



Unlock Data to Optimize Industrial Processes

John Hedengren
Brigham Young University
EGI Technical Conference



BYU PRISM
MODEL OPTIMIZE CONTROL

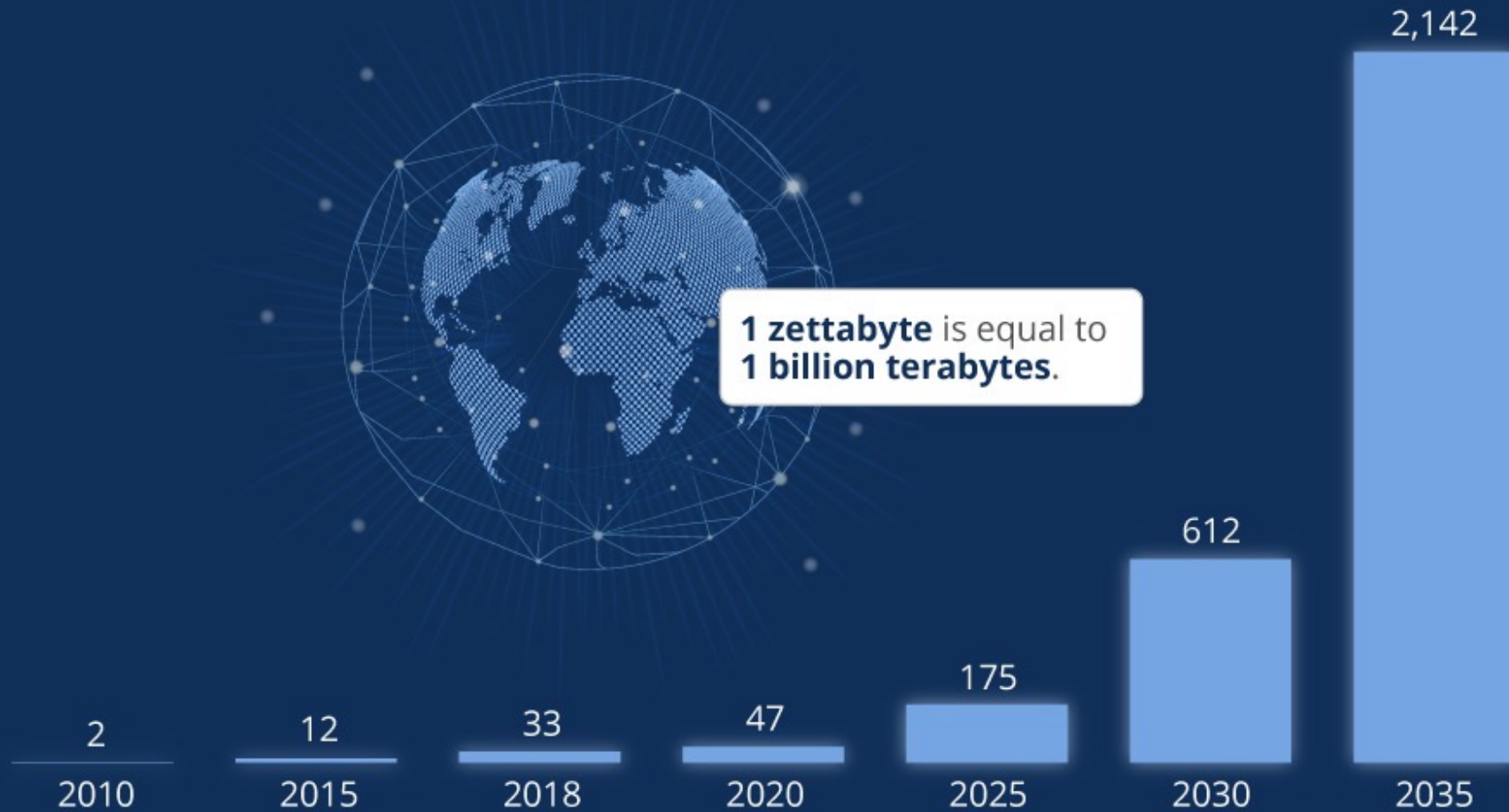


Outline

- Trends in Data and Computing
- Data-Driven Engineering Tutorials
- Research Overview
- Conclusions

Current Trends in Data

Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)



25 GB/hour



150,000 points/sec



51,200 GB/hr

Source: Simafore, Fortune, RTInsights, Cisco

Current Trends in Data



25 GB/hour



150,000 points/sec

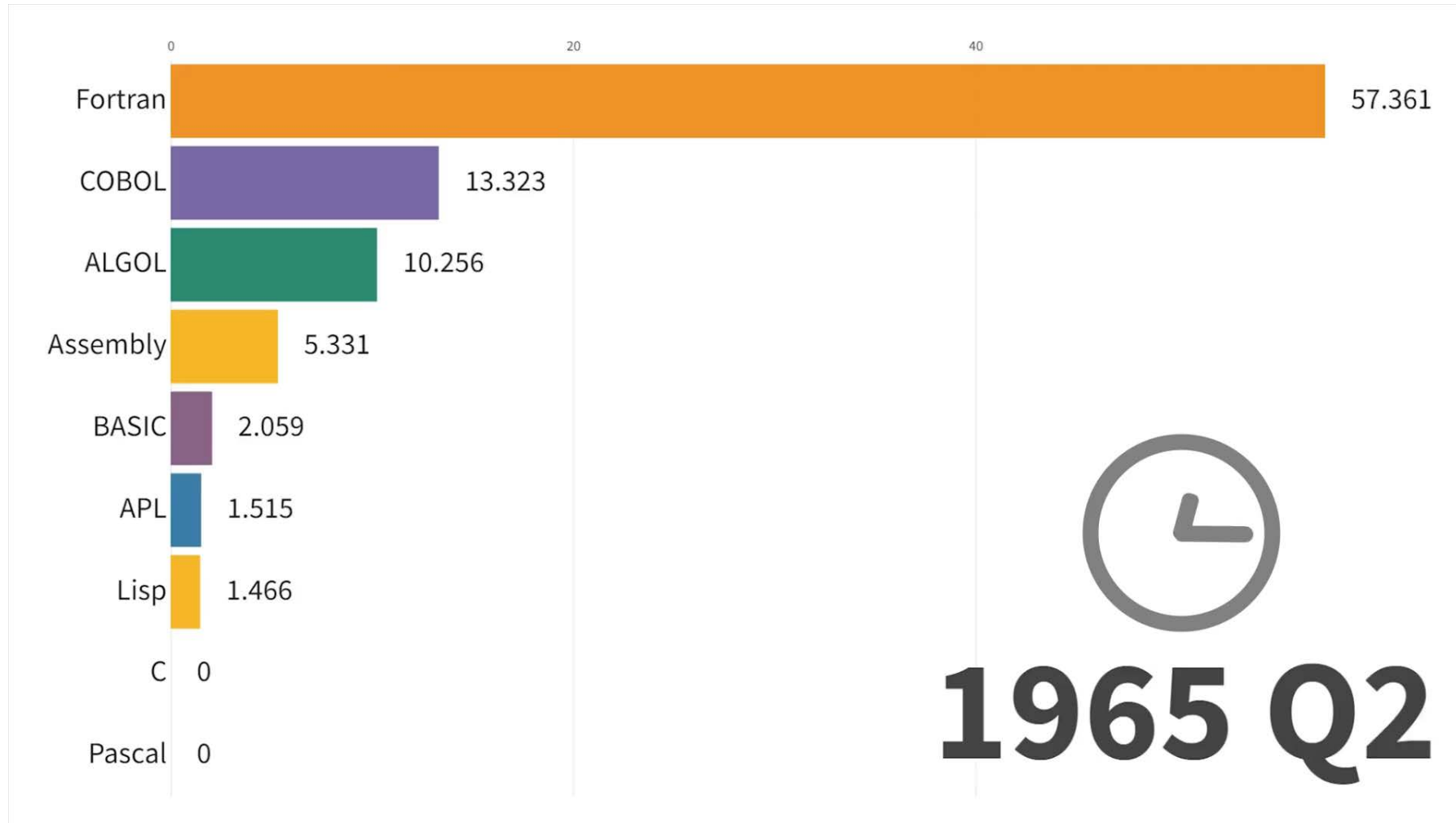


51,200 GB/hr

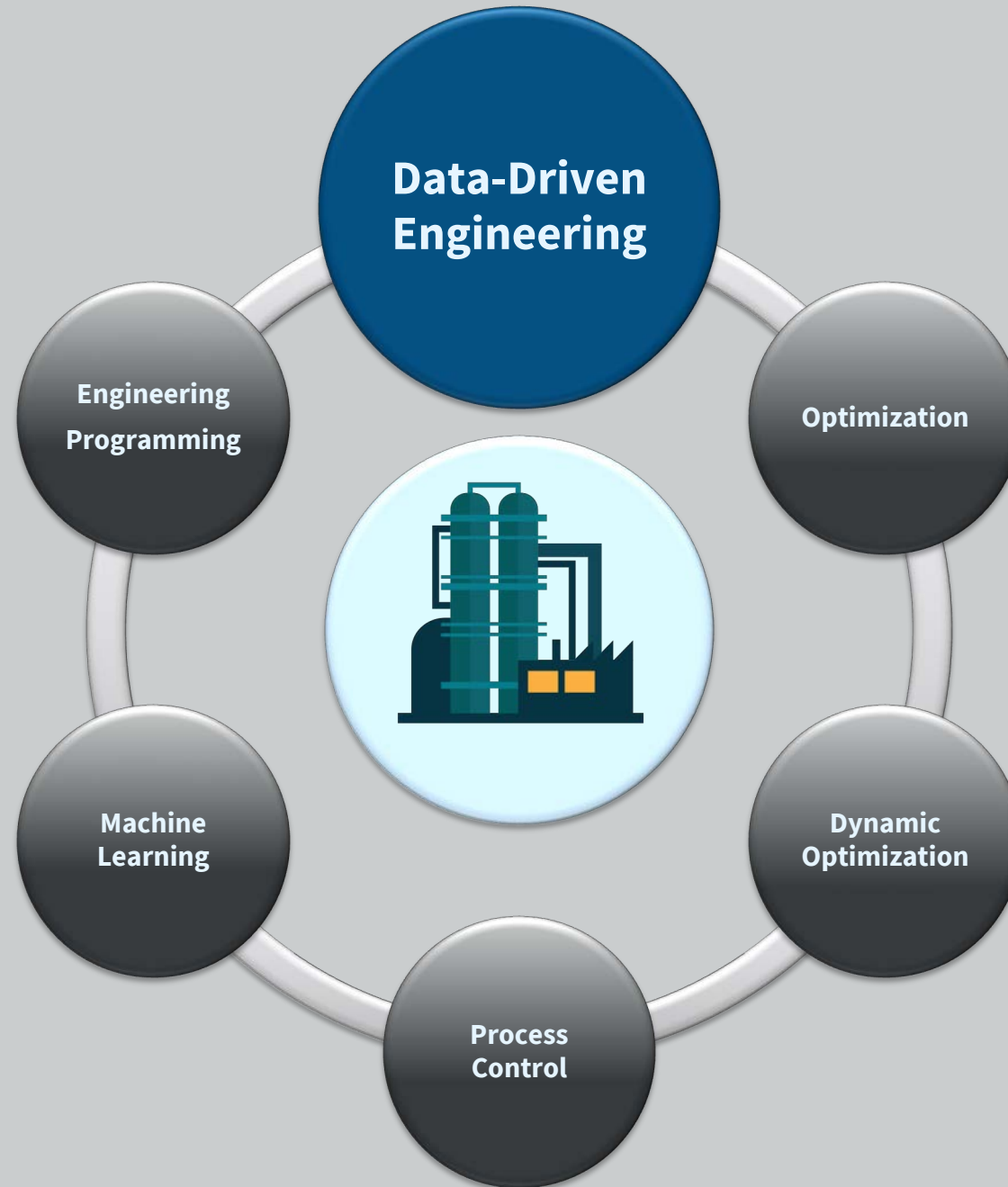
Source: Simafore, Fortune, RTInsights, Cisco

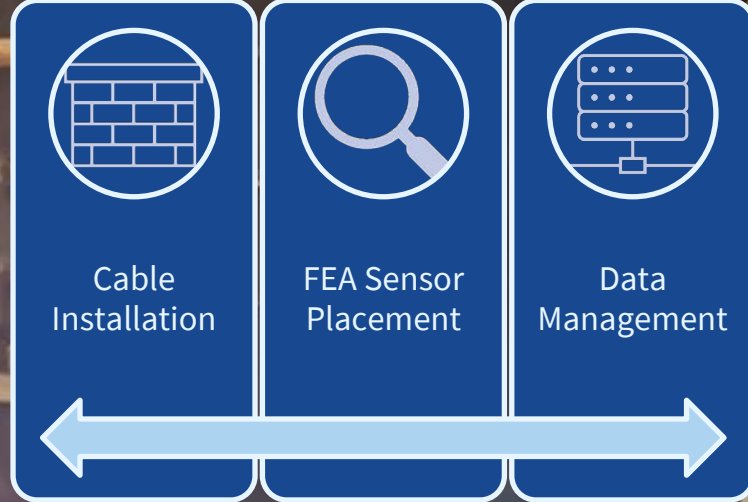
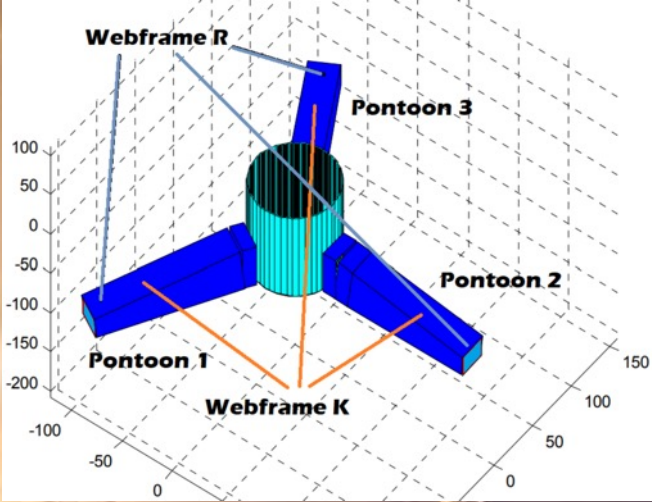
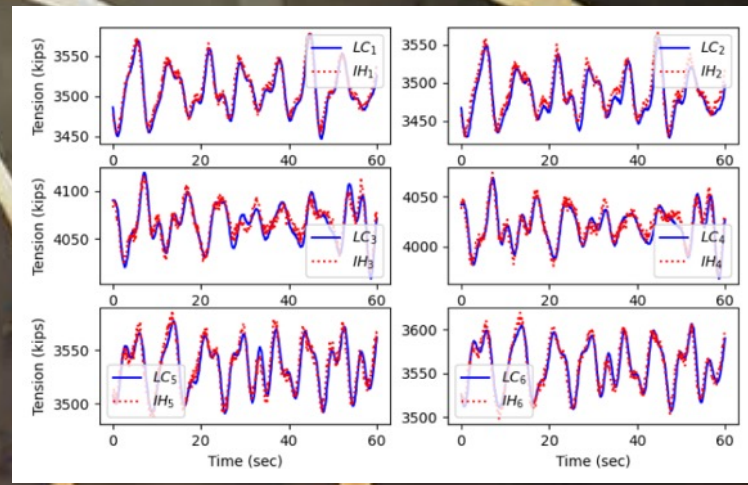
Current Trends in Computing

- +22% projected growth in programming jobs over next decade
 - Development, QA, Analysis, Testing



Source: Data is Beautiful Most Popular Programming Languages 1965 - 2022





AI to Enhance HI (Human Intelligence)

FIGURE 1. Achievement distribution for students under conventional, mastery learning, and tutorial instruction.

CONVENTIONAL 1-30*

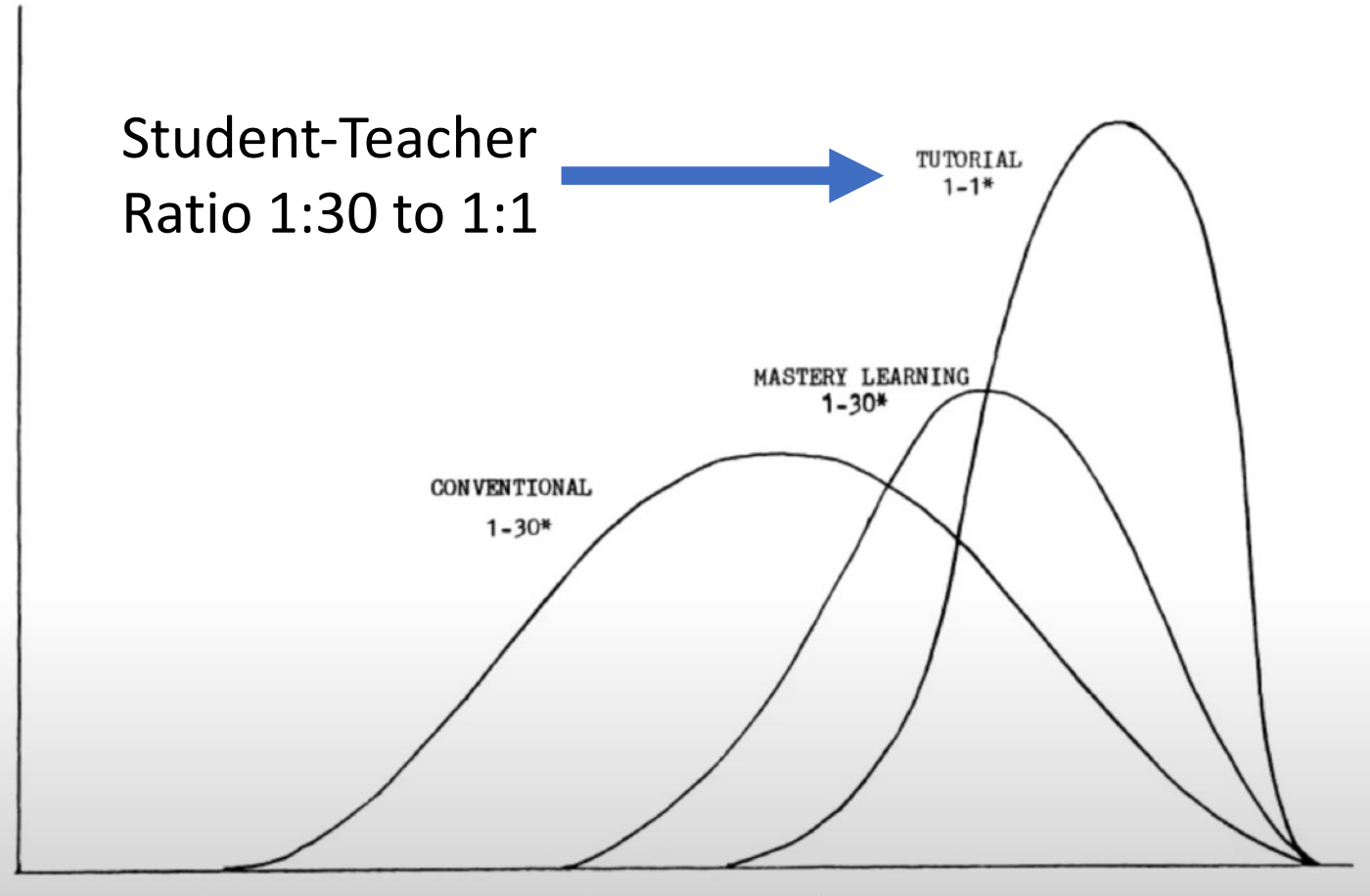
MASTERY LEARNING 1-30*

TUTORIAL 1-1*

*Teacher-student ratio

Summative Achievement Scores

June/July 1984



Salman Khan, TED Talk 2023

Summative Achievement Scores



 GPT-3.5

 GPT-4 

ChatGPT

Brainstorm incentives

for a customer loyalty program in a small book...

Plan a trip

to explore the Madagascar wildlife on a budget

Make a content strategy

for a newsletter featuring free local weekend e...

Come up with concepts

for a retro-style arcade game

Send a message

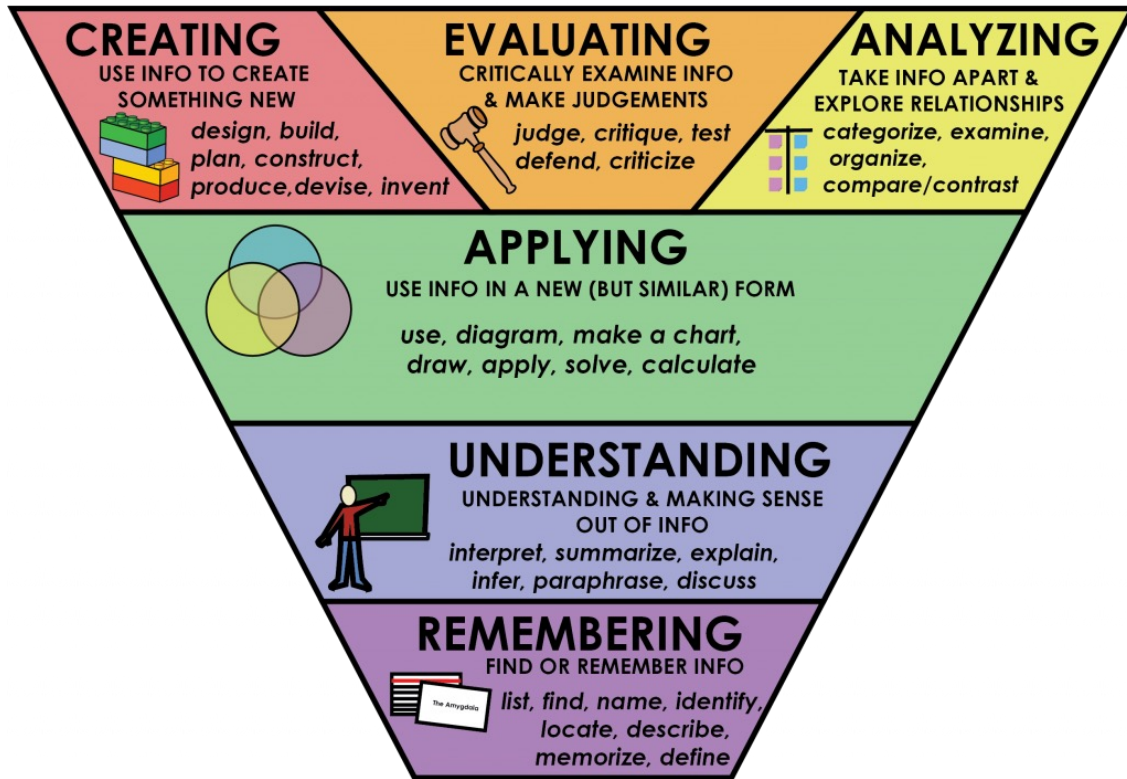


NEW

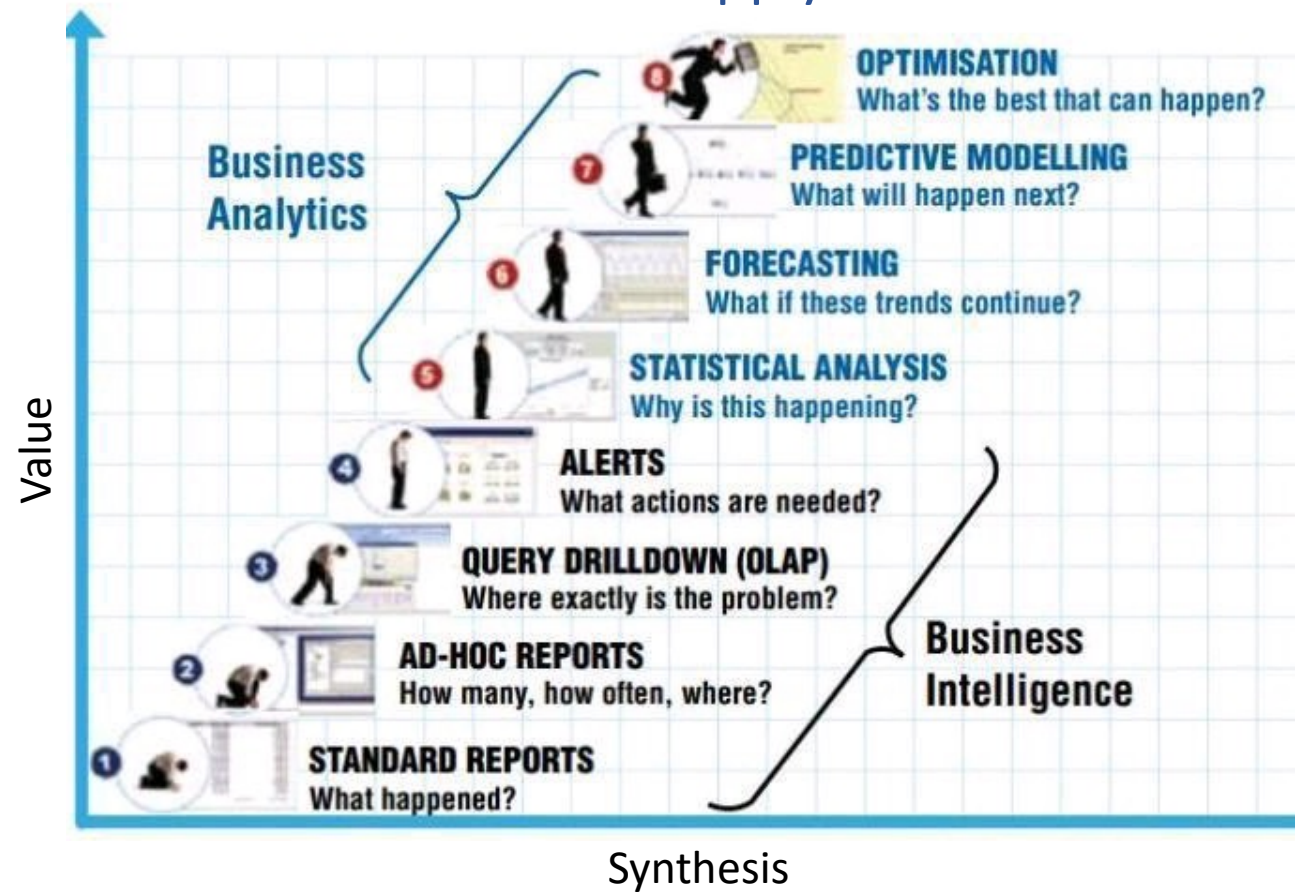


Generative AI for Prediction and Optimization

Learn



Apply





Data-Driven Engineering
apmonitor.com/dde

Data-Driven Engineering



Install 

1

Basics

2

Tuple

3

List

4

Set

5

Dict

6

Numpy

7

Pandas



Data-Driven Engineering

1 Python Basics

- 🔁 LOOPS
- 📄 PRINT

Data-Driven Engineering

5 Python Dict 🐍

- 📁 CREATE
- 📄 PRINT

Data-Driven Engineering

2 Python Tuple

- 📄 PRINT
- 📄 ITERATOR

Data-Driven Engineering

6 Python NumPy

- 📁 INSTALL + IMPORT
- 1/2 3/4 ID, 2D, 3D ARRAYS

Data-Driven Engineering

3 Python List 🐍

- 💡 CREATE 🗑️ REMOVE
- 📄 ITEMS 🛠️ EXTEND
- 🔍 FIND ✂️ POP

Data-Driven Engineering

7 Python Pandas

- 📁 INSTALL 📄 SORT
- 👤 IMPORT 🖼️ DATAFRAME
- 1/2 3/4 SERIES ↺️ TRANSPOSE
- 💬 INDEX 🔍 INSPECT
- 🔍 RETRIEVE ✂️ CLEANSE
- 🔪 SLICE 📊 STATISTICS
- 📏 LENGTH 📈 PLOTS
- 📊 pandas 🐍

Data-Driven Engineering

4 Python Set 🐍

- 💡 CREATE
- 📄 PRINT
- 📄 UNPACK
- 👤 COPY
- 🗑️ REMOVE
- 👤 INSPECT
- 🔑 METHODS

Python Introduction

6



NumPy

<https://apmonitor.com/dde>

Data-Driven Engineering



6 Python NumPy



INSTALL + IMPORT



1D, 2D, 3D ARRAYS



EXPORT + IMPORT



UNARY OPERATIONS



BINARY OPERATIONS



SCAN ME

- 1 Basics
- 2 Tuple
- 3 List
- 4 Set
- 5 Dictionary
- 6 NumPy
- 7 Pandas



ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021



Prompt Learning








- Help me find the error in my code without showing the answer.
- Explain each line of this Python code to a Matlab user.
- Generate a similar example.
- How can I make this more Pythonic?
- Test my knowledge of Numpy linspace with a quiz.
- Summarize what we've discussed so far.
- Translate this Python code to Matlab.
- I'm interested in _____. Why is this important to know?
- Generate a lesson plan on Numpy.



Data-Driven Engineering



Data Import

- 1 TEXT 
- 2 AUDIO 
- 3 VIDEO 
- 4 DATABASE 
- 5 SENSORS 
- 6 CLOUD 
- 7 WEB SCRAPING 

Data-Driven Engineering

Text Data

CSV

Data-Driven Engineering

MicroPython

Thonny

ESP32

Data-Driven Engineering

Audio Data

Data-Driven Engineering

Cloud

aws

Data-Driven Engineering

Video Data

Data-Driven Engineering

Invoice	Order
Invoice_id	Order_id
Customer_id	Total
Order_id	Product_id
Date_time	Customer_id
Status	Date_time
Total	Remark

Database

SQLite

CONNECT, UPDATE, SELECT, FETCH

Data-Driven Engineering

Web Scrape

BeautifulSoup⁴

pandas

TEXT, TABLES, IMAGES

Data Import and Access



Text Data Analysis

<https://apmonitor.com/dde>



Data-Driven Engineering



Text Data

CSV

IMPORT AND ANALYZE




SCAN ME

- 1 Text
- 2 Audio
- 3 Video
- 4 Database
- 5 Sensors
- 6 Cloud
- 7 Web Scraping

Data-Driven Engineering

DATA TRANSFER

- 1 MODBUS
- 2 MQTT 
- 3 OPC UA
- 4 WebSocket

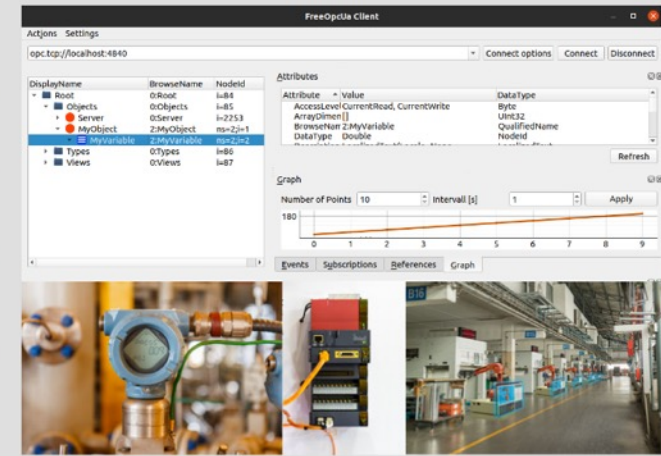
Data-Driven Engineering



Object type	Access	Size	Address Space
Coil	Read-write	1 bit	00001 – 09999
Discrete input	Read-only	1 bit	10001 – 19999
Input register	Read-only	16 bits	30001 – 39999
Holding register	Read-write	16 bits	40001 – 49999

Data-Driven Engineering

Server & Client



Data-Driven Engineering



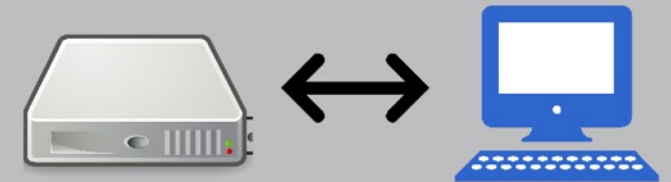
Broker & Clients

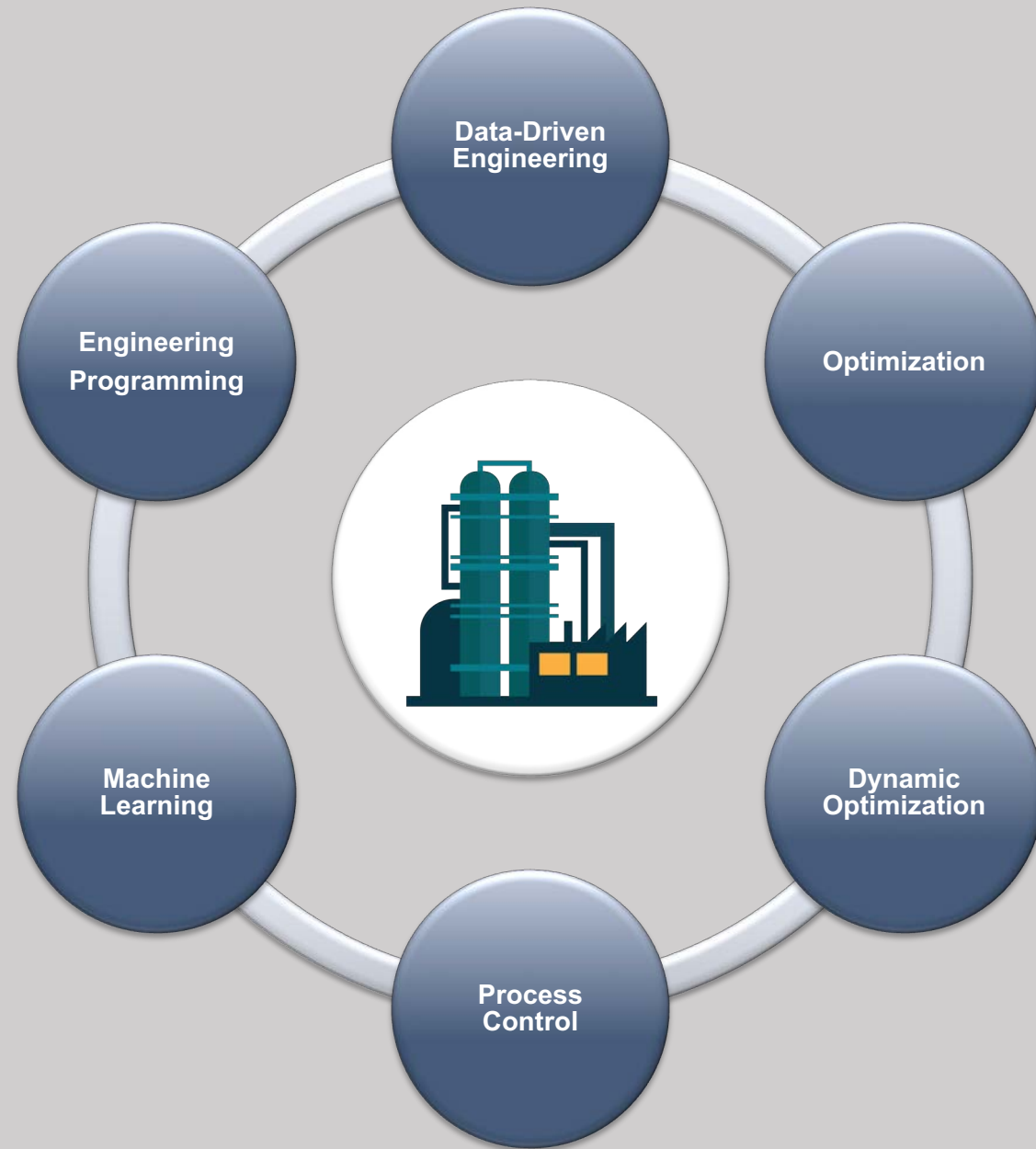


Data-Driven Engineering



Server & Client





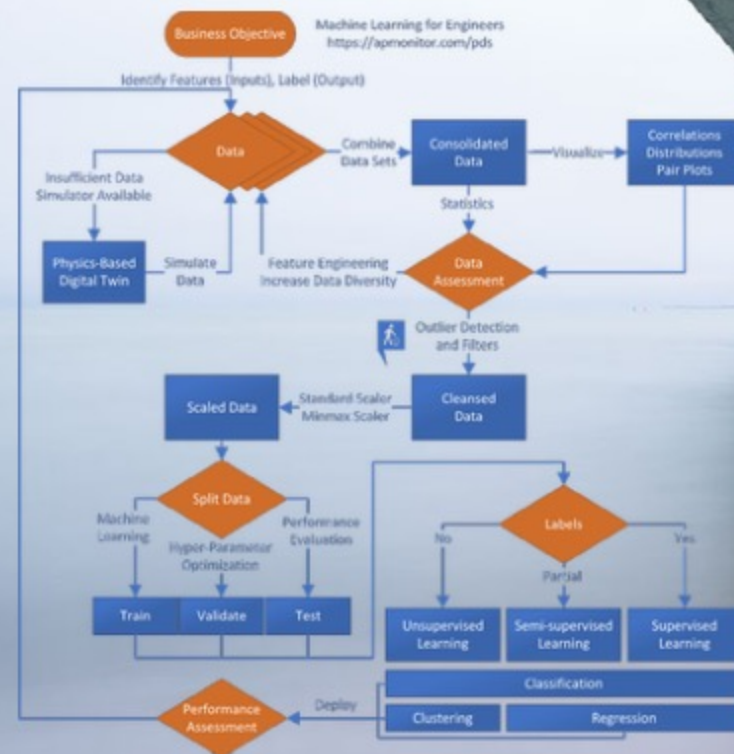


Foundations Applications Projects

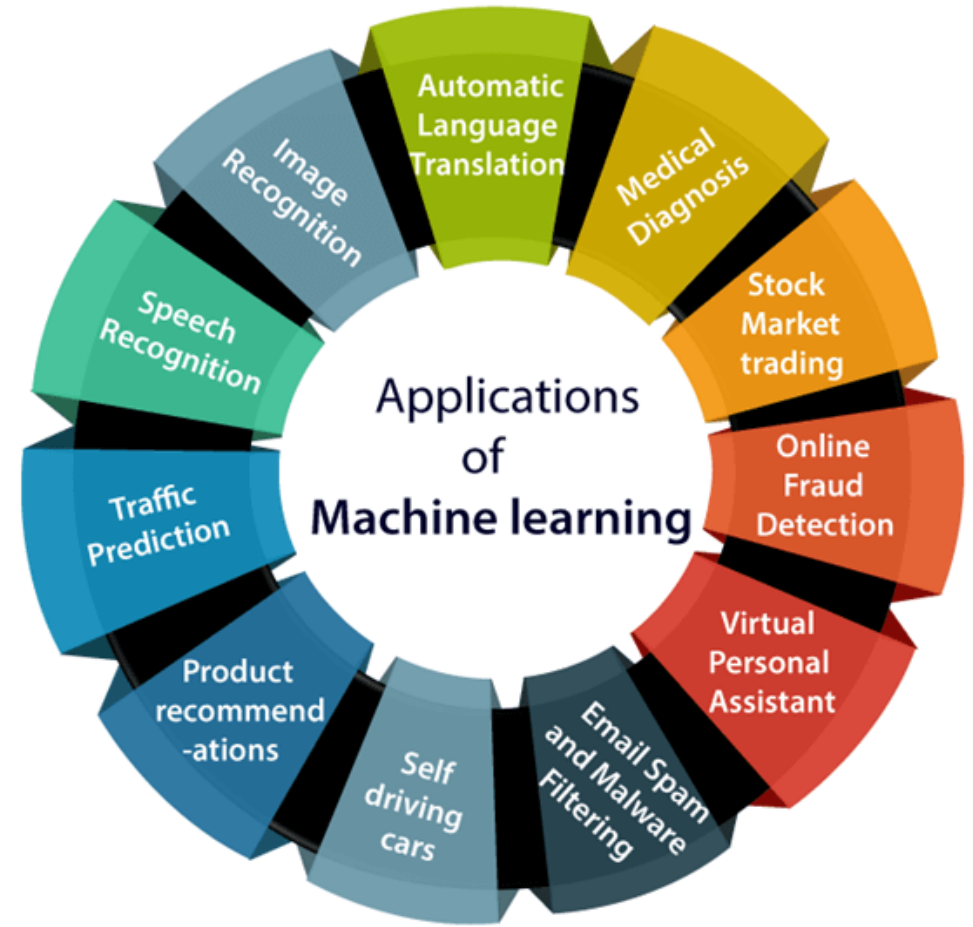
Machine Learning for Engineers



<https://apmonitor.com/pds>

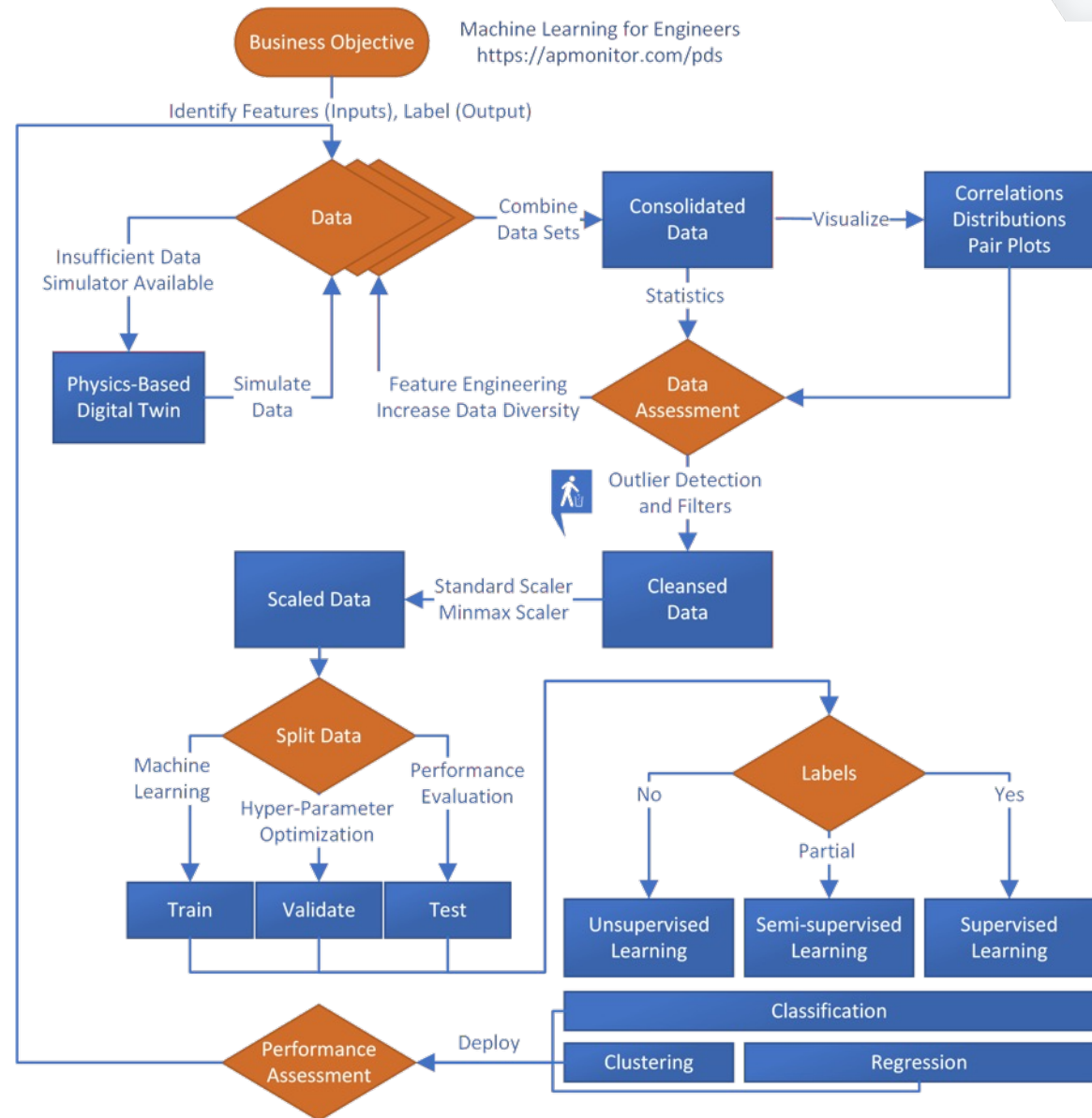


Machine Learning Applications

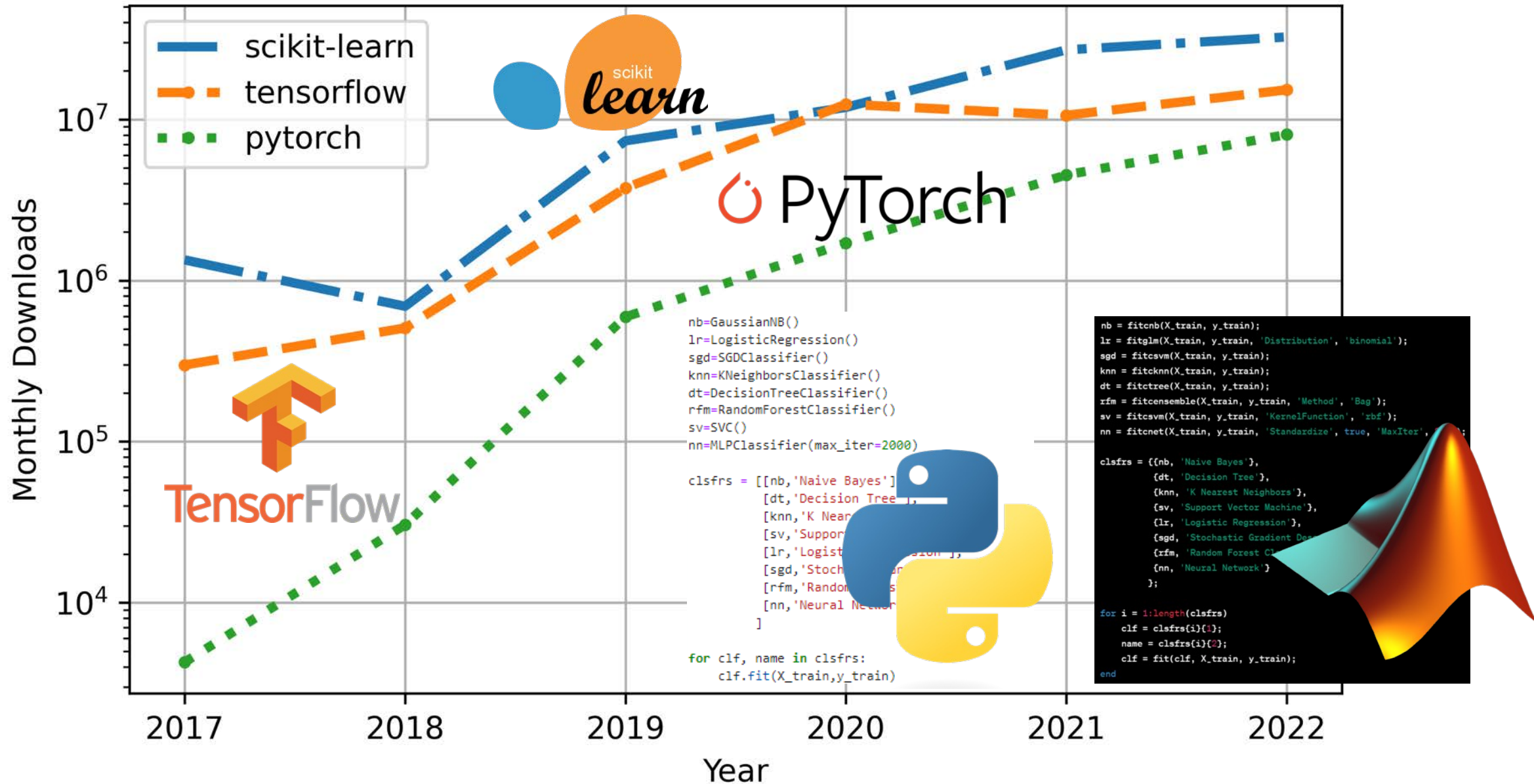


<https://www.javatpoint.com/applications-of-machine-learning>

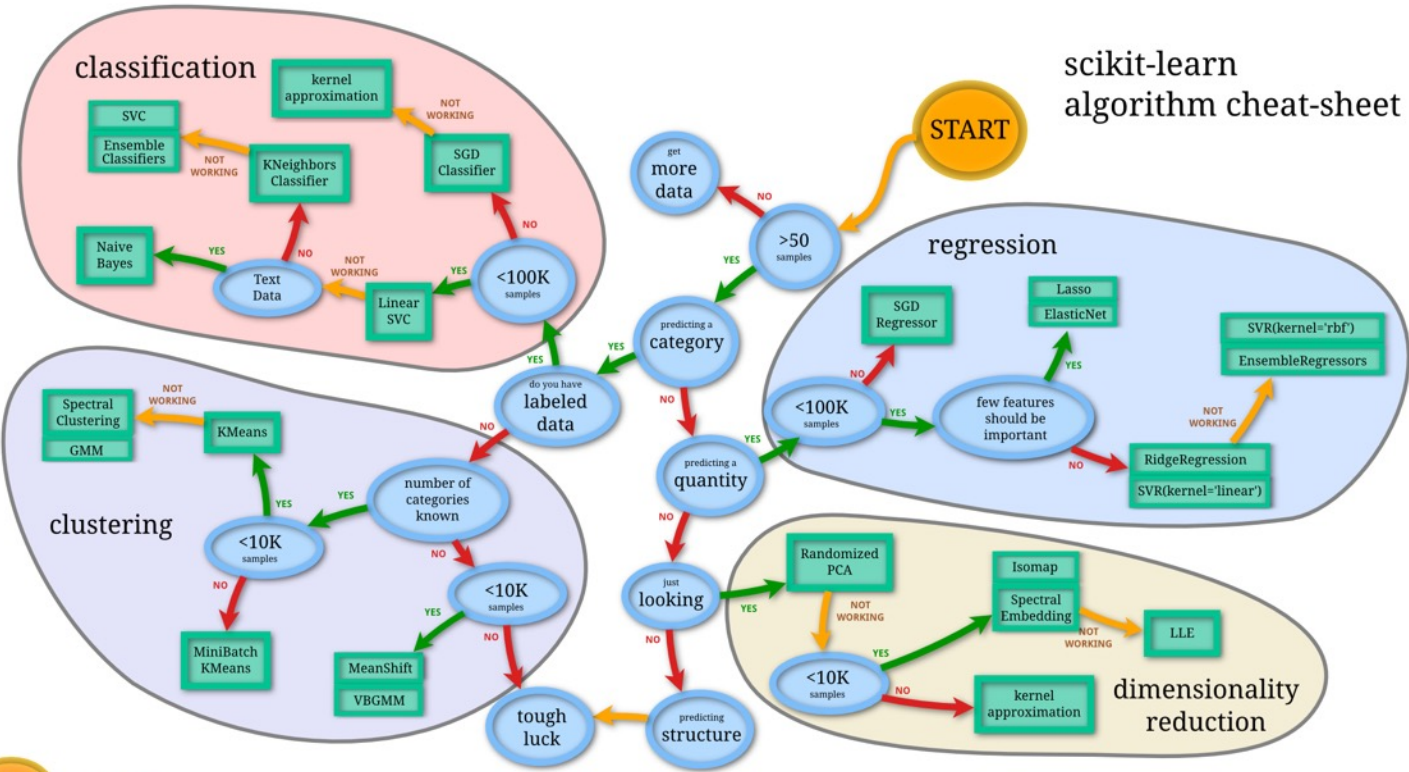
Machine Learning Roadmap



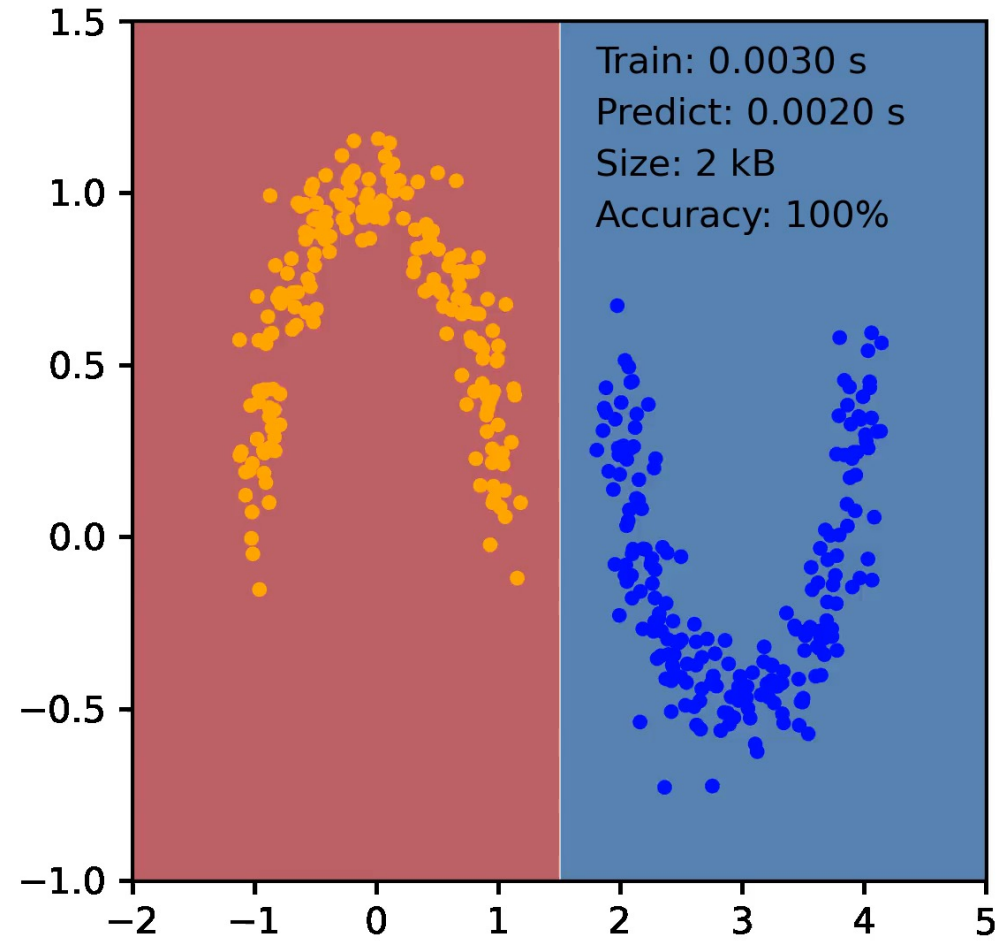
Data-Driven Modeling Languages



Navigate Machine Learning

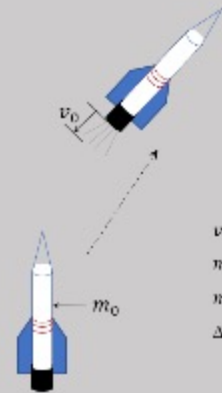
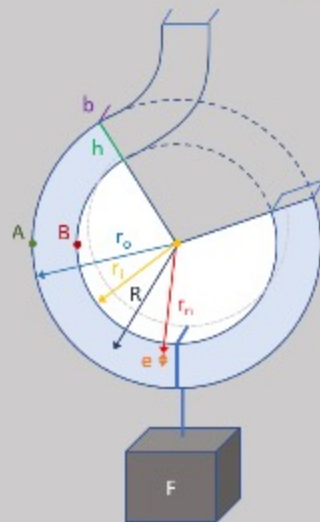
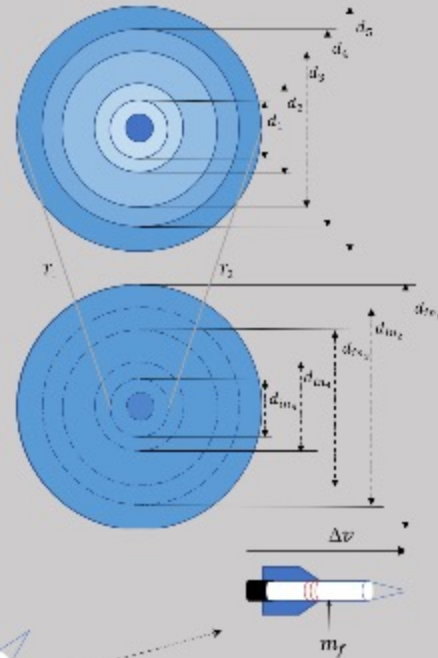
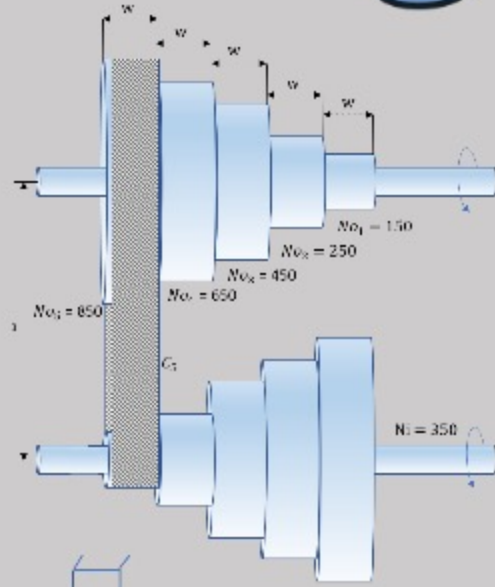


AdaBoost



Engineering Optimization

Introduction
Modeling
Unconstrained
Discrete
Genetic Alg
Constrained
Robust
Dynamic

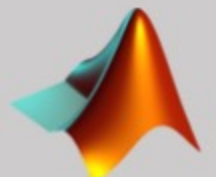


$v_0 \rightarrow$ exhaust velocity
 $m_0 \rightarrow$ initial or wet mass
 $m_f \rightarrow$ final or dry mass
 $\Delta v \rightarrow$ total change in velocity or final velocity



Optimization
Methods for
Engineering Design

Parkinson | Balling | Hedengren
Brigham Young University
Second Edition

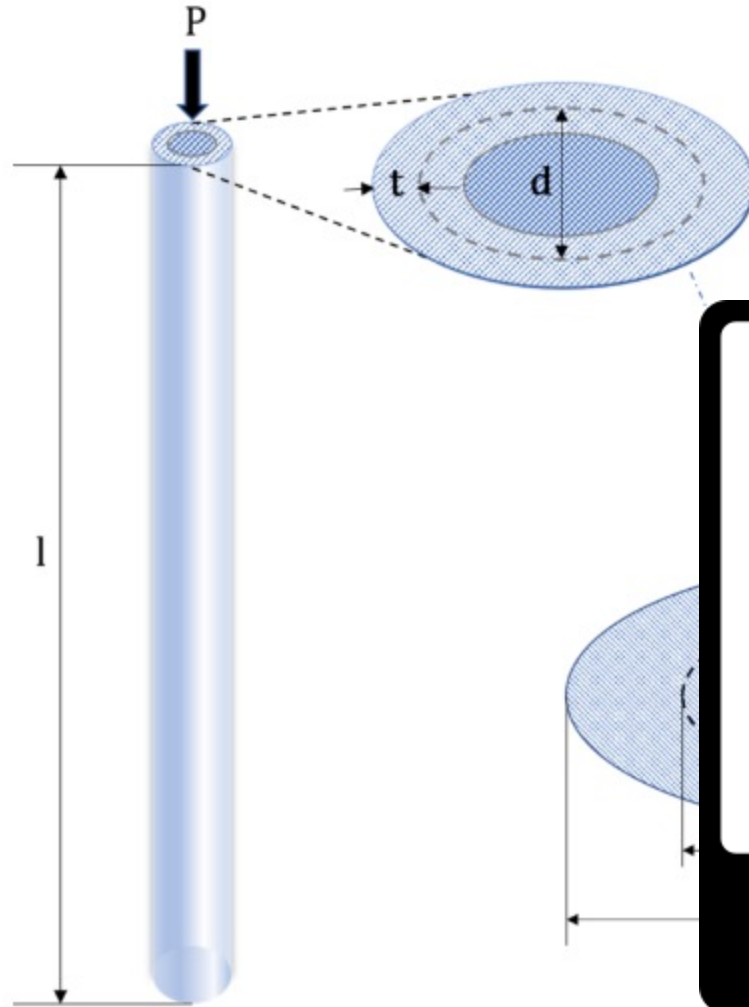




Engineering Optimization



Tubular Column



SCAN ME

ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests





Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

ew chat

optimize tubular colu  

umpy Array Modification.

ear conversations

grade to Plus

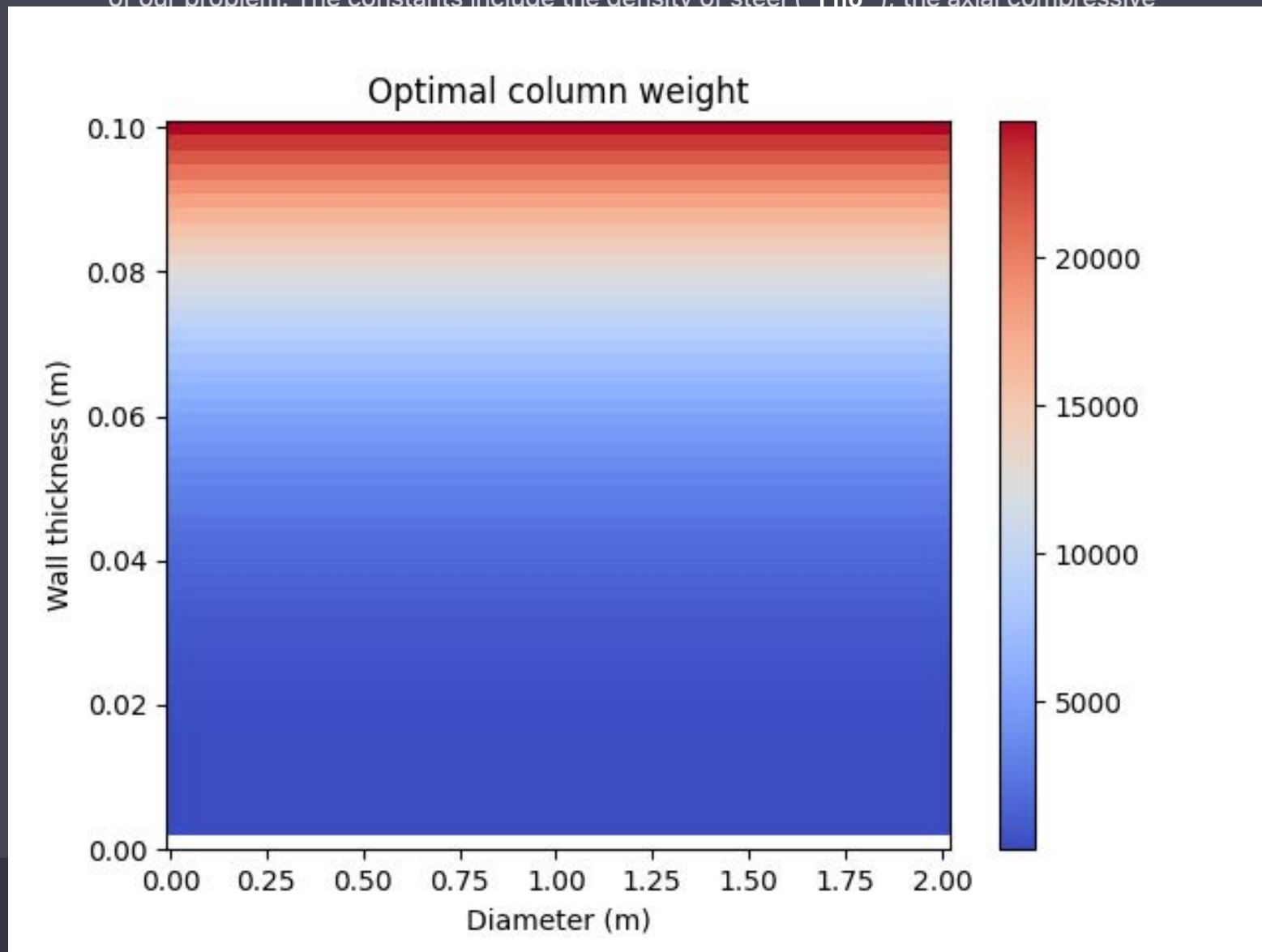
NEW


rk mode

dates & FAQ

g out

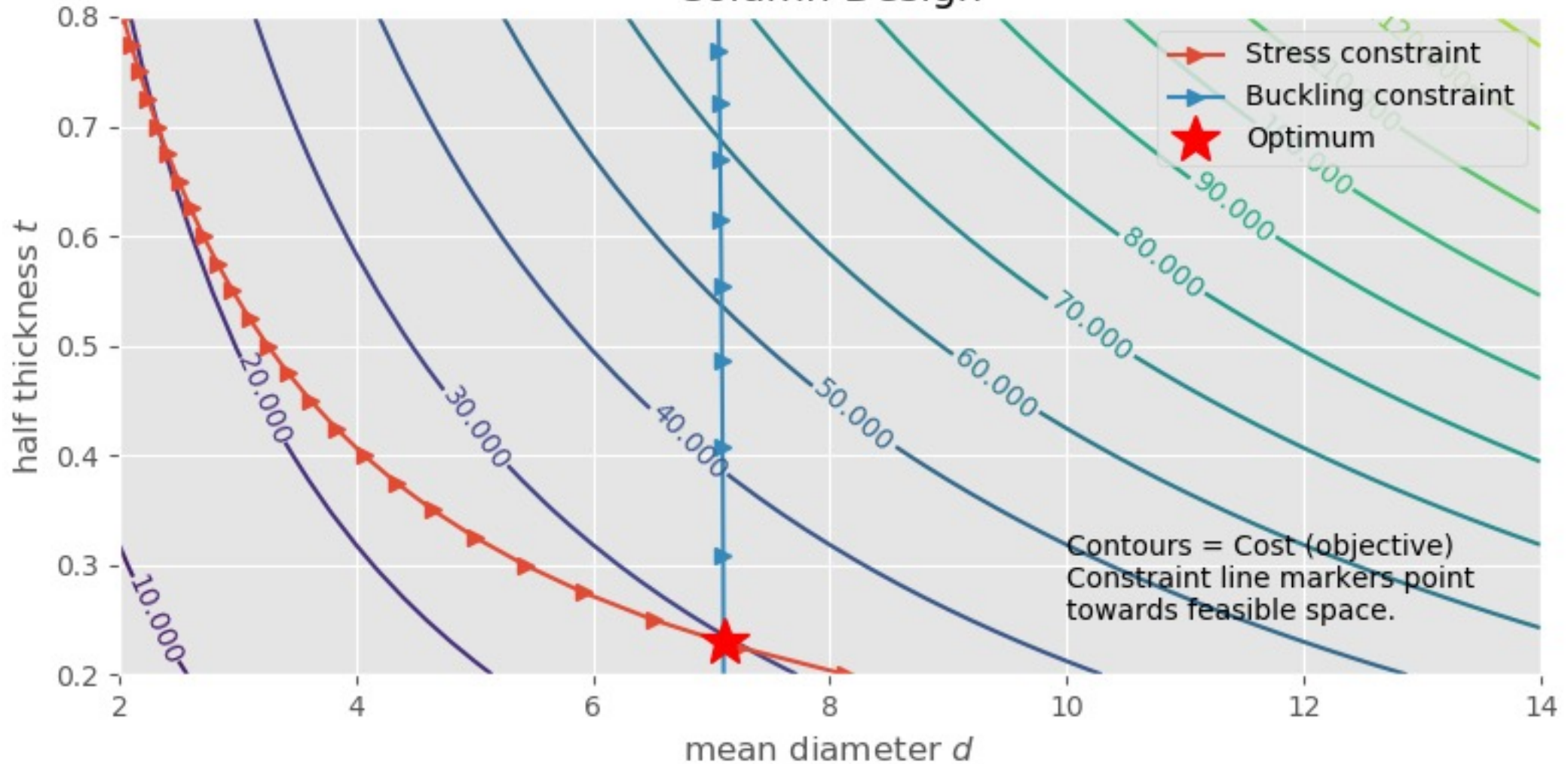
of our problem. The constants include the density of steel (ρ), the axial compressive



 Regenerate response

Create a col 

Column Design



Regenerate response

AI



Python and
MATLAB

Process Dynamics
and Control

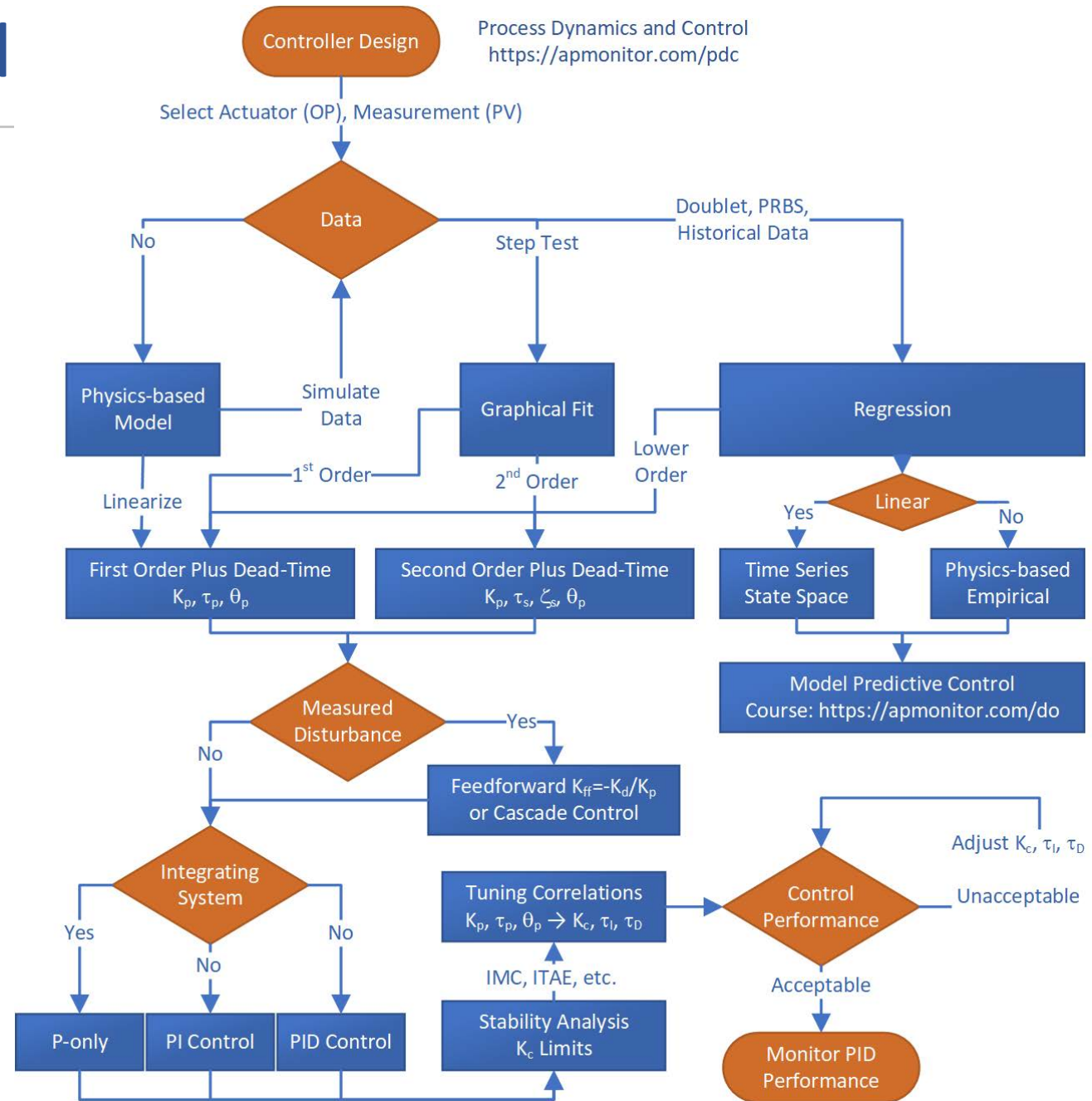
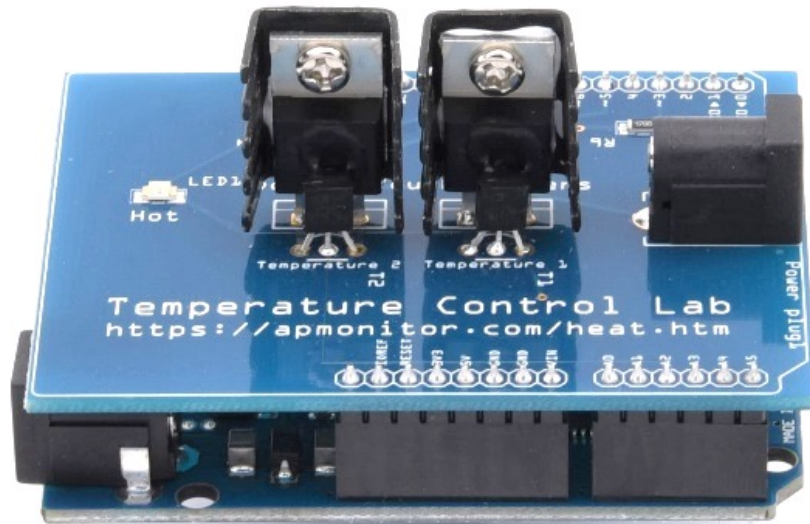


<https://apmonitor.com/pdc>

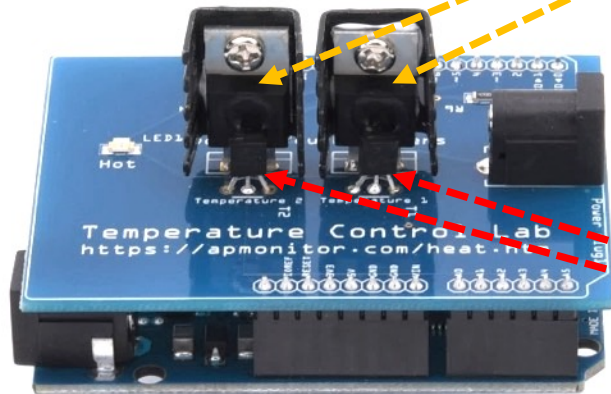
Theory
Simulation
Project



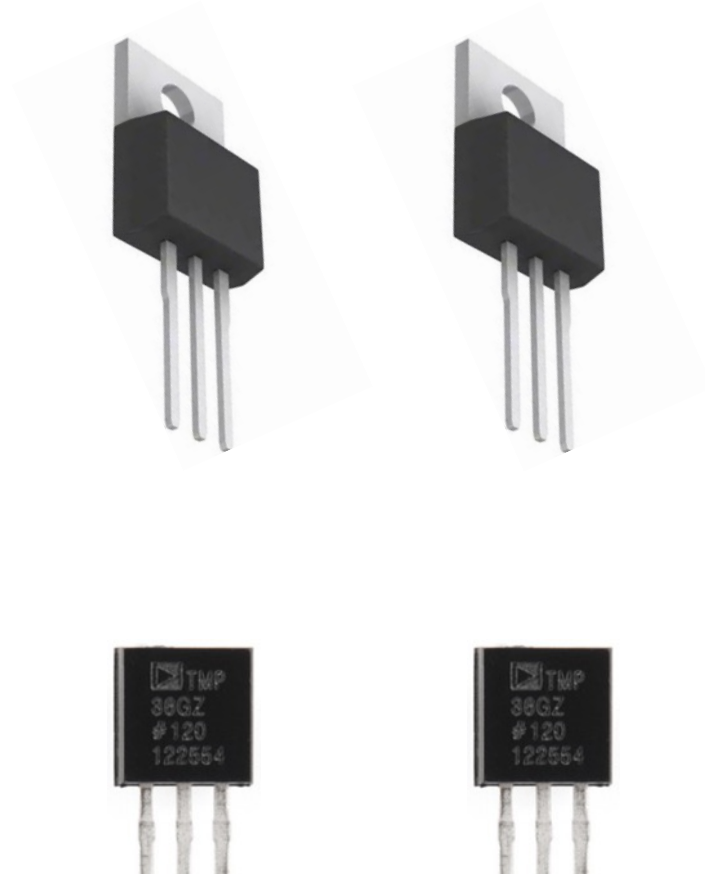
Process Dynamics and Control



Temperature Control Lab



apmonitor.com/heat.htm

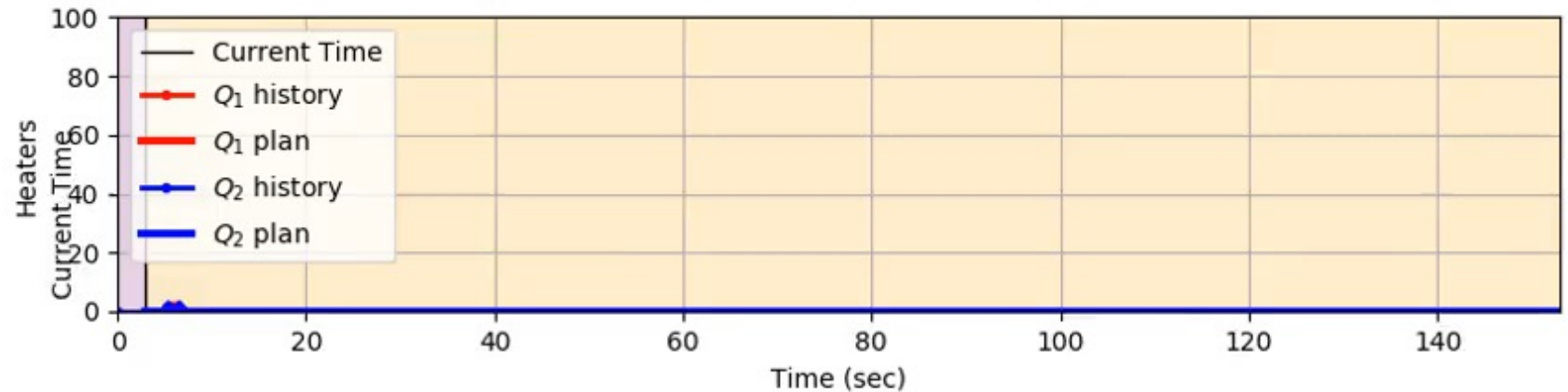
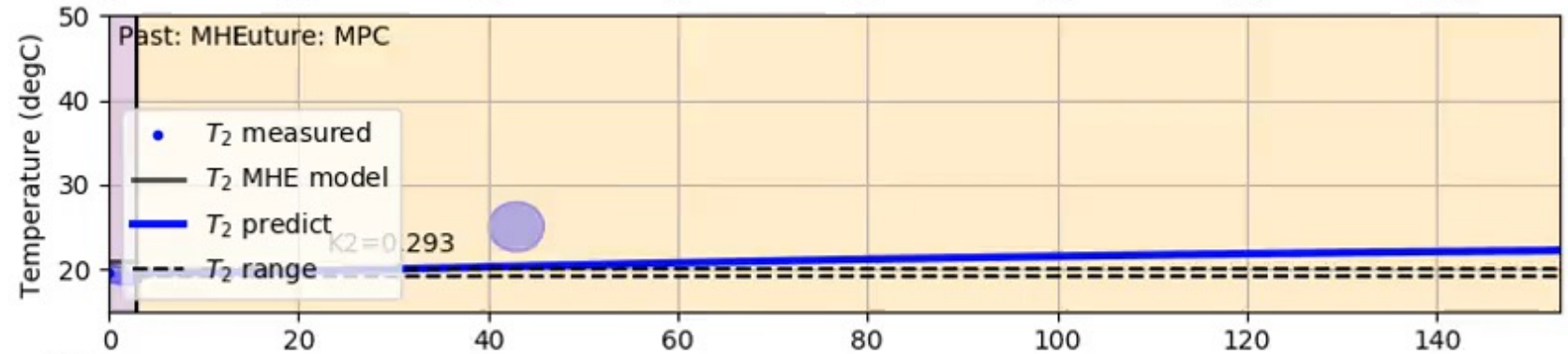
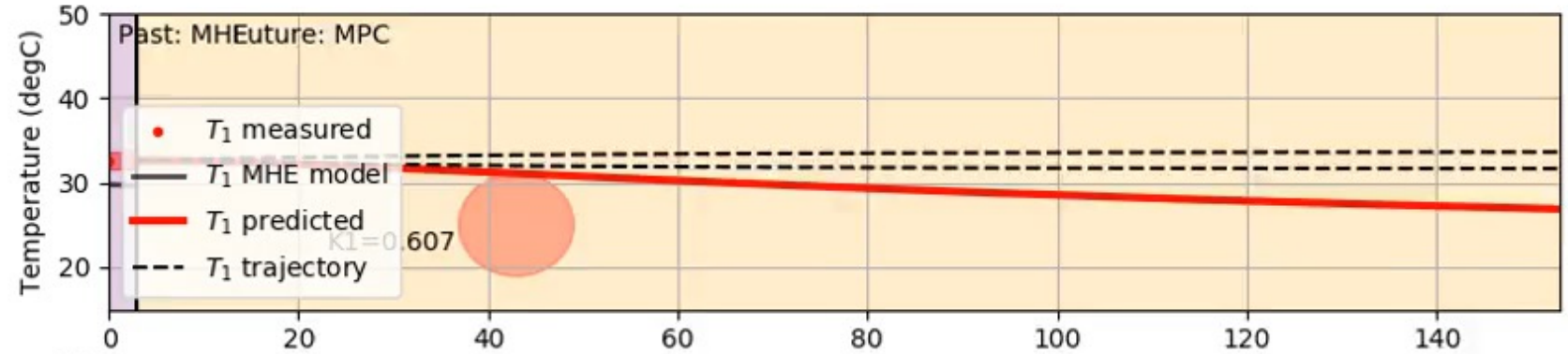
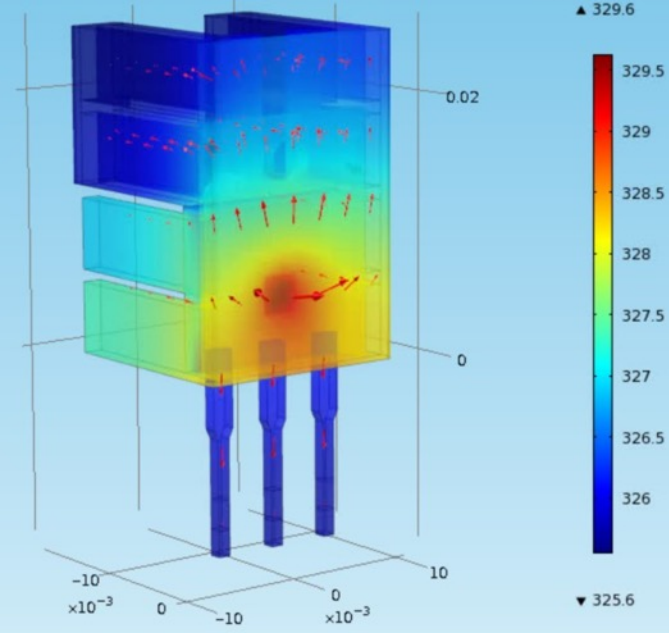
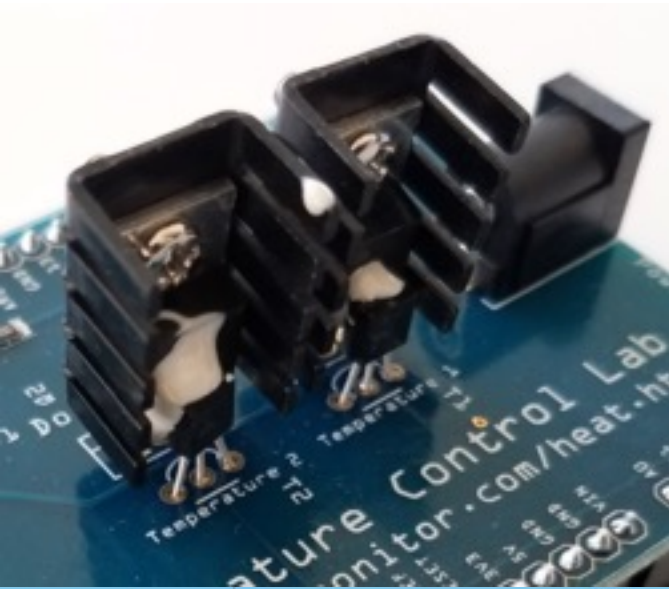


Actuator
Heaters



Sensors
T (°C)

Benchmark: Temperature Control Hardware



Data Availability



25 GB/hour



150,000 points/sec



51,200 GB/hr

Source: Simafone, Fortune, RTInsights, Cisco

Application: Flight Optimization

<http://prismweb.groups.et.byu.net/360/>



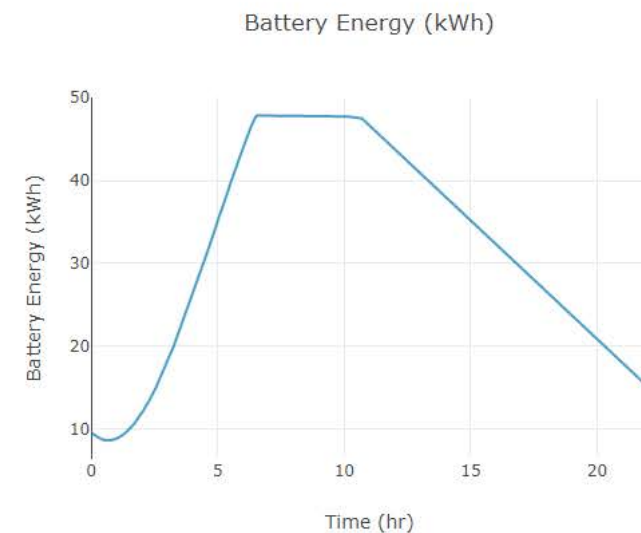
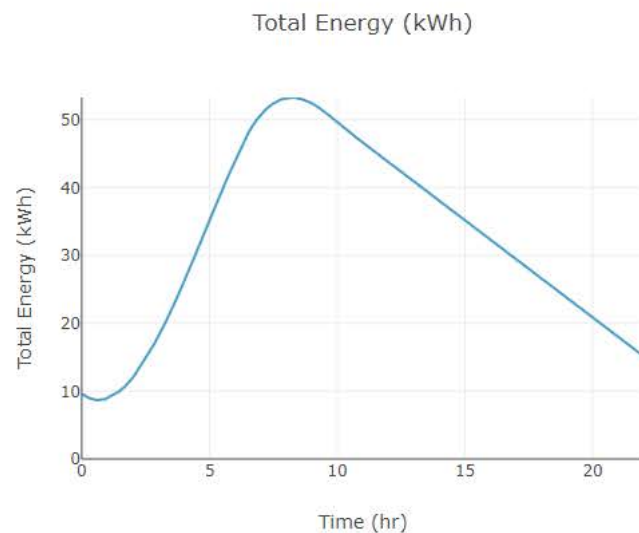
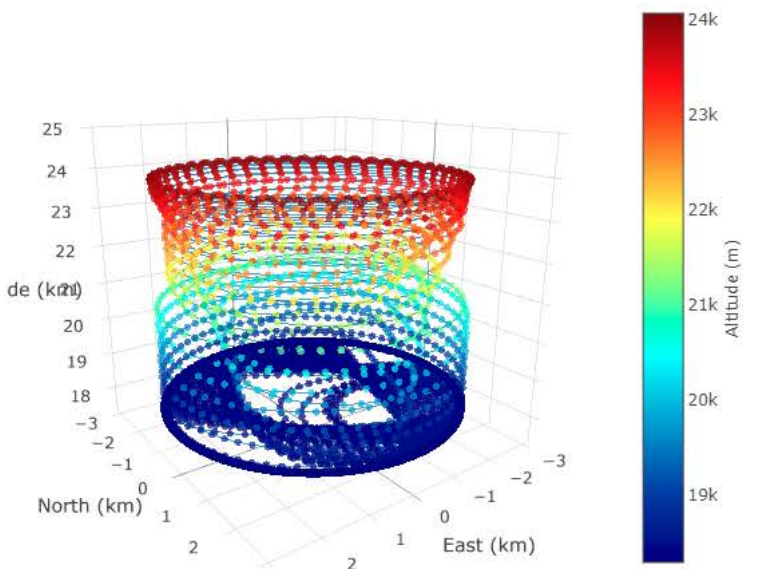
Select Variable
Altitude (m) x ▾

2D Hide Wind Hide Sun
 3D Show Wind Show Sun

Select Variable
Total Energy (kWh) x ▾

Select Variable
Battery Energy (kWh) x ▾

<https://github.com/BYU-PRISM/hale-trajectory>



Application: Biomechanics

Data-Driven Engineering

BIOMECHANIC Regression

predicted

measured

scikit learn

measured	predicted
2.2	2.2
2.4	2.4
2.6	2.6
2.8	2.8
3.0	3.0
3.2	3.2
3.4	3.4
3.6	3.6
3.8	3.8
4.0	4.0



MileSplit50: Jane Hedengren Moves To The No. 1 Spot



Behind The Cost Of Track Recruiting, As



Twenty Years Ago, Track Recruiting



Jason Vigilante Tabbed As



On The Line: Woodbridge, Addison



MileSplit50: Jane Hedengren Moves To

MileSplit50

XC RANKINGS

INDIVIDUAL - GIRLS



SEE MORE →

- 1 Jane Hedengren *Provo, UT*
- 2 Elizabeth Leachman *Boerne, TX*
- 3 Ellie Shea *, MA*
- 4 Sadie Engelhardt *Ventura, CA*
- 5 Isabel Allori *Fort Collins, CO*

f SHARE

🐦 TWEET

LIVE EVENT COVERAGE

2023 Dave Sanders Invitational
CO 2023-09-22

2023 Live in Lou XC Classic
KY 2023-09-29

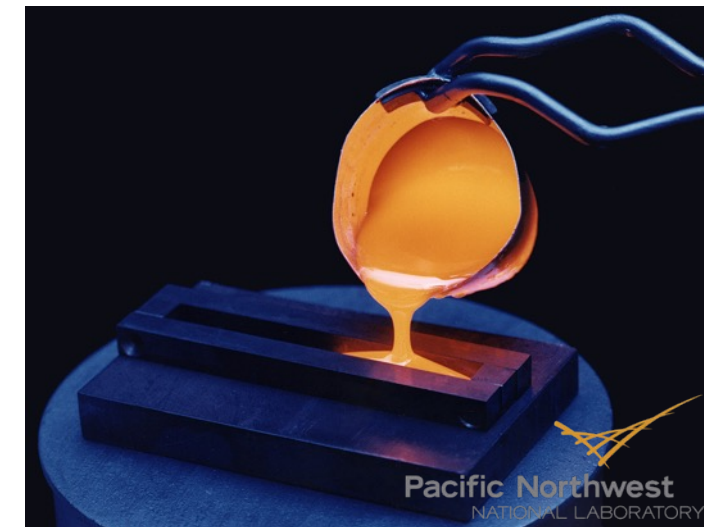
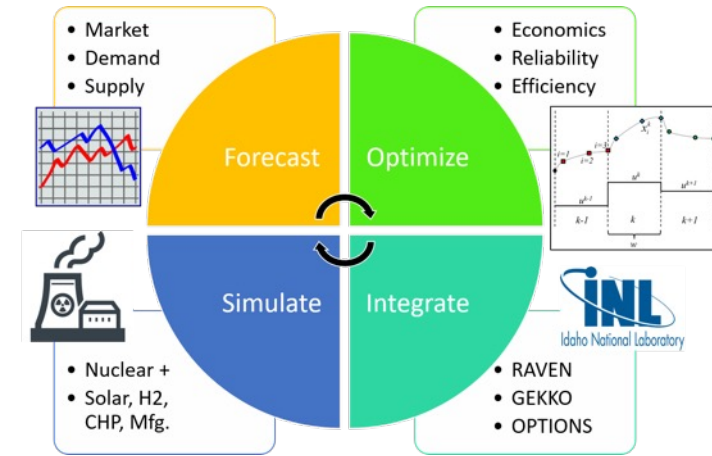
2023 FSU Pre-State Invitational
FL 2023-09-29

2023 McQuaid Invitational
NY 2023-09-30

Application: Drilling Automation

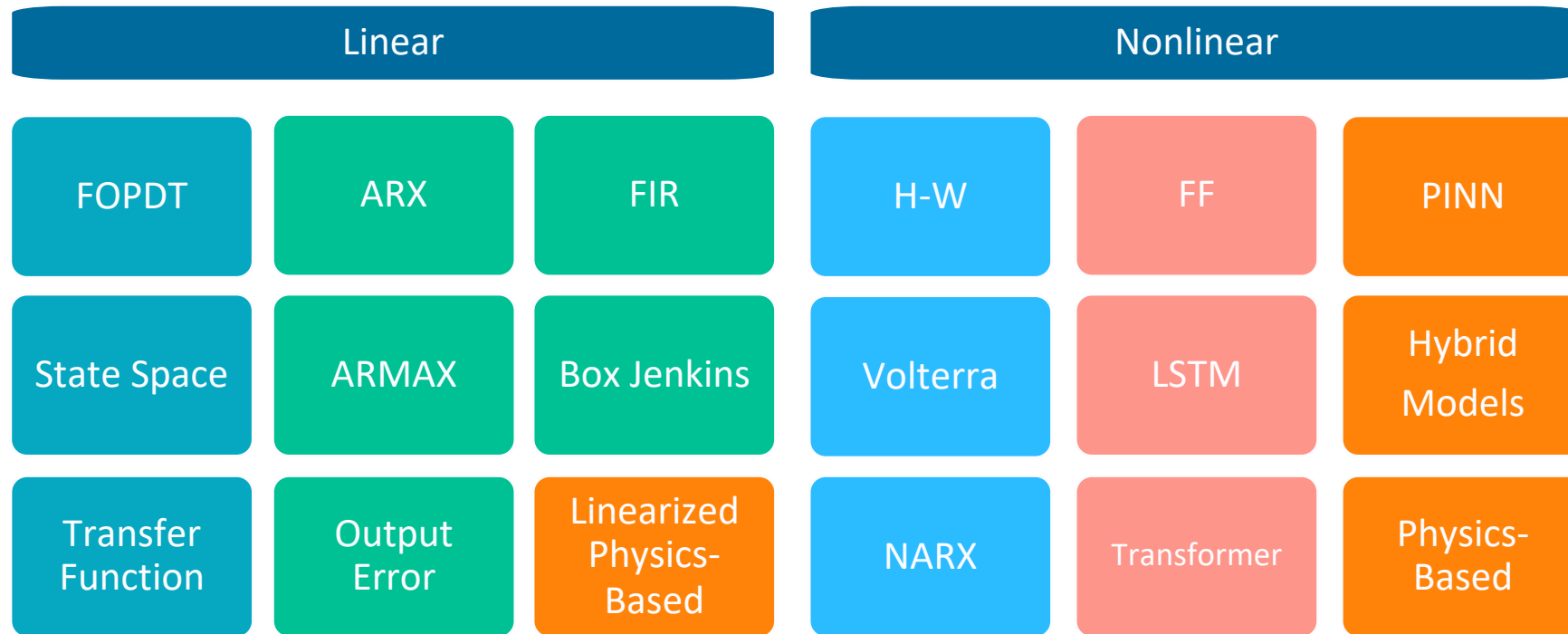


Physics-Informed, Data-Driven Modeling



Gunnell, L., Manwaring, K., Lu, X., Reynolds, J., Vienna, J., Hedengren, J.D., Machine Learning with Gradient-based Optimization of Nuclear Waste Vitrification with Uncertainties and Constraints, Processes, 10(11), 2365, Nov 2022, DOI: 10.3390/pr10112365.

Model Types



SysID Add-on Overview

Application

- Browser-based application
- Google-like search
- Tools for common functions
- Save and collaborate

Time Series Analytics

- Diagnostics analytics
- Monitoring and alerts
- Predictive analytics

The image shows a laptop displaying the Seeq Workbench interface. A large time series plot is visible, showing data points and a fitted model. Overlaid on the right side of the laptop is a 'SysID' configuration panel. This panel includes sections for 'Time Series Settings', 'Manipulated Variables (MV)' (with 'F_sw' selected), 'Measured Variables (CV)' (with 'CA1, CA2, T1, T2' selected), 'Model Structure', 'Training Conditions' (with 'Manual Condition 1' and 'Manual Condition 2' selected), and 'Validation Conditions' (with 'Manual Condition 3' selected). Below these settings is a 'TRAIN' and 'VALIDATION' plot showing 'Value' vs 'Time' with data points and model fits. At the bottom of the SysID panel is a table for variable styling:

Name	Unit	Color	Line Style
<input checked="" type="checkbox"/> CA1_pred	-	Blue	—
<input type="checkbox"/> CA2_pred	-	Red	—
<input type="checkbox"/> T1_pred	-	Green	—
<input type="checkbox"/> T2_pred	-	Purple	—
<input checked="" type="checkbox"/> CA1	-	Blue	—
<input type="checkbox"/> CA2	-	Red	—
<input type="checkbox"/> T1	-	Green	—

Advanced Analytics

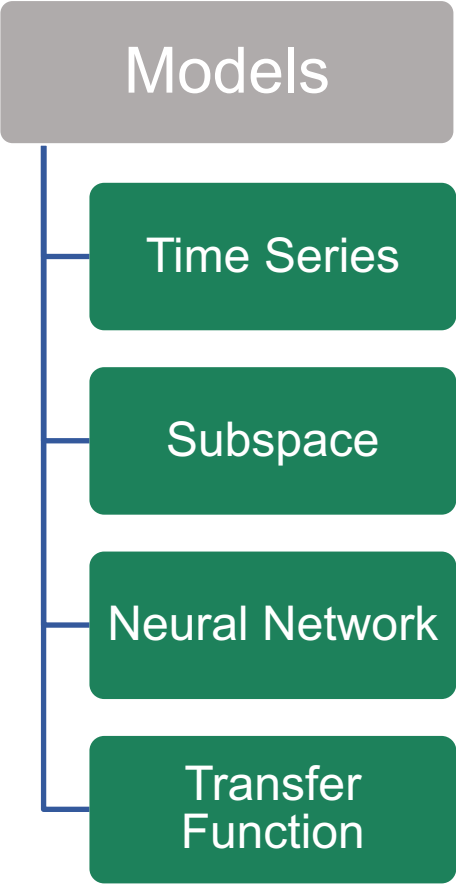
- Data cleansing
- Pattern recognition
- Scalable calculations
- Machine learning



`pip install seeq-sysid`



System Identification (SysID) Add-on



The screenshot displays the Seeq software interface for system identification. At the top, the Seeq logo is on the left, and navigation tabs for 'TIME SERIES', 'SUBSPACE', 'NEURAL NETWORK', and 'TRANSFER FUNCTION' are on the right. The main content area is split into two panels:

- Time Series Settings (Left Panel):** A dark green sidebar containing:
 - Time Series Settings** (with close and share icons)
 - Manipulated Variables (MV):** A dropdown menu with 'Select'.
 - Measured Variables (CV):** A dropdown menu with 'Select'.
 - Model Structure:** A dropdown menu with a downward arrow.
 - Training Conditions:** A dropdown menu with 'All Data'.
 - Validation Conditions:** A dropdown menu with 'All Data'.
 - Buttons for 'TRAIN', 'VALIDATE', and 'PLOT MODEL' at the bottom.
- Data Visualization (Right Panel):** A plot area with a light blue grid. The y-axis is labeled 'Value' (range -1 to 4) and the x-axis is labeled 'Time' (range 0 to 5). The plot is divided into 'TRAIN' (Time 0-2) and 'VALIDATION' (Time 2-5) regions. Below the plot is a table with columns for Name, Unit, Color, and Line Style, which currently shows 'No Data Available'.

Time Series Settings

Manipulated Variables (MV)

Select

Controlled Variables (CV)

Select

Model Structure

Training Conditions

All Data

Validation Conditions

All Data

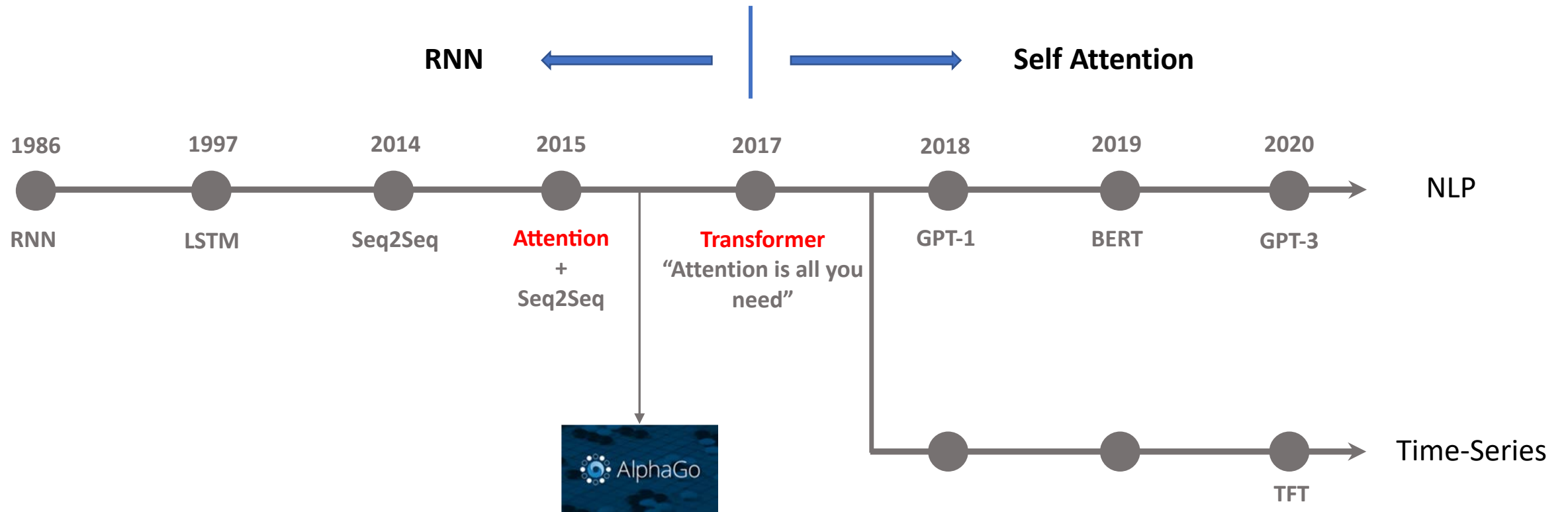
Identify Validate Push Model



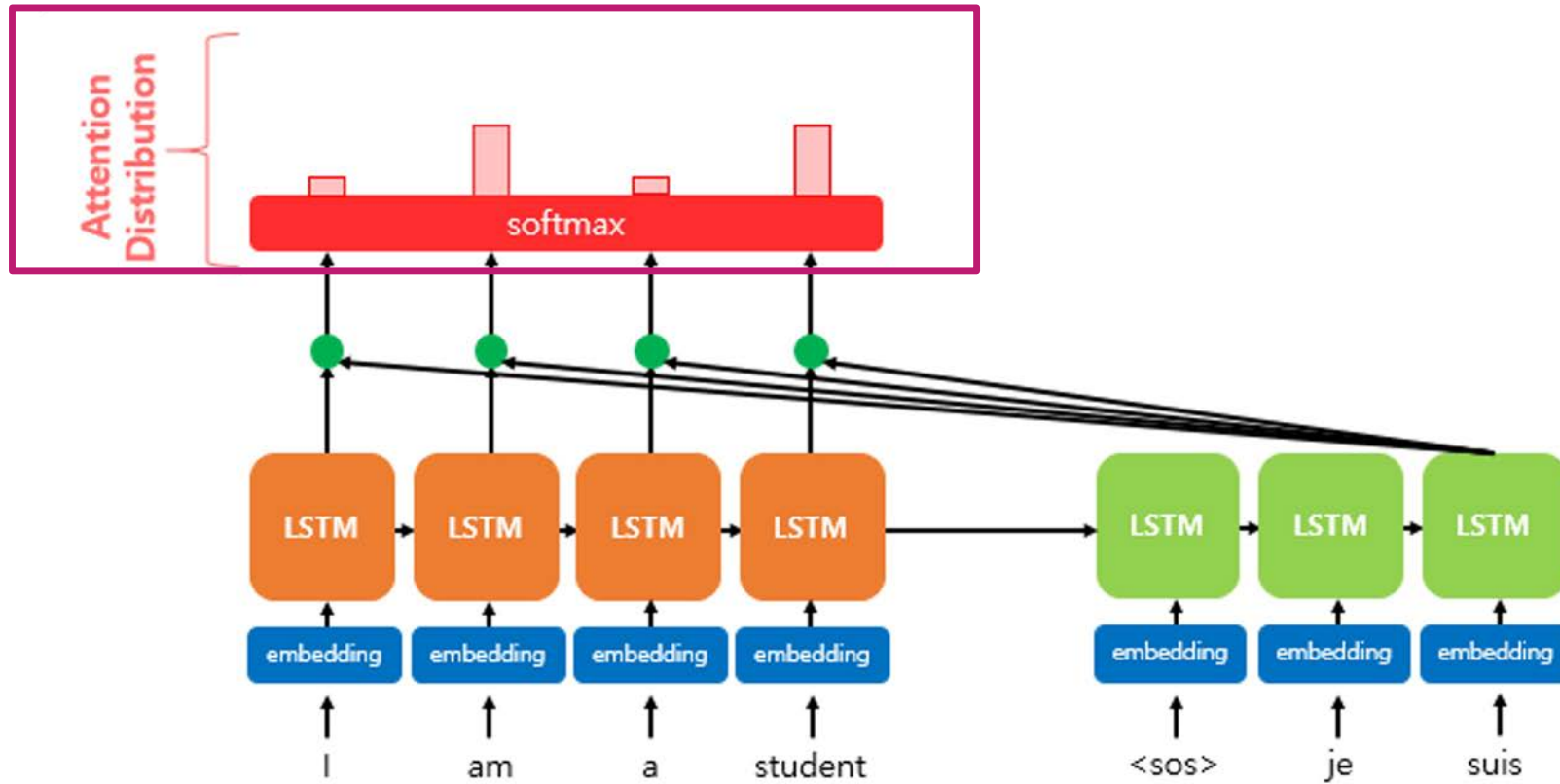
<input type="checkbox"/>	Name	Unit	Color	Line Style
No Data Available				

1. Select Data
2. Select Model
3. Identify
4. Push Model

Neural Network Models for Sequence Data

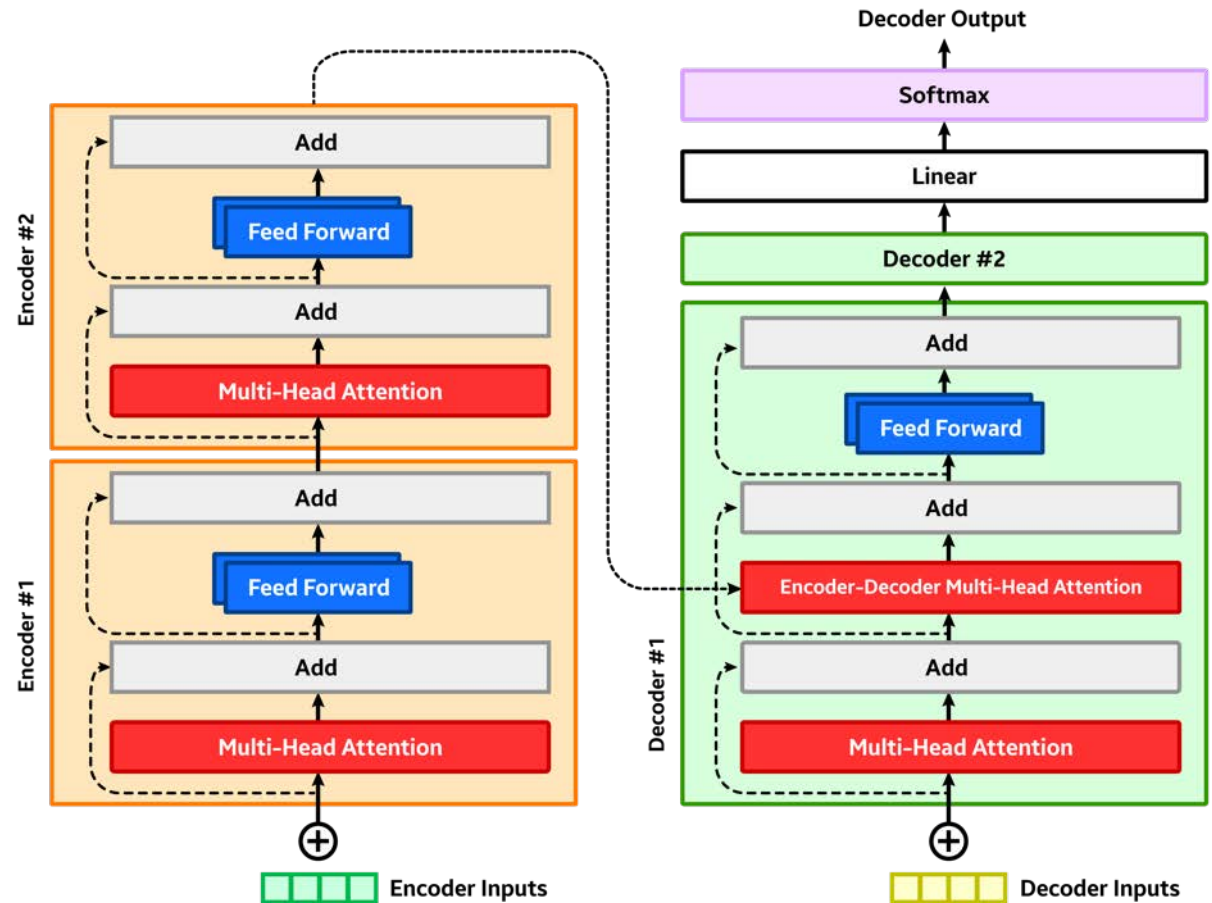


Attention Mechanism



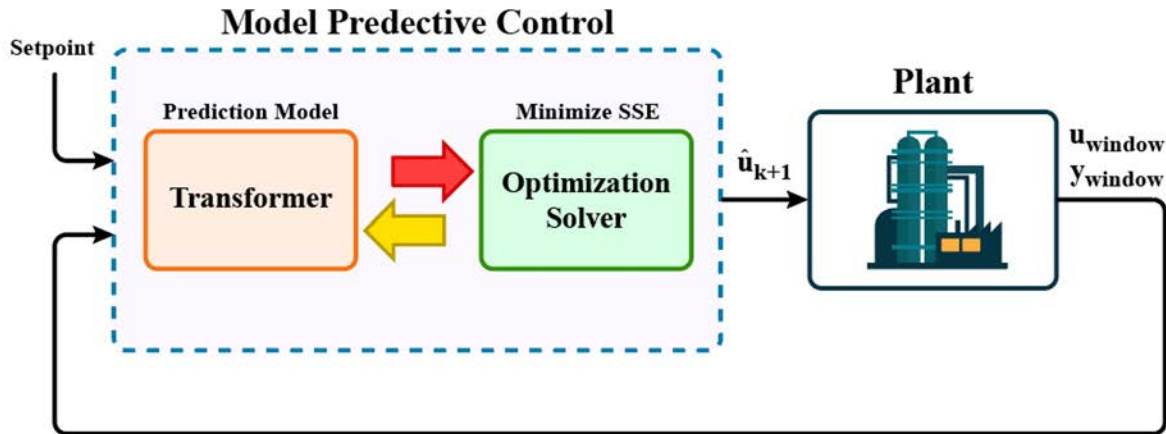
Transformer Architecture (Self-attention)

- Short processing time
- No vanishing gradient
- Captures irregular temporal dependency



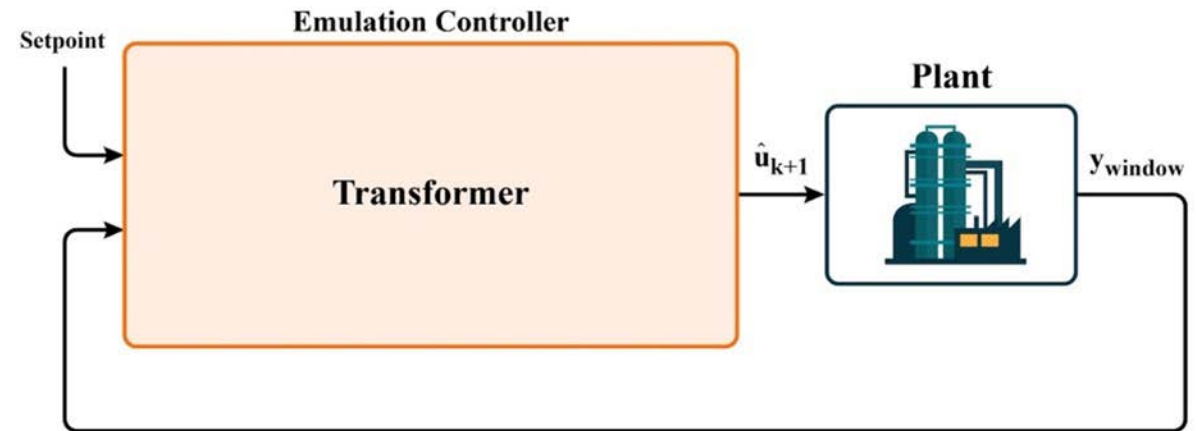
Model Predictive Control (Two options)

Surrogate MPC



Trained by Open-loop data

Emulation MPC



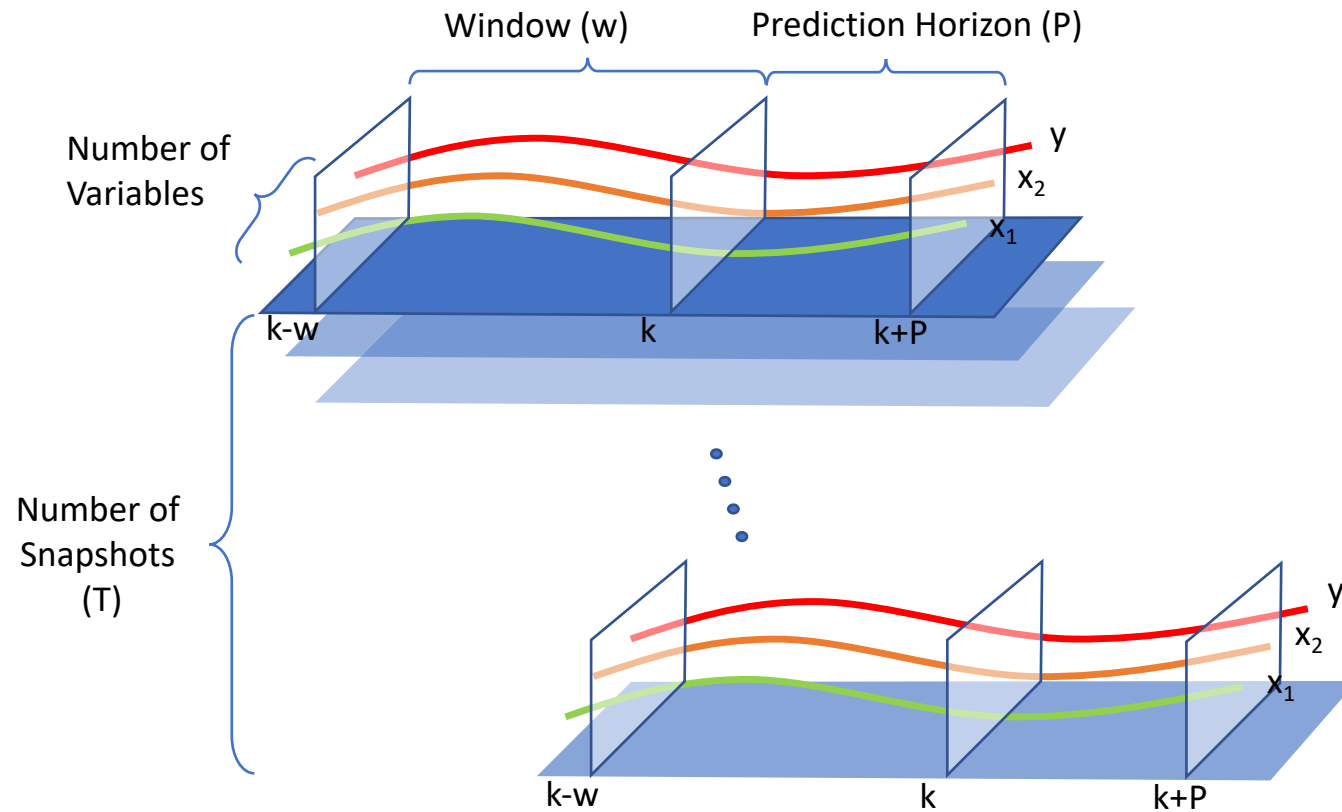
Trained by Closed-loop data

Fast (No optimization step)

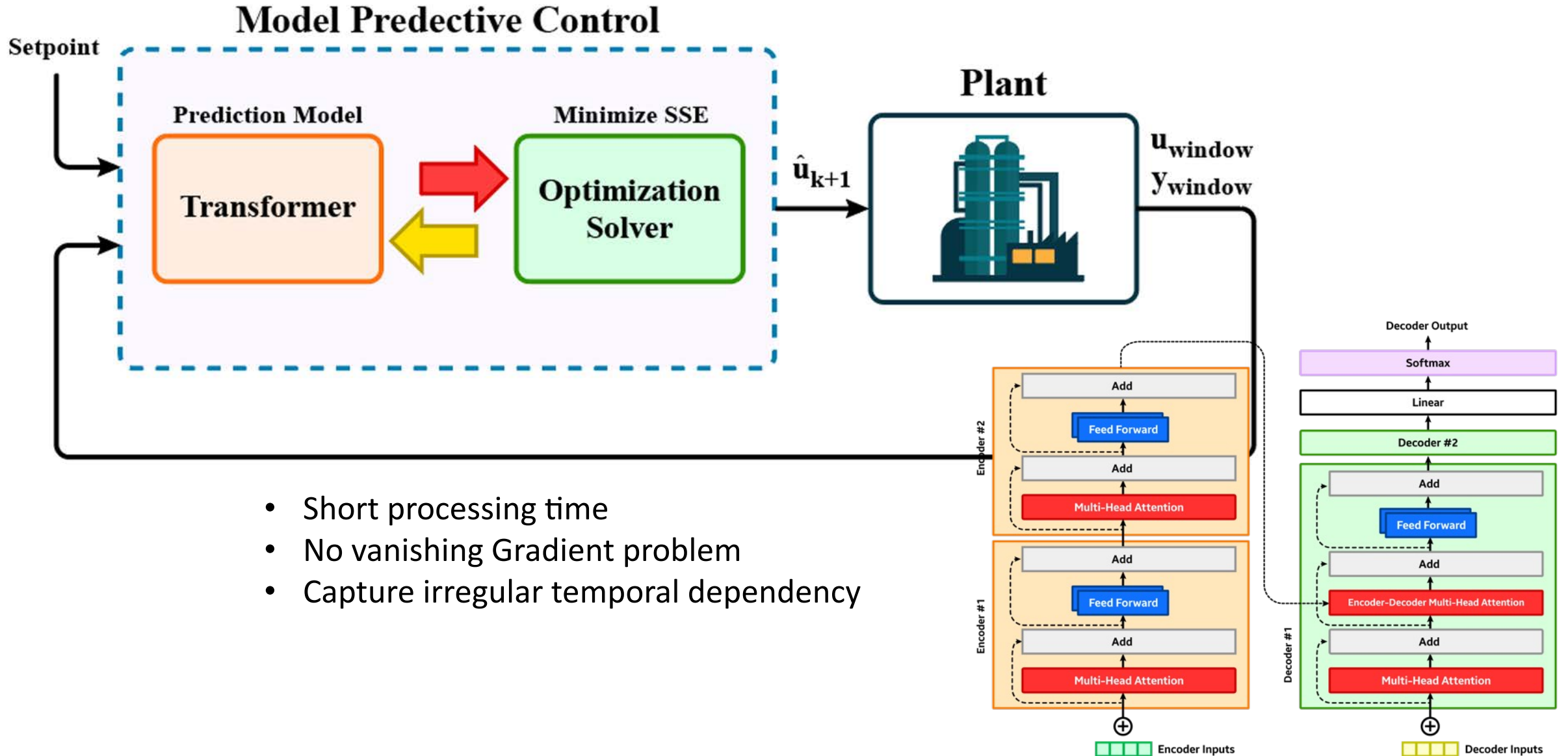
No online correction - can't guarantee the performance

Training Data Preparation

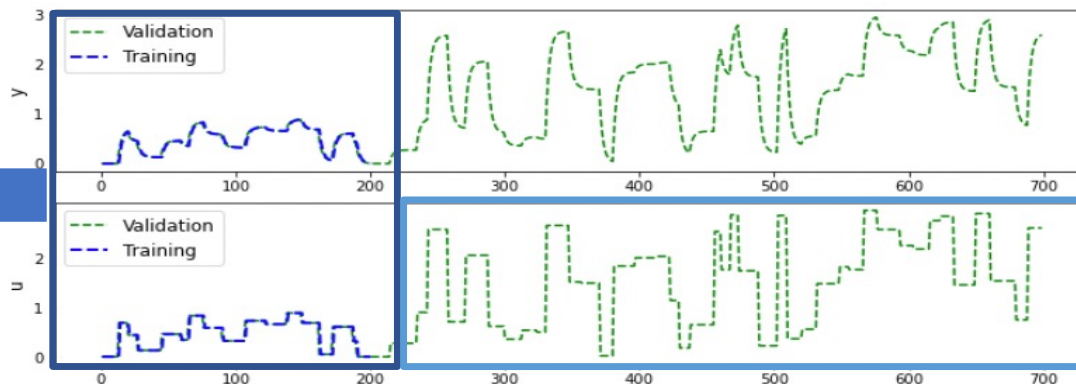
Receding Window Snapshots



Future of Data-Driven Control



Physics Informed Neural Network (PINN)

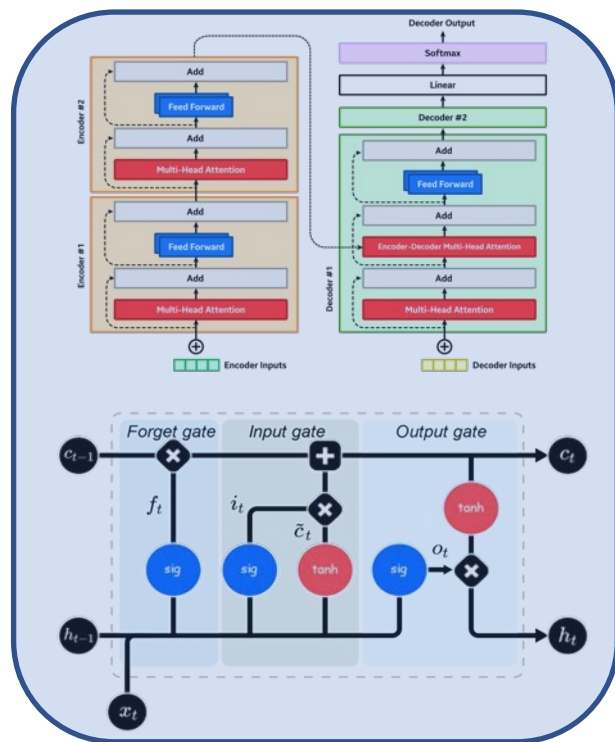


Training

y_{meas} ,
 u_{meas}

u_{meas}

Evaluating



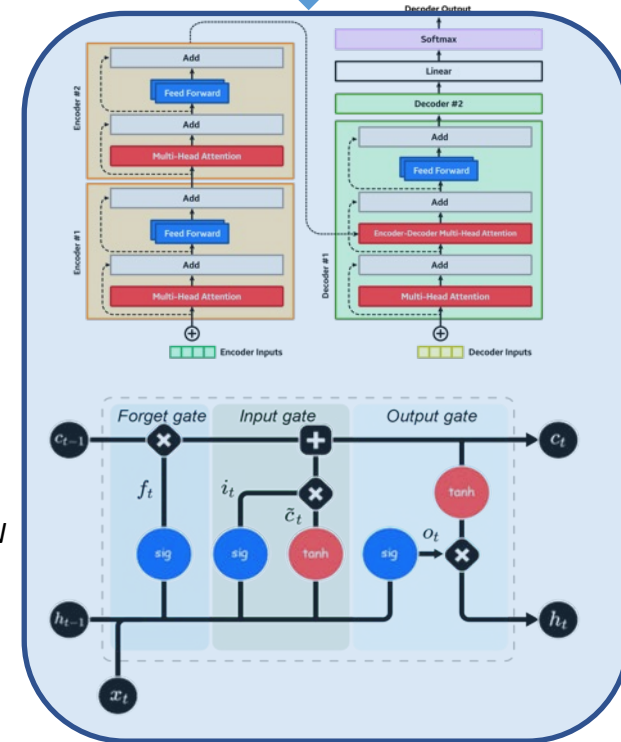
$$MSE_y = \frac{1}{N_y} \sum_{i=1}^{N_y} |y_{NN}(t_i, u_i) - y_{meas}|^2$$

+

$$MSE_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_i, \dot{y}_{NN}, y_{NN}, u_{meas})|^2$$

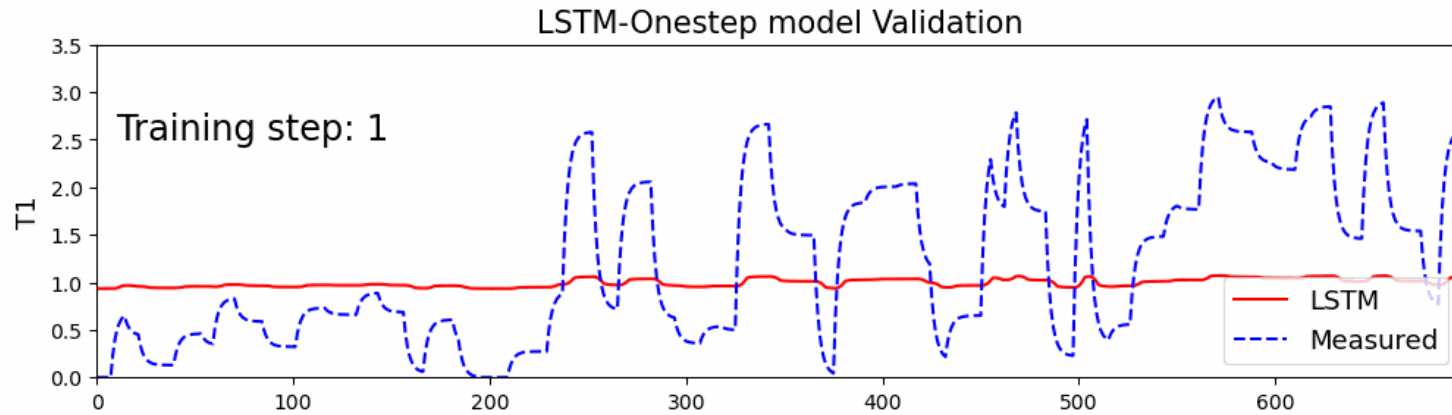
Combined Loss function

y_{NN}

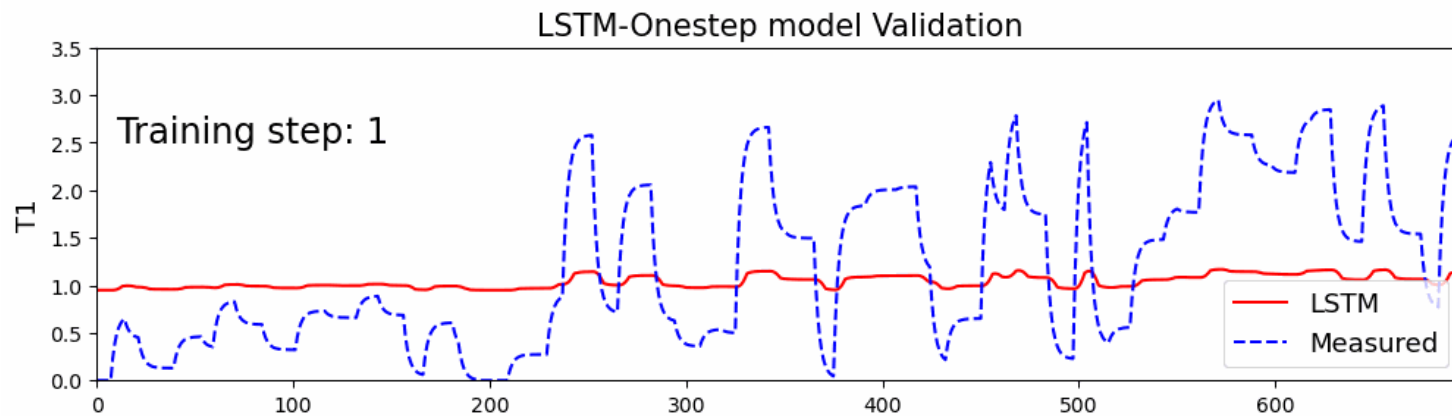


Physics Informed Neural Network (PINN)

PINN Off



PINN On



BYU PRISM
MODEL OPTIMIZE CONTROL

ROOM

Logos: Facebook, SUSTAINABLE ENERGY SOLUTIONS, RODMAX, Schlumberger, equinor, bp, USTAR, astro, Reliable controls, U.S. DEPARTMENT OF ENERGY, NOV, Seeq



BYU
BRIGHAM YOUNG UNIVERSITY
FOUNDED 1875
PROVO, UTAH

BYU Chemical Engineering
IRA A. FULTON COLLEGE OF ENGINEERING



SCAN ME





BYU PRISM
MODEL OPTIMIZE CONTROL