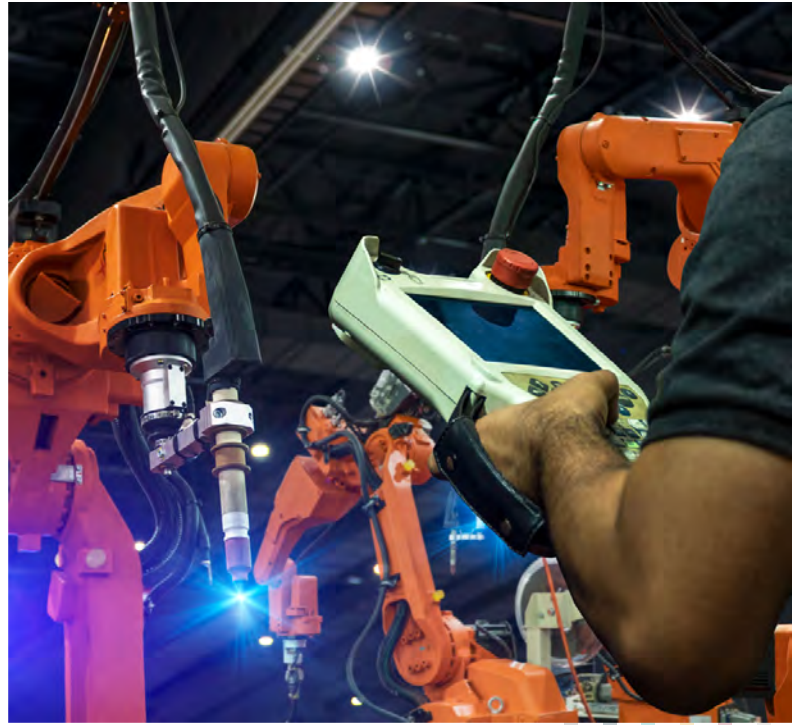
When engineers and scientists develop a deeper understanding of machine learning methods, they

can apply machine learning in a more judicious way. They can more quickly assess where machine learning can help accelerate their processes and where other modeling methods remain a better option.

According to Marzouk, “Adoption is growing fast but what's happening is still a lot of ad hoc experimentation. Many industries are somewhere between the initial discovery phase and the phase where they

begin to encounter limitations. You don't have to become a machine learning expert to apply these new tools effectively. Engineers who more fully understand where machine learning can help - and where it can't - can achieve real gains from these tools, whatever their industry."

We encourage leaders in engineering and physical sciences to help their teams develop a deeper understanding of the capabilities of machine learning. By applying these new tools thoughtfully to create specific models in their respective disciplines, they can help make better predictions that mitigate risk and catapult their work ahead.



SOURCES

MIT xPRO, "Machine Learning, Modeling, and Simulation: Engineering Problem-Solving in the Age of AI," faculty research and course presentations, 2020.

<https://xpro.mit.edu/programs/program-v1:xPRO+MLx/>

Deloitte Digital, "Asset Monitoring & Predictive Maintenance," May 2017.

<https://www2.deloitte.com/content/dam/Deloitte/us/Documents/about-deloitte/us-a-turnkey-iot-solution-for-manufacturing.pdf>

Digital McKinsey, "Smartening Up with Artificial Intelligence," April 2017.

https://deepsense.ai/wp-content/uploads/2018/04/170419_mckinsey_ki_final_m.pdf

Nature, "Recent Advances and Applications of Machine Learning in Solid-State Materials Science," August 2019.

<https://www.nature.com/articles/s41524-019-0221-0>

SAS, "The Machine Learning Primer," 2017.

https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/machine-learning-primer-108796.pdf

SPD Group, "AI and Machine Learning in Manufacturing: The Complete Guide."

https://spd.group/machine-learning/ai-and-ml-in-manufacturing-industry/#Designing_a_jet_engine,
January 2020.

towards data science, "How Do You Teach Physics to Machine Learning Models?" August 2018.

<https://towardsdatascience.com/how-do-you-combine-machine-learning-and-physics-based-modeling-3a3545d58ab9>

WIRED, "Machine Learning Takes the Guesswork Out of Design Optimization," August, 2019.

<https://www.wired.com/brandlab/2019/08/machine-learning-takes-guesswork-design-optimization/>

Nature Energy, "Data-driven Prediction of Battery Cycle Life Before Capacity Degradation," March 2019.

<https://www.nature.com/articles/s41560-019-0356-8>