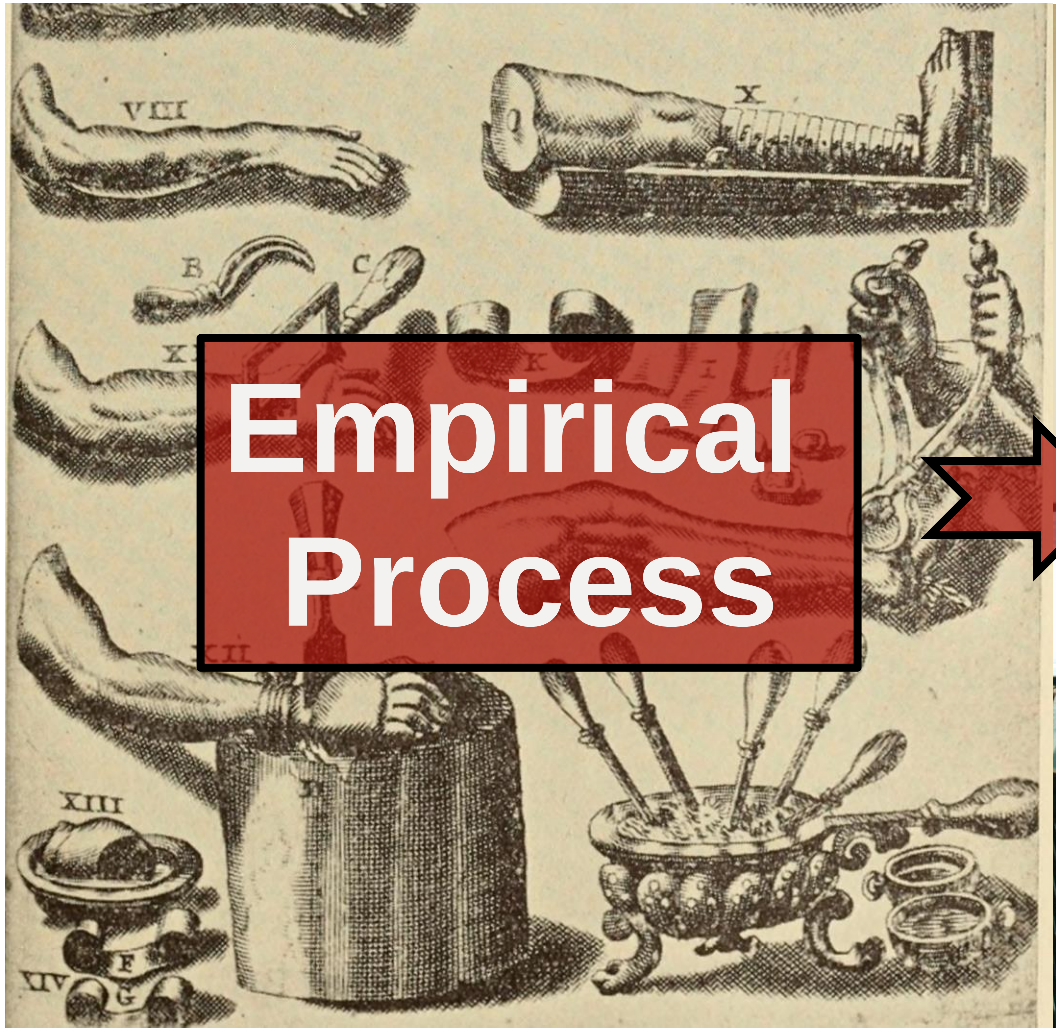
LiCO: Simplifying and accelerating deep learning application development

David Ellison, PhD
2018-05-24

Lessons From Medicine: ‘Modernizing’ Deep Learning Model Development



Empirical Process → **Experimental Process**

Experimental Process for Deep Learning

Design	Execute
Planned	Controlled
Ethical	Reproducible

AI Use Cases that Lenovo is Exploring Today

For better quality control on

Manufacturing Lines

Mark III Systems

Lenovo

For proactive actions improving

Water & Energy Conservation

NC STATE UNIVERSITY

Lenovo

To advance particle physics at

CERN's Large Hadron Collider

UCL

Lenovo

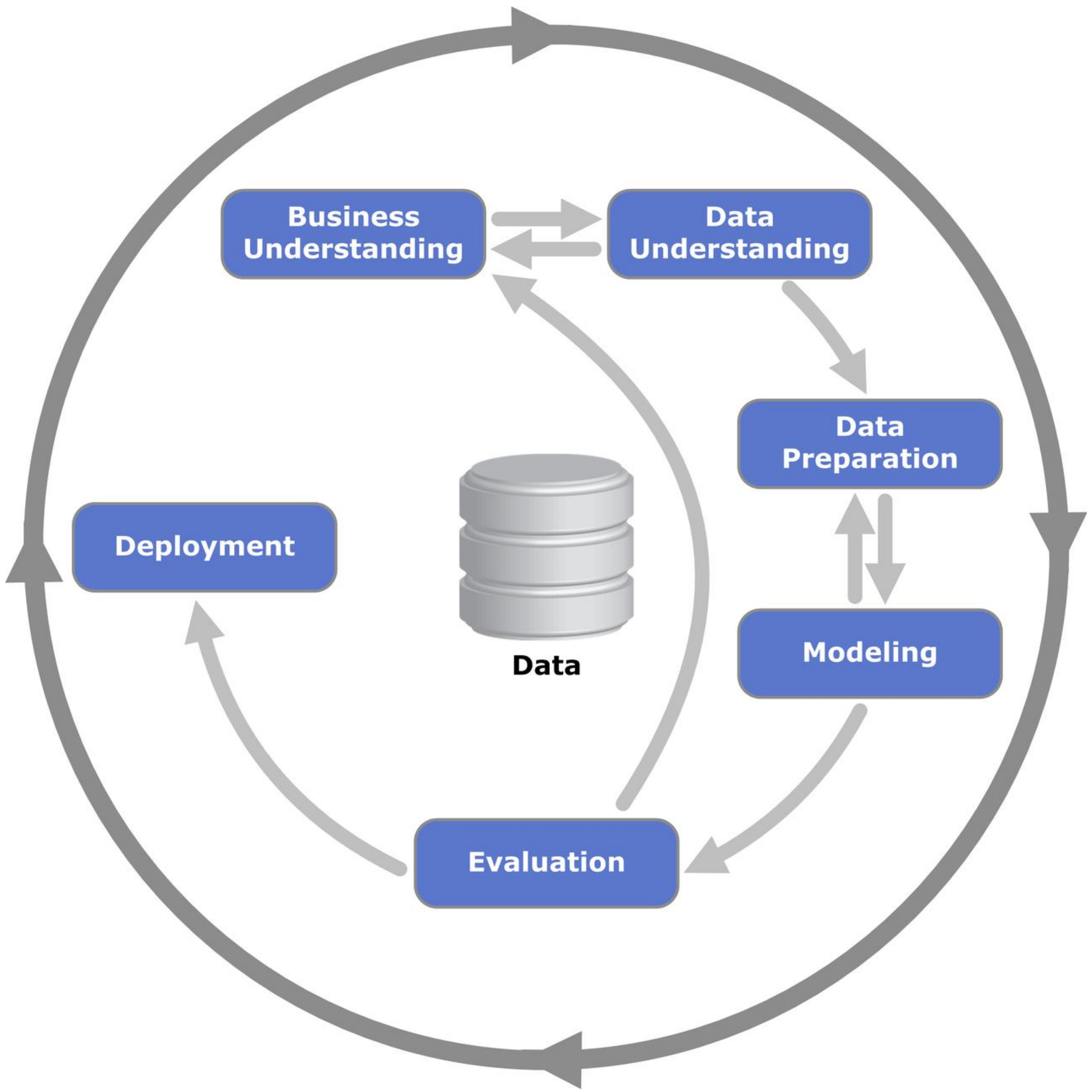
A more efficient and effective

Radiology Practice

CodaLab

Lenovo

Current Models of Development (CRISP-DM)



Kenneth Jensen

xkcd

Progression vs. Safety

Leonard McCombe/Time Life Pictures/Getty Images

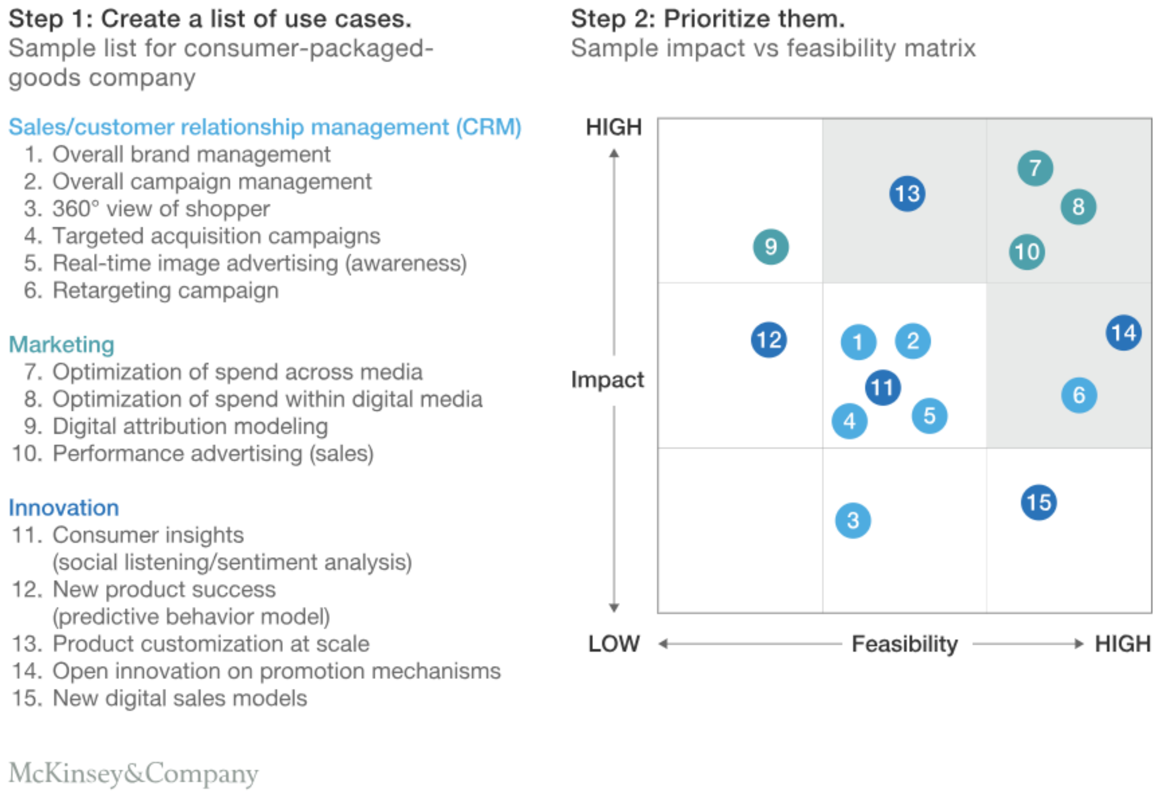
CNBC & Guyana News

Experimental Process for Deep Learning

Design	Execute
Planned	Controlled
Ethical	Reproducible

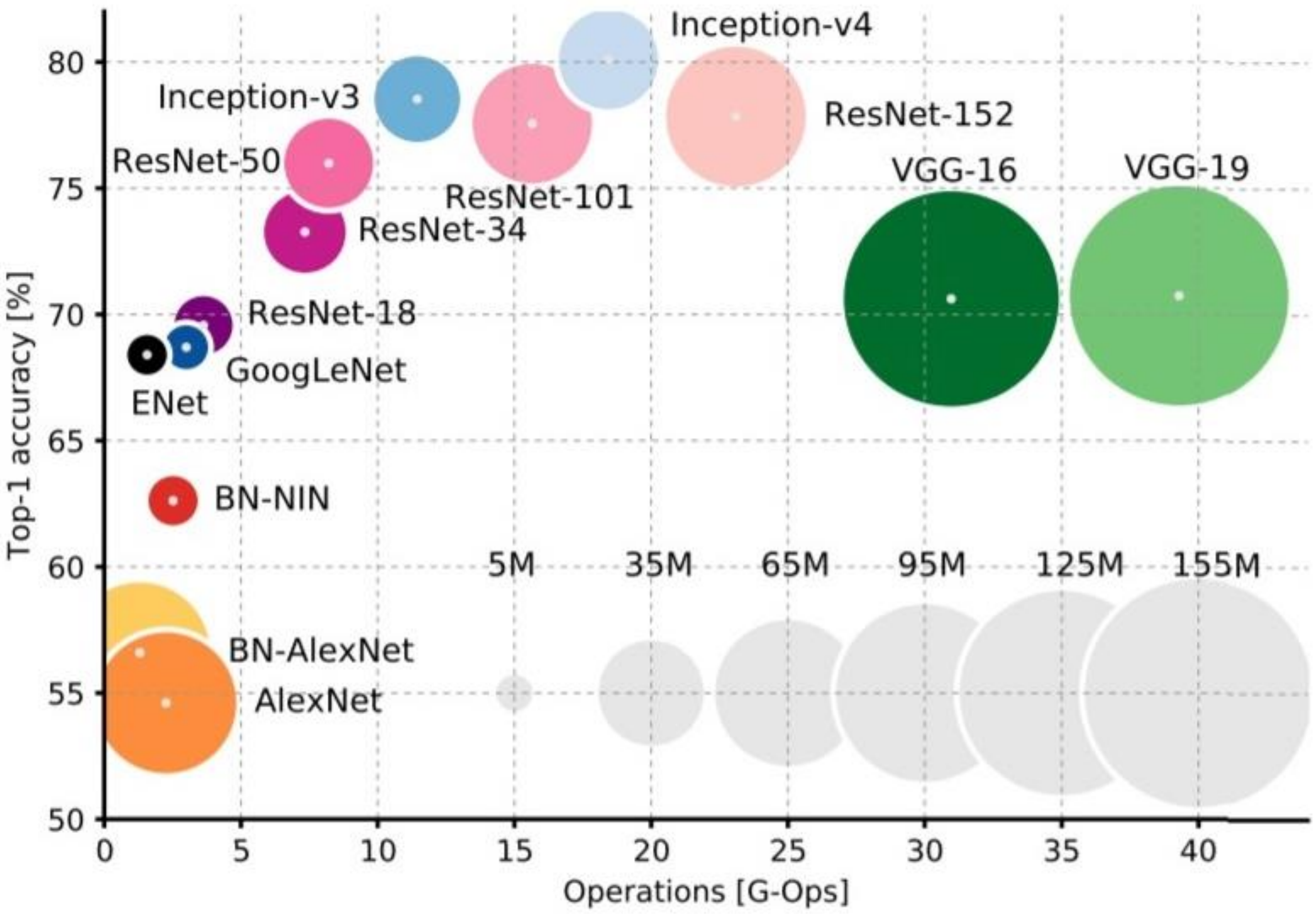
How Will We Measure Success?

1. Choose metrics first
2. Get data and access early
3. Start collecting historical data ASAP
4. Plan for data restructuring



McKinsey Co 2018, Ten Red Flags
Signaling Your Analytics Program Will Fail

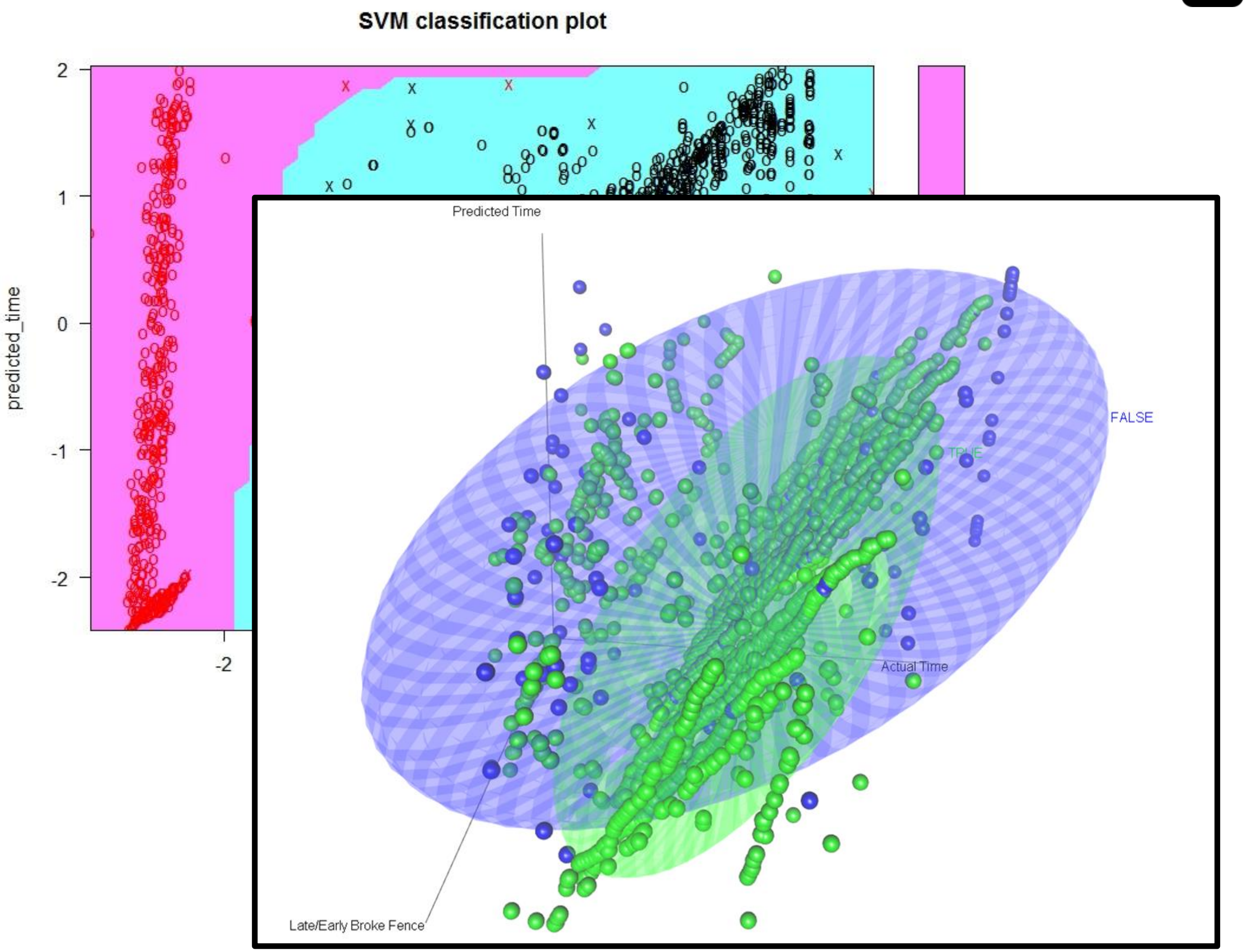
Trade-off Between Metrics



Canzaini et al 2017 An Analysis of Deep Neural Network Models for Practical Applications

Deploy Simplest Model Possible

1. Statistical learning
2. Machine learning
3. Deep learning



David Ellison, unpublished data

How Will the Deep Learning Model be Used?

Know how to:

1. Get data to your model
2. Evaluate your model
3. Use results of your model

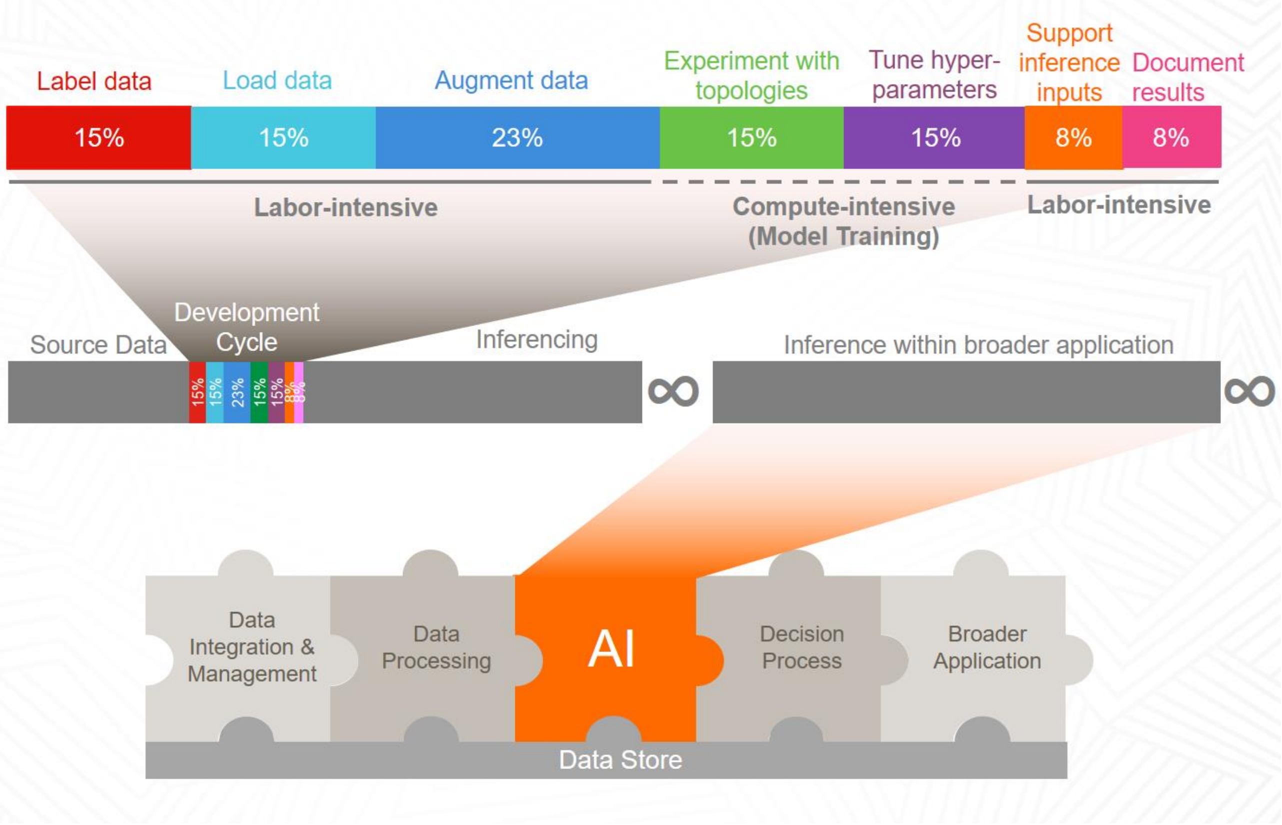


Image from Intel Customer Engagements

Metrics and Decision-Making

1. What metrics will be used to evaluate ethics?
   - Excluding obviously unethical features is not enough
   - What is the balance of fairness and safety?

2. Who decides what is ethical?
   - Programmers/Management
   - External board

3. At what point is it unethical NOT to use an algorithm?

Top Image: Ohio State Bar, individuals pictured are actors/models
Bottom Image: Uber dash cam of Tempe, AZ accident

Safety in AI

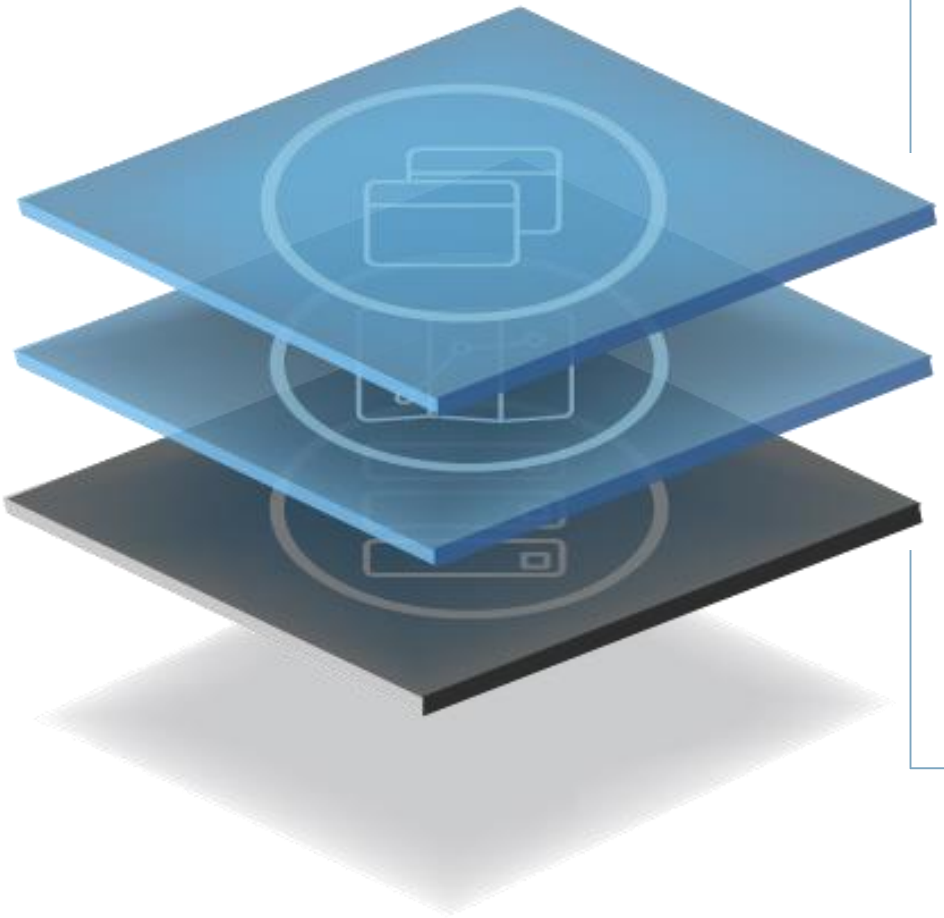
- Avoid Negative Side Effects
  - Cleaning robot should not knock things over to clean faster. Should not manually specify rules
- Avoiding Reward Hacking
  - Ensure the cleaning robot won't game its own reward function. Might cover over messes, or hide not to tell humans about the mess
- Scalable Oversight
  - How to control for big, infrequent mistakes that require human intervention. How does the cleaning robot it treat candy wrappers vs. lost cellphones
- Safe Exploration
  - Explore new space without obvious mistakes. Cleaning robot attempting to wet mop electric outlets
- Robustness to Distributional Shifts
  - Adapting to a change in environment, e.g. office cleaning vs factory cleaning

Amodei et al, 2016 Concrete Problems in AI Safety, p3

Experimental Process for Deep Learning

Design	Execute
Planned	Controlled
Ethical	Reproducible

LiCO Stack for **Artificial Intelligence**



- Web Portal
- Pre-trained Models
- Workflow Templates

- Distributed AI Training
- AI Frameworks

TensorFlow Caffe mxnet

- Open-Source Cluster Management
- Optimized HW Libraries

intel NVIDIA

LiCO for **Cluster Deployment**

Easily manage AI & HPC clusters, workloads & users.

Efficiently manages **workflows** for AI training tasks

- Easy-to-use GUI & templates to simplify job management
- Leverage pre-trained models, monitor training & manage job history for rapid deployment.
- Multi-user support & intuitive admin controls.

Simplifies open source deployment in enterprises

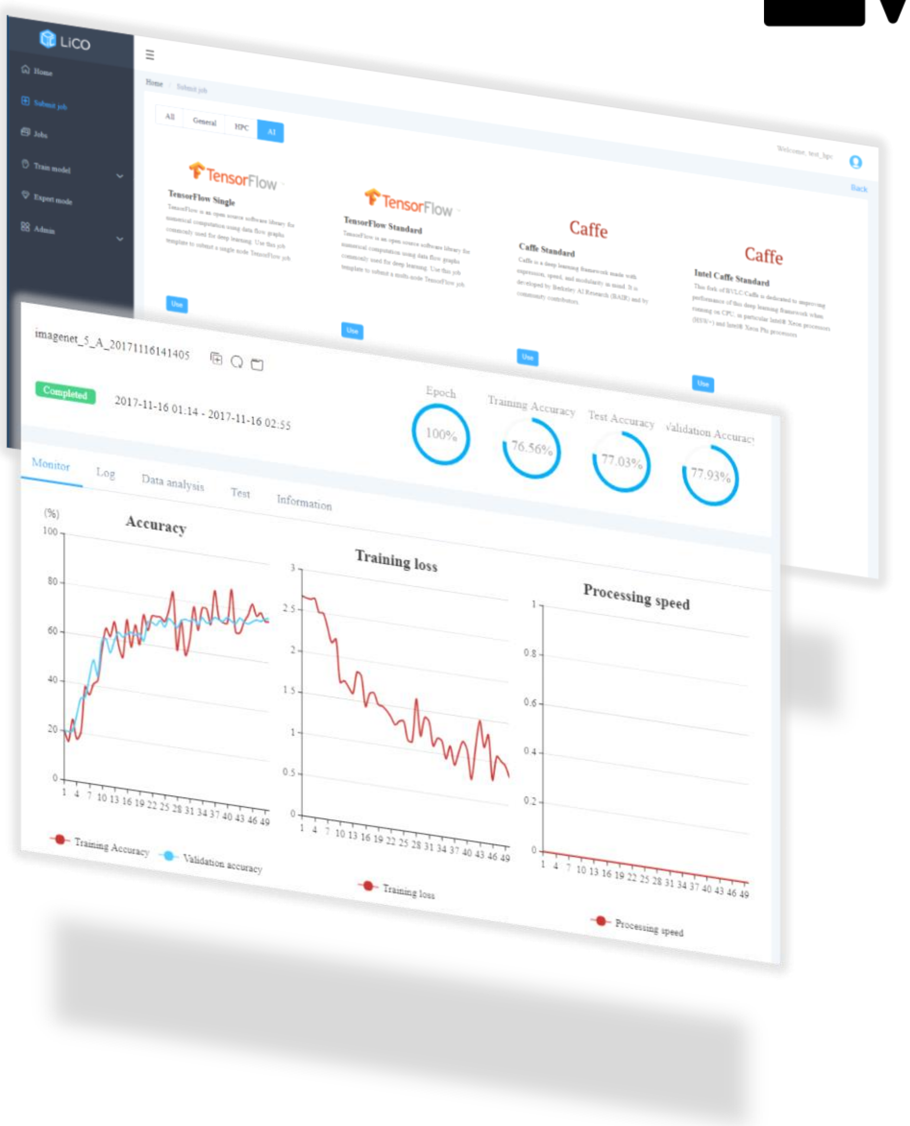
- Validated stacks reduce time-to-development
- Choice of multiple AI frameworks through containers

Offers **flexibility** with hardware infrastructure

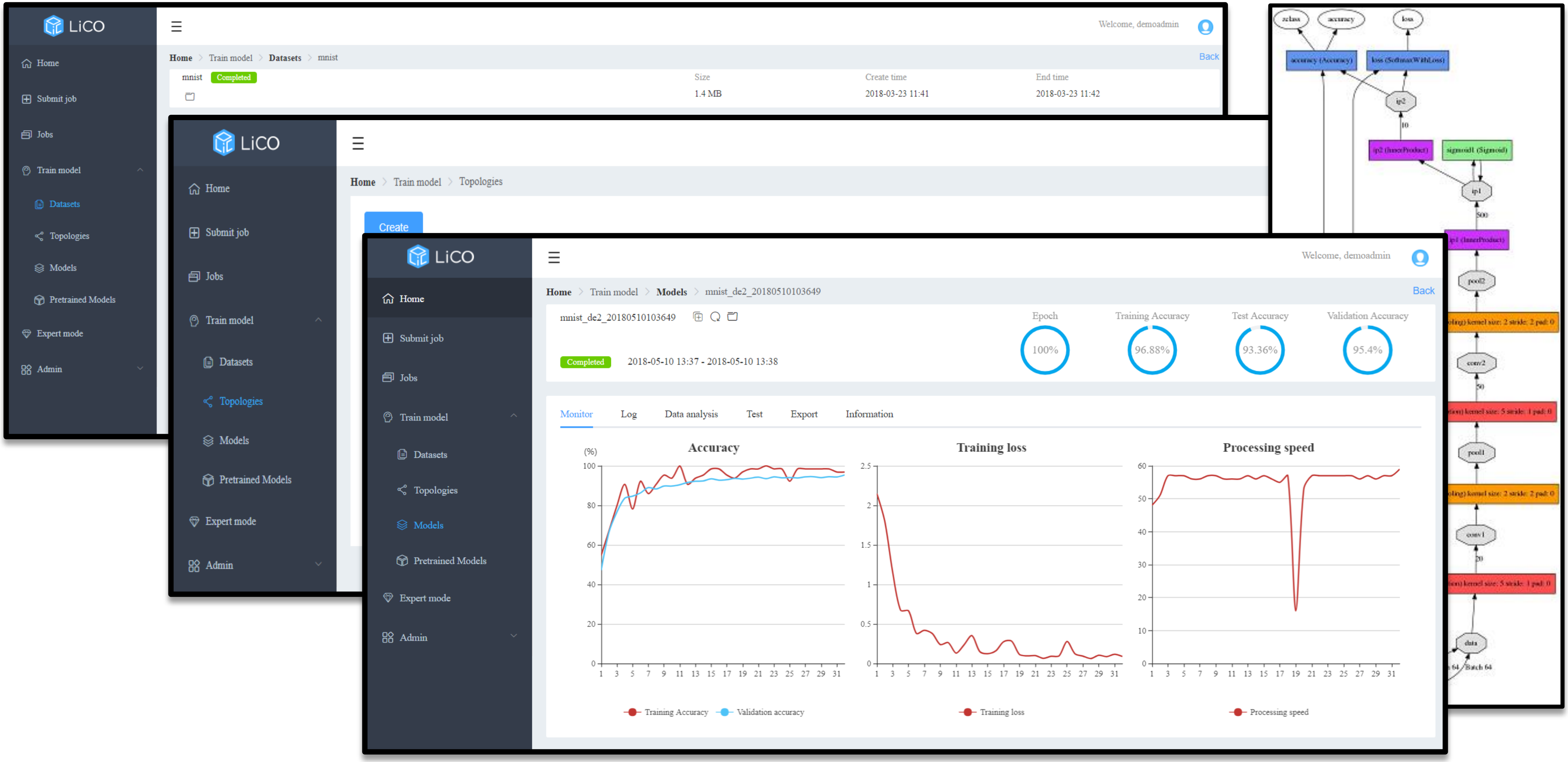
- Supports both **NVIDIA** GPUs and **Intel** processors to suit varying workloads

Optimally uses hardware resources for better TCO

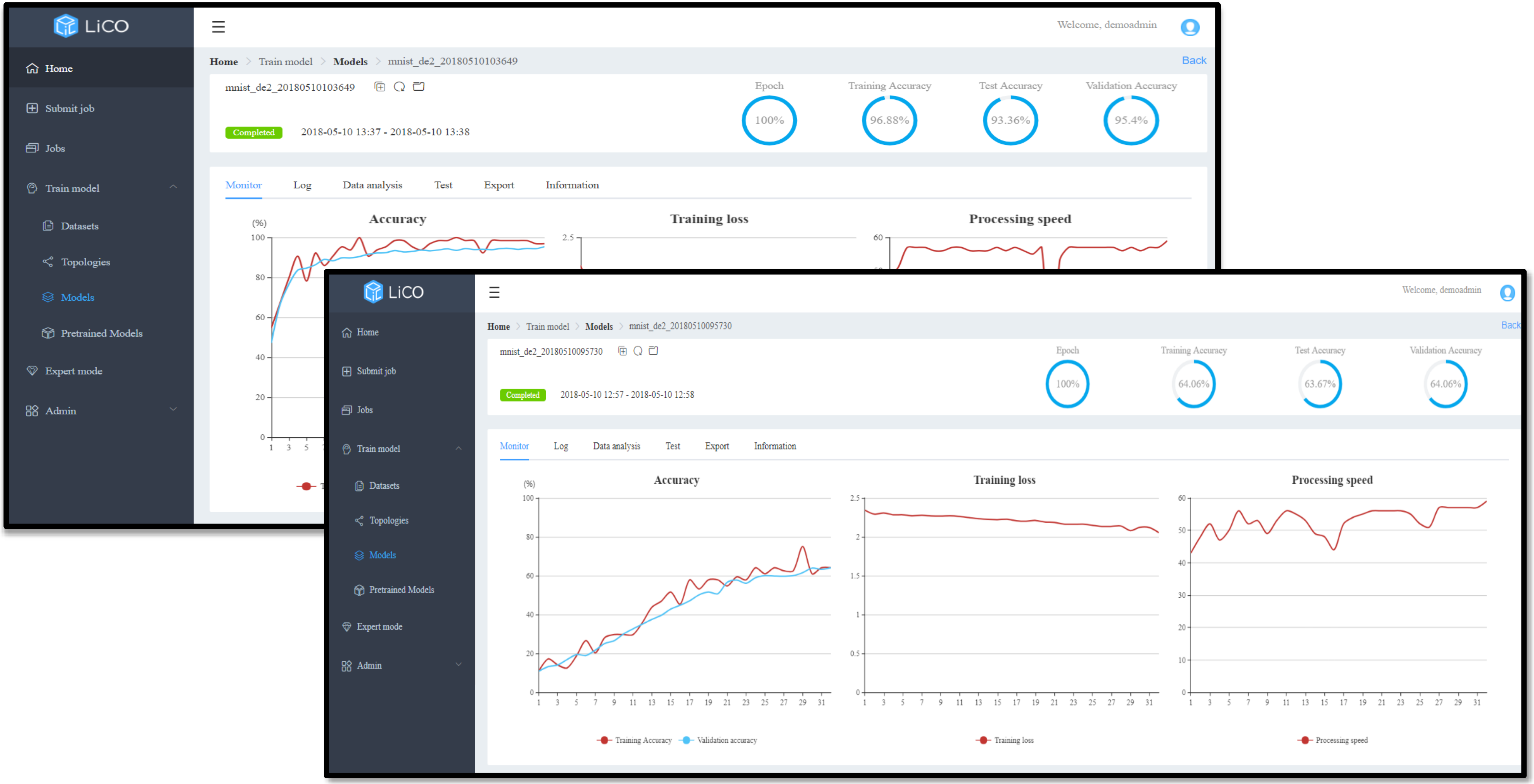
- Improve scaling efficiency for distributed model training
- Manages cluster resources among multiple jobs



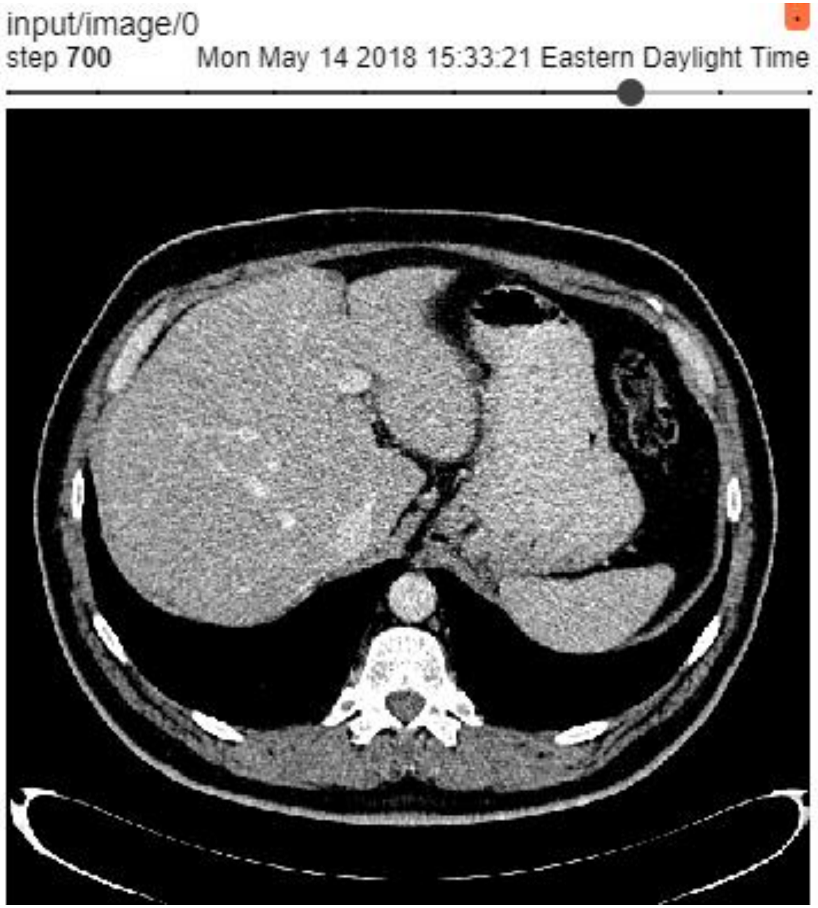
Track Data Sets, Topologies, Models

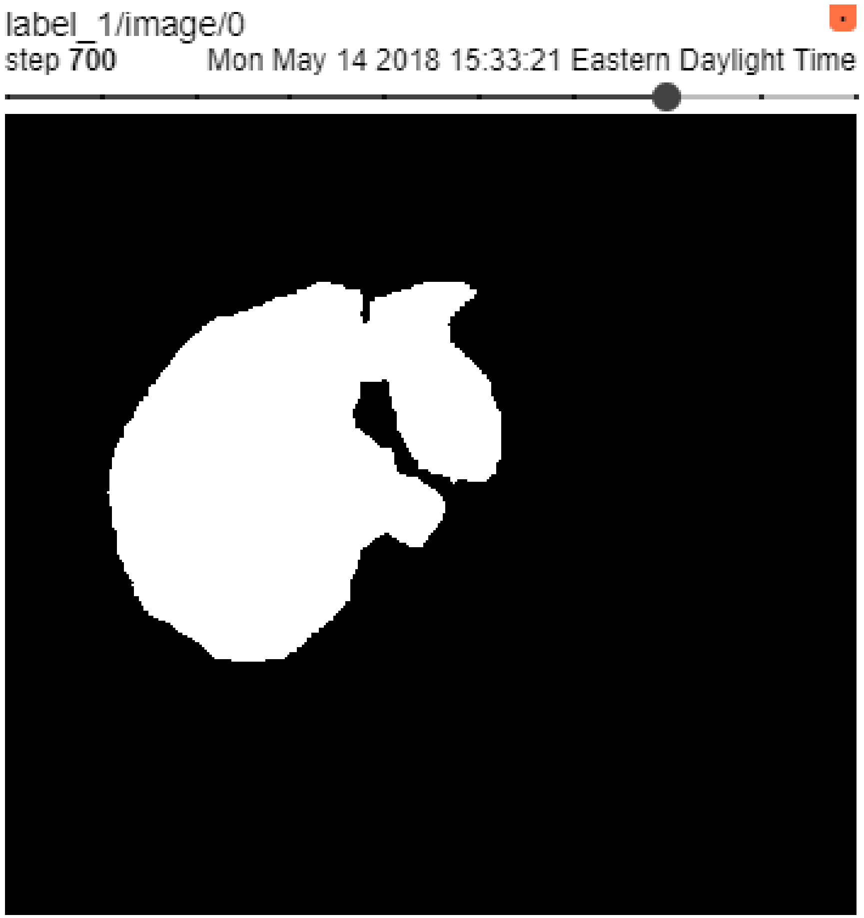


Complete Model History

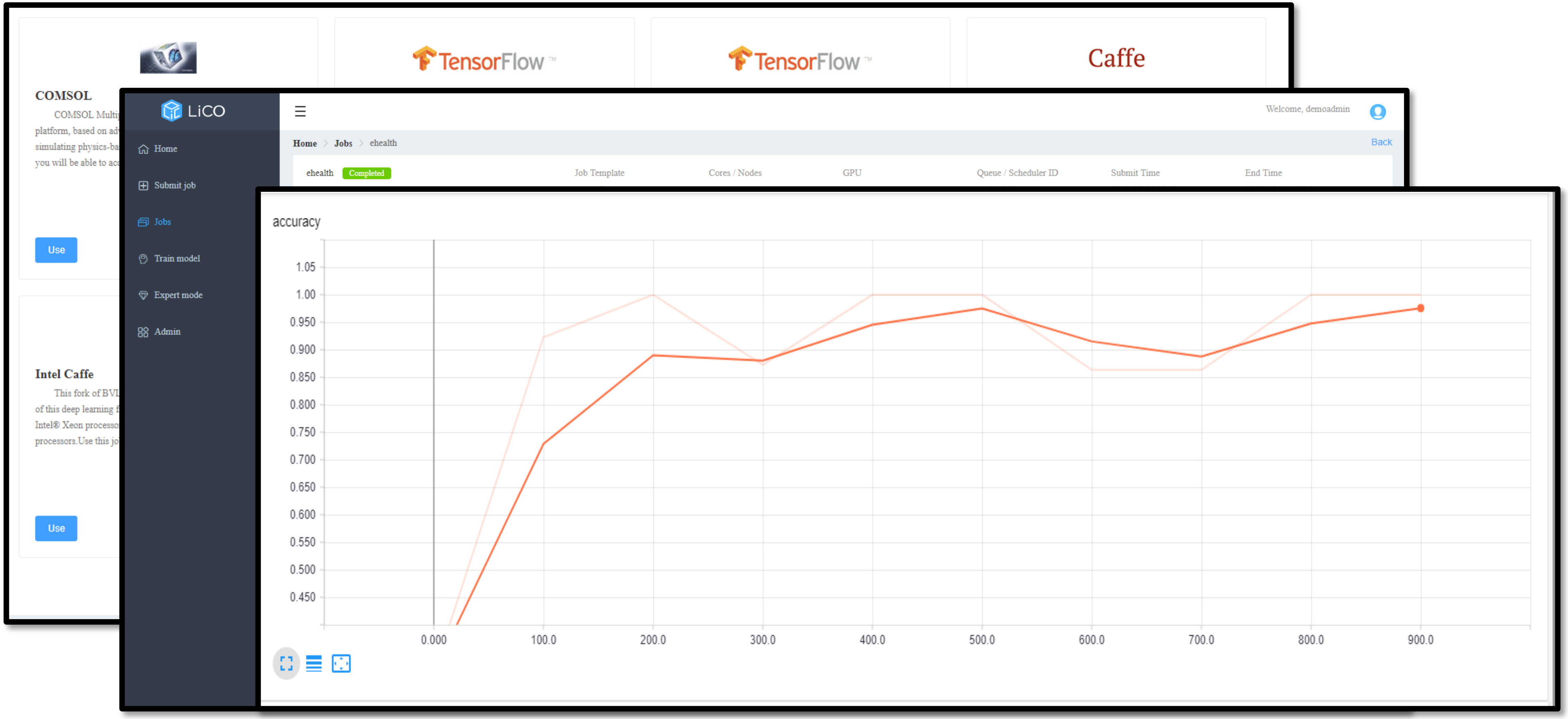


Consistent Visual Inspections

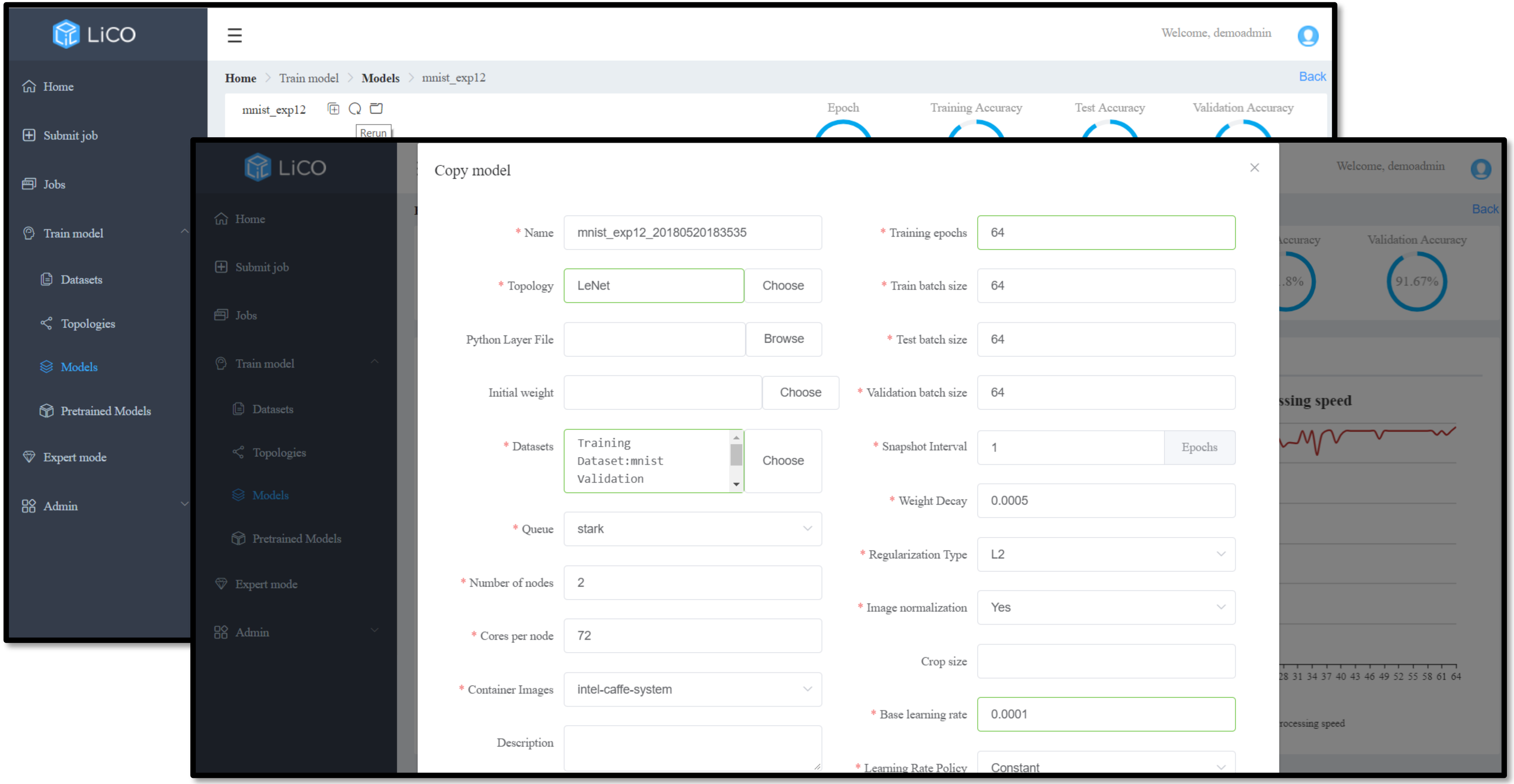




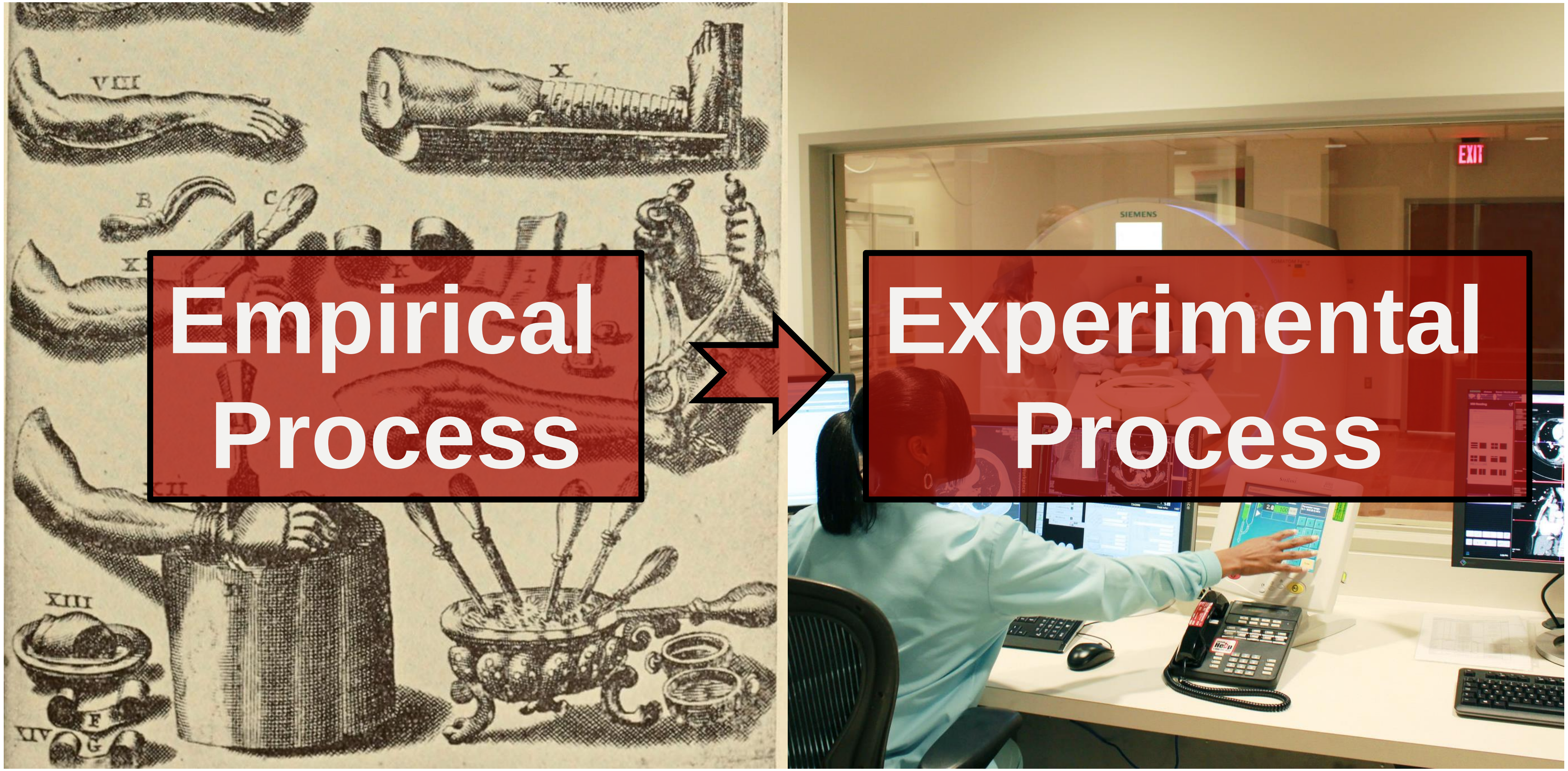
Flexibility Across Frameworks



Others Can Validate & Adjust Model



Lessons From Medicine: ‘Modernizing’ Deep Learning Model Development



Experimental Process for Deep Learning

Design	Execute
Planned	Controlled
Ethical	Reproducible

AI Use Cases that Lenovo is Exploring Today

For better quality control on

Manufacturing Lines

For proactive actions improving

Water & Energy Conservation

To advance particle physics at

CERN's Large Hadron Collider

A more efficient and effective

Radiology Practice

AI: Better quality control for Manufacturing

Breakthrough innovation in Manufacturing with Lenovo AI

Mark III Systems

Lenovo

Current State of Manufacturing:

- Better Quality Control directly related to:
  - More yield at higher speed
  - Lower production costs by faster adjustments to process

Where AI can help:

- Leverage cameras and sensors through-out the production lifecycle
- Better manage quality through product age of customizations

AI Inference and Training:

- Image / Pattern recognition and analysis
- ML training of products at different stages of production

See demo in action at SC17, Booth 1353

AI: For Proactive Actions Improving Drought Management

Lenovo AI helping conserve water and energy for food security

Current State and Challenges :

- Agriculture accounts for 70% of the total global freshwater withdrawals
- As water availability for agriculture becomes scarce and adds cost, it is critical to identify areas that will have impacts on certain crops to manage drought

NC STATE UNIVERSITY

Lenovo

Where AI can help:

- <u>Predict drought prone areas</u>: AI and deep learning can help identifying patterns from large amounts of geospatial data that will most likely face water challenges

AI Inference and Training:

- Deep learning training with satellite images for feature extraction of agriculture land, crops, and water resources to access drought conditions

AI: To Advance Particle Physics in the Large Hadron Collider

Lenovo AI progressing advanced Particle Physics

Current State and Challenges :

- Reconstructing particle trajectories from thousands of sensor measurements in the detector is an important data analytics task
- Traditional computational methods demand high amounts of resources and don’t scale well as the LHC pushes to higher collision frequencies

Where AI can help:

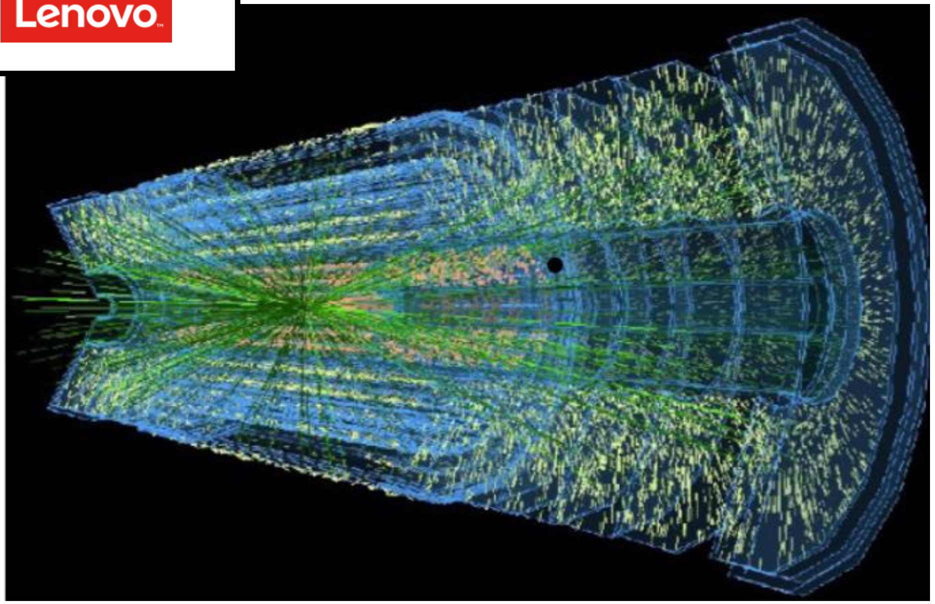
- Pattern recognition: Reconstructing particle trajectories using imaging data from the collider much more efficiently than traditional methods

AI Inference and Training:

- Machine learning methods using binary image data from experiments combined with integral transforms for pattern recognition

UCL

Lenovo



Lenovo

AI: A More Efficient and Effective Radiology Practice

Lenovo Helping advance Cancer Research

Current State of Radiology:

- Central to patient care
- Top 4 area of expense for hospitals
- 93% of US Radiology groups under staffed

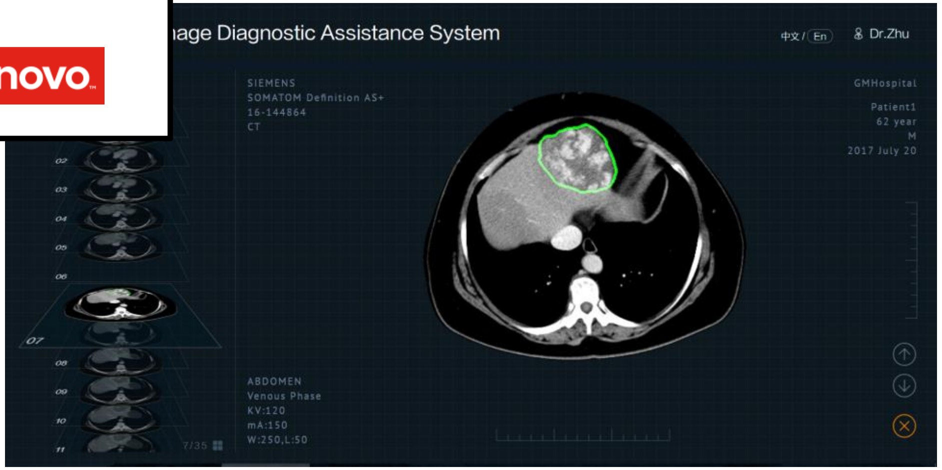
Where AI can help:

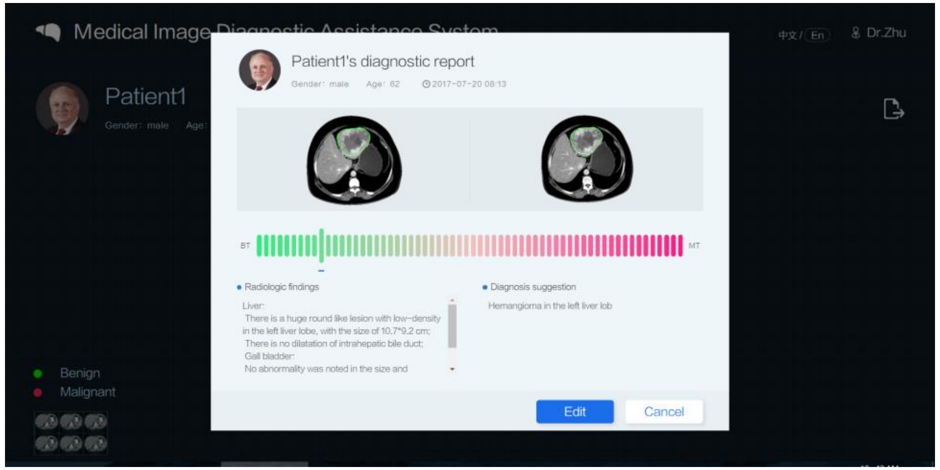
- Better Efficiency: AI helps prioritize highest risk patients, optimizes radiologist time and lead to better patient outcome
- Improve Effectiveness: AI can offer a second screening to physician finding to reduce avoidable errors

AI Inference and Training:

- Image / Pattern recognition and analysis
- ML training with millions of patient images and outcomes

See demo in action at SC17, Booth 1353



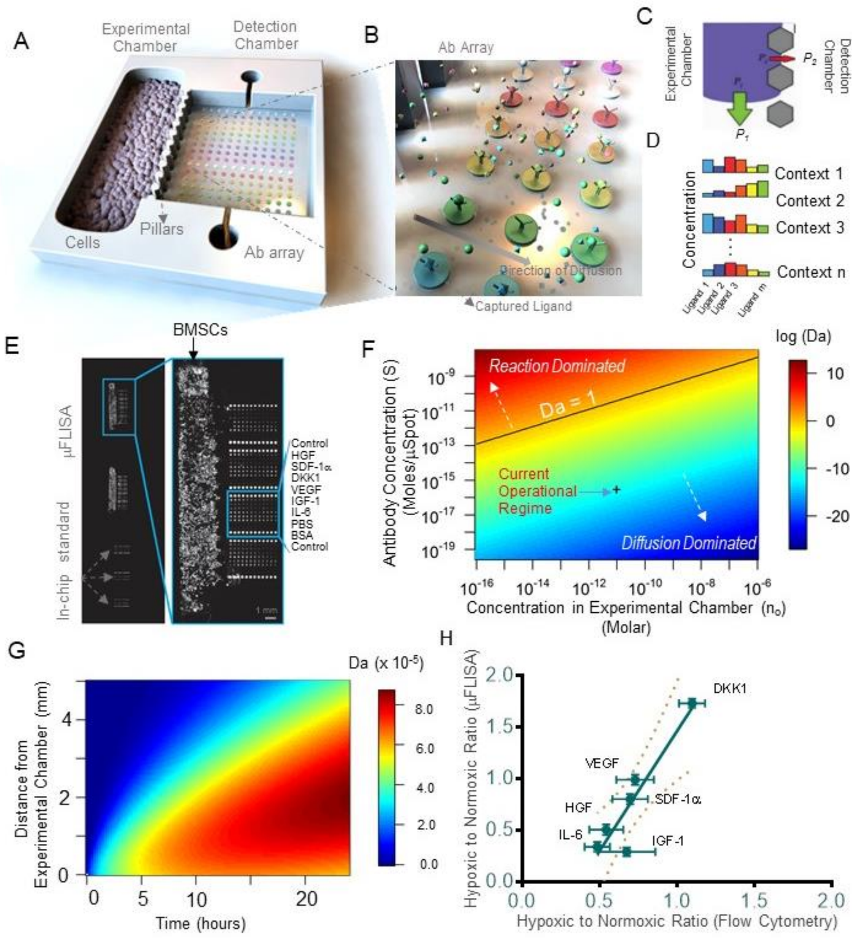


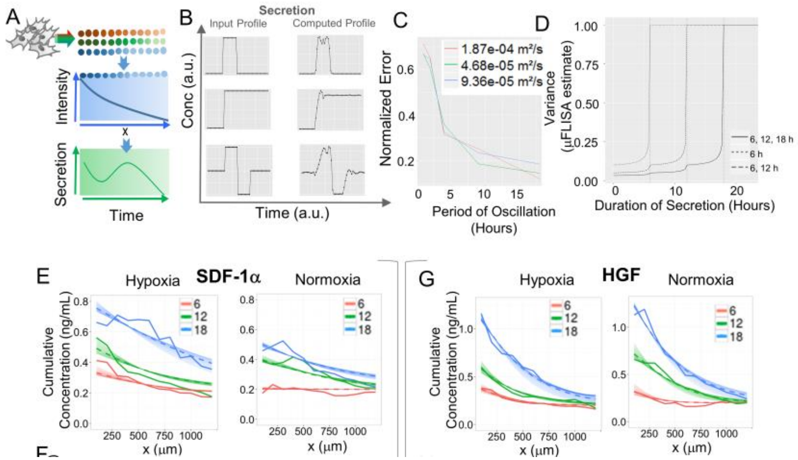
thanks.

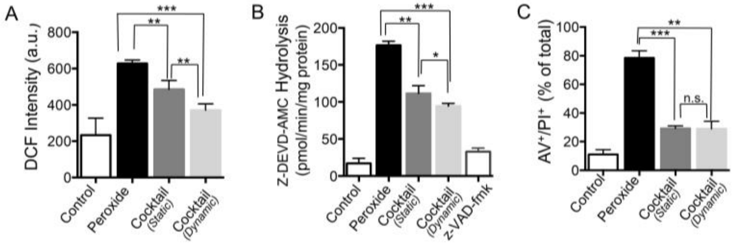
Different is better

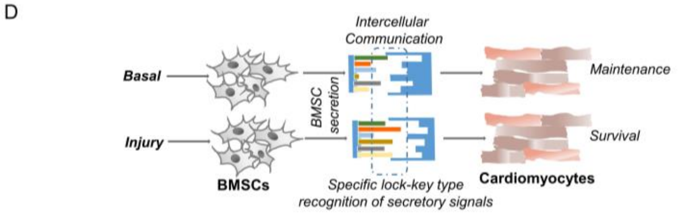
Appendix

Dynamic Analysis of Cell Secretions for Cardiac Repair



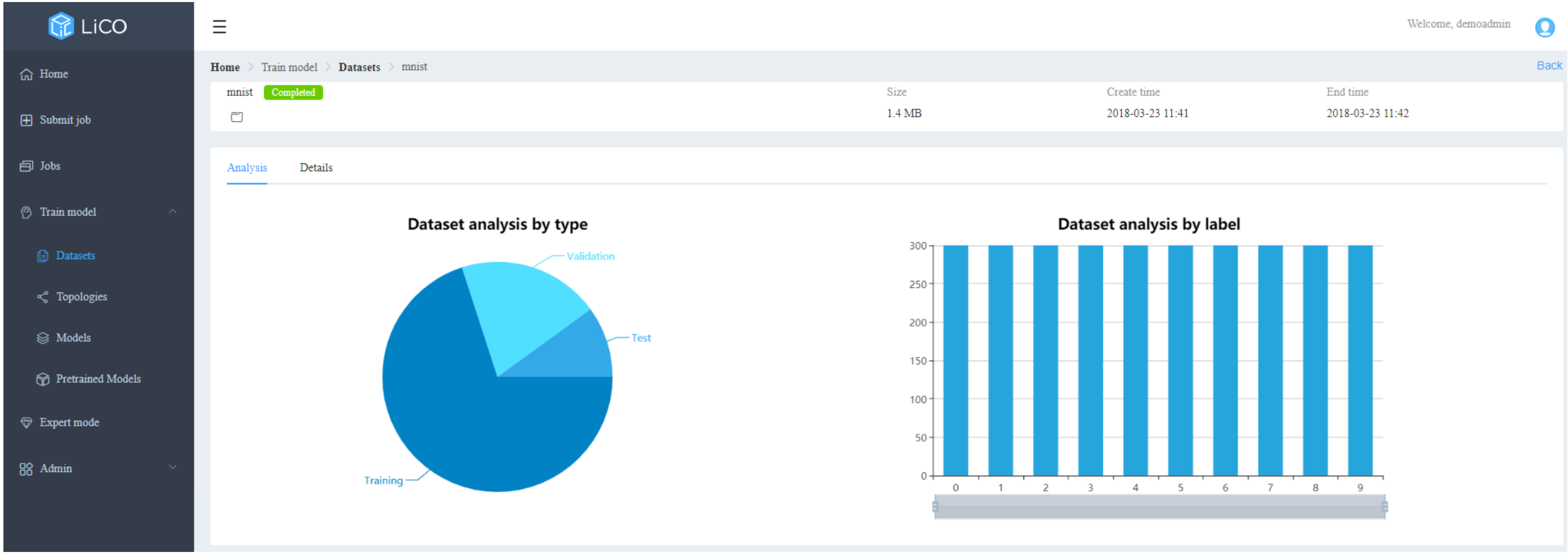






Ellison, Gupta, Suhail, et al. microFLISA: A New Experimental and Computational Platform For Analysis of Dynamic Secretomes Uncovers a Cardioprotective Secretion Signature. Under Review with Nature BME.

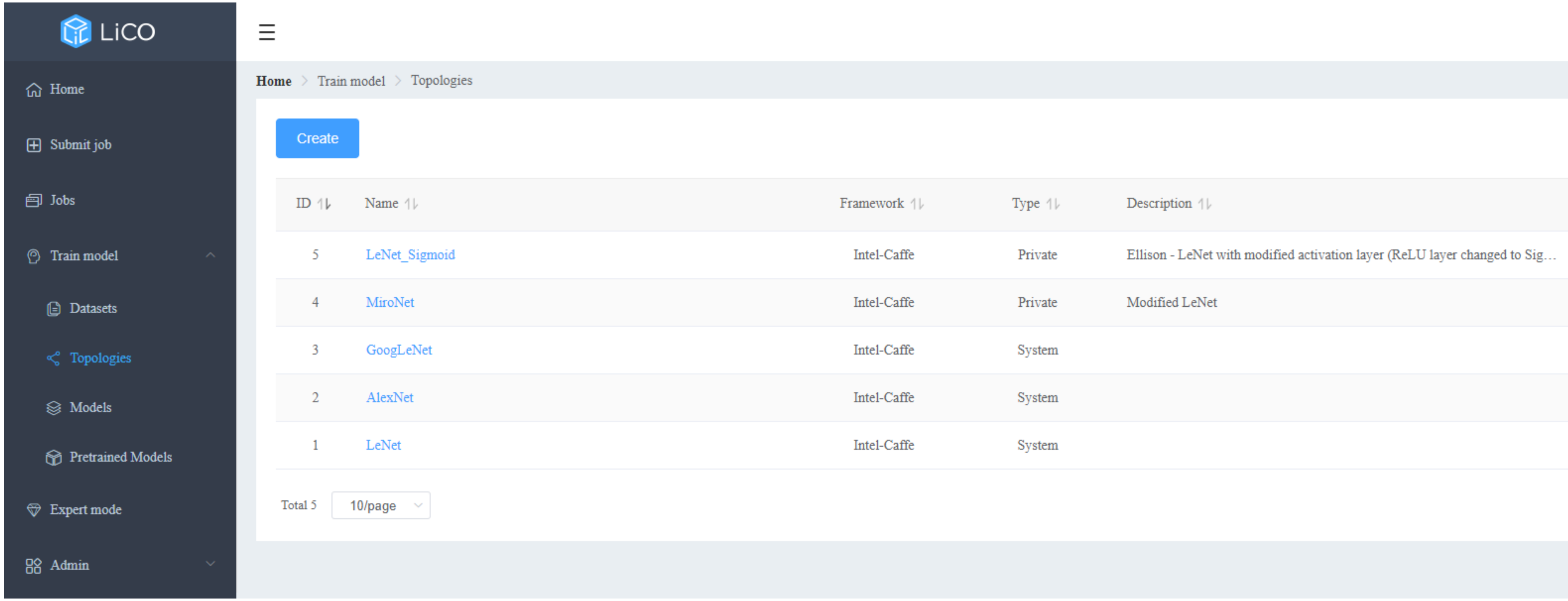
Controlled: Track Data Sets

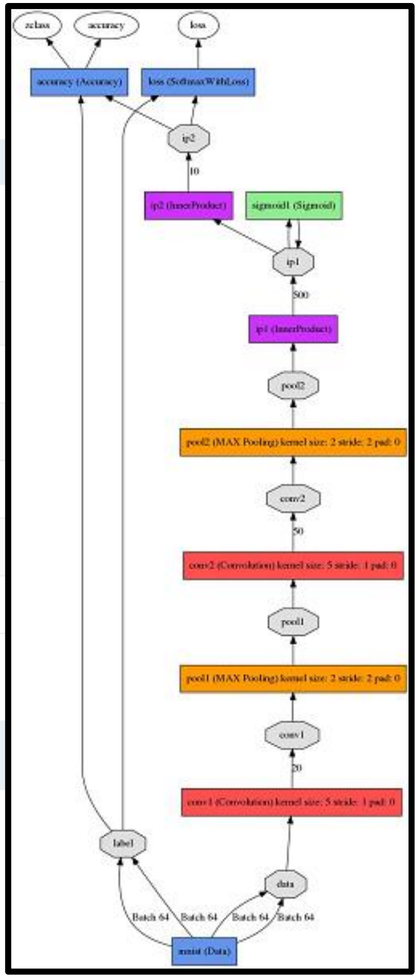


Principle: Keep Snap Shots of Your Data

- Minimally ensuring expected class balance and distributions
- Extremely large or streaming data poses more complex problems

Controlled: Track Topologies

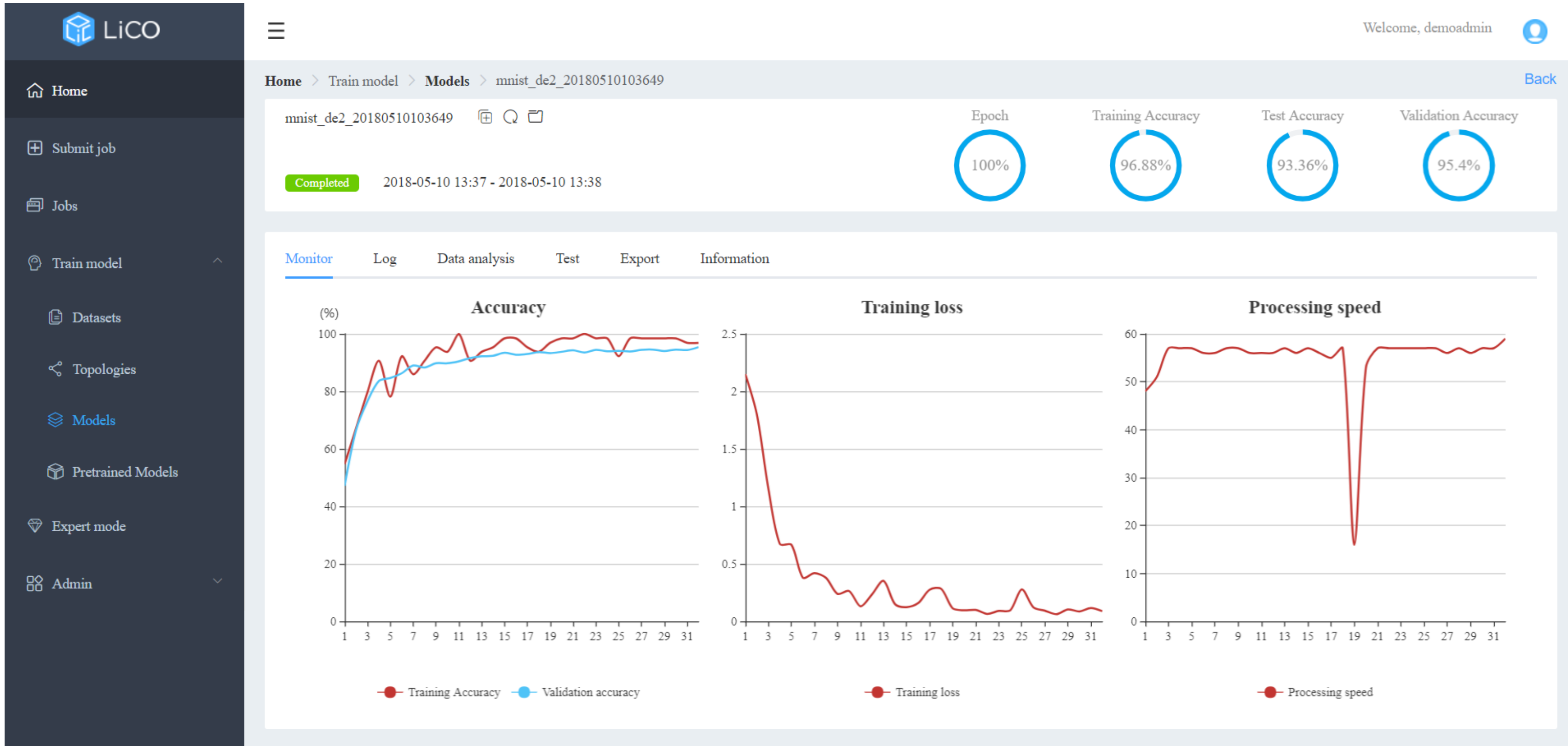




Principle: Record and Check Neural Network Topologies

- Changes in activation functions can have unexpected results
- Quick visualization for sanity checks
- Enable faster transfer learning

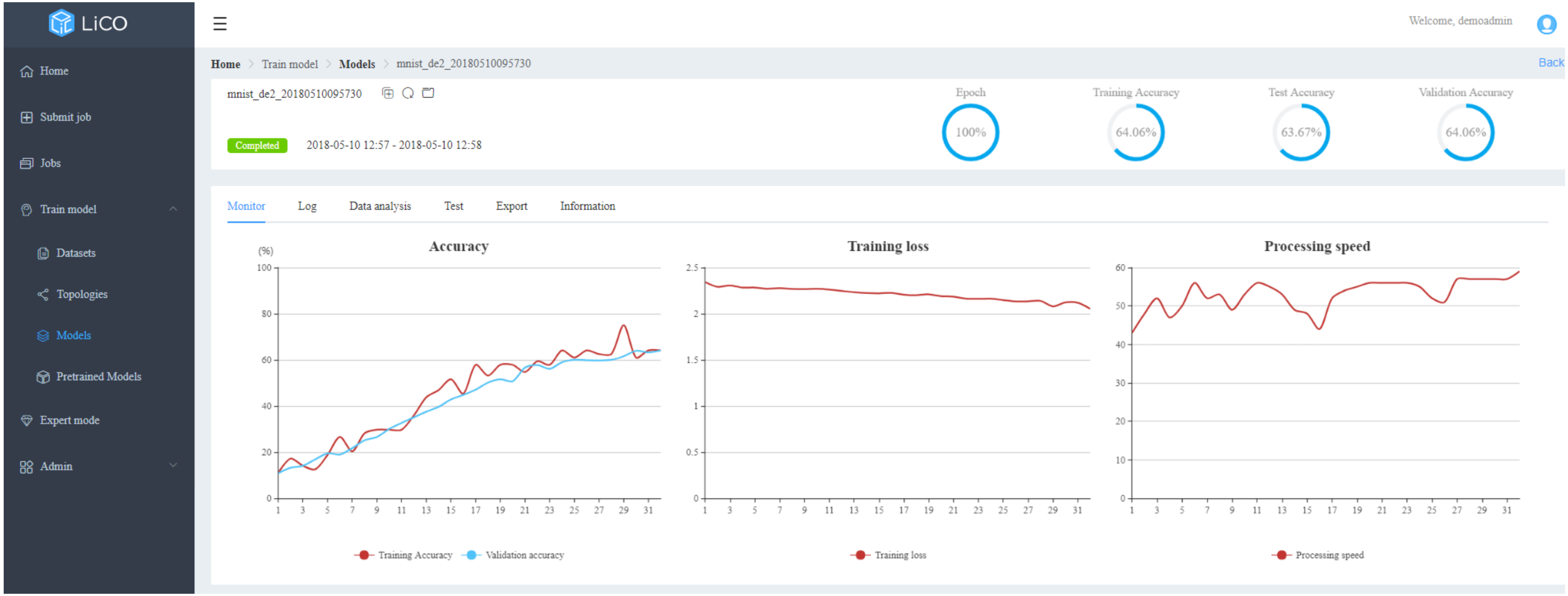
Controlled: Consistent Job Runs/Models



Principle: Robustly Record Successful Models

- Establish metrics and consistently measure the across models
- Look for and record anomalies
- Simple to systematically do hyper parameter tuning

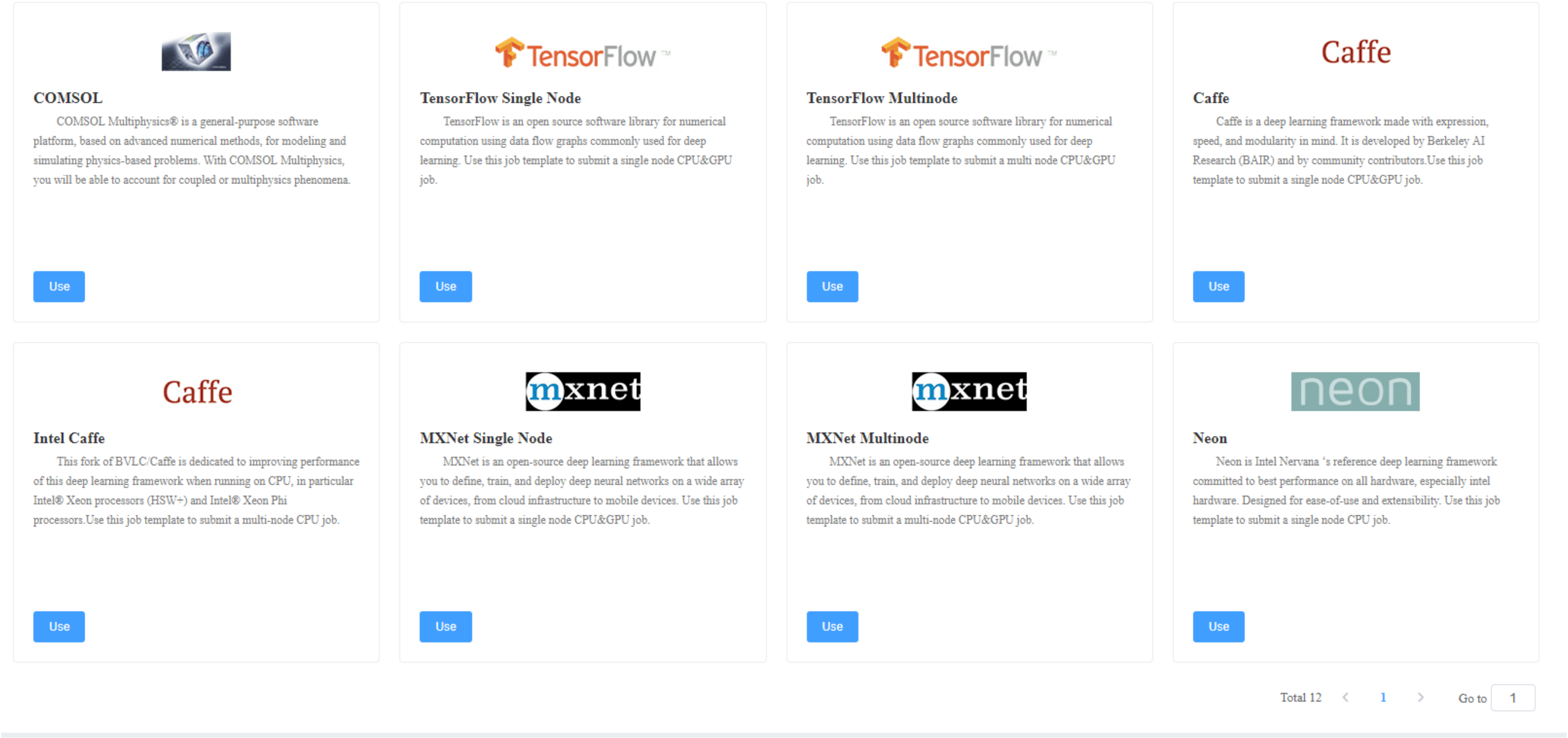
Controlled: Remember Poor Models



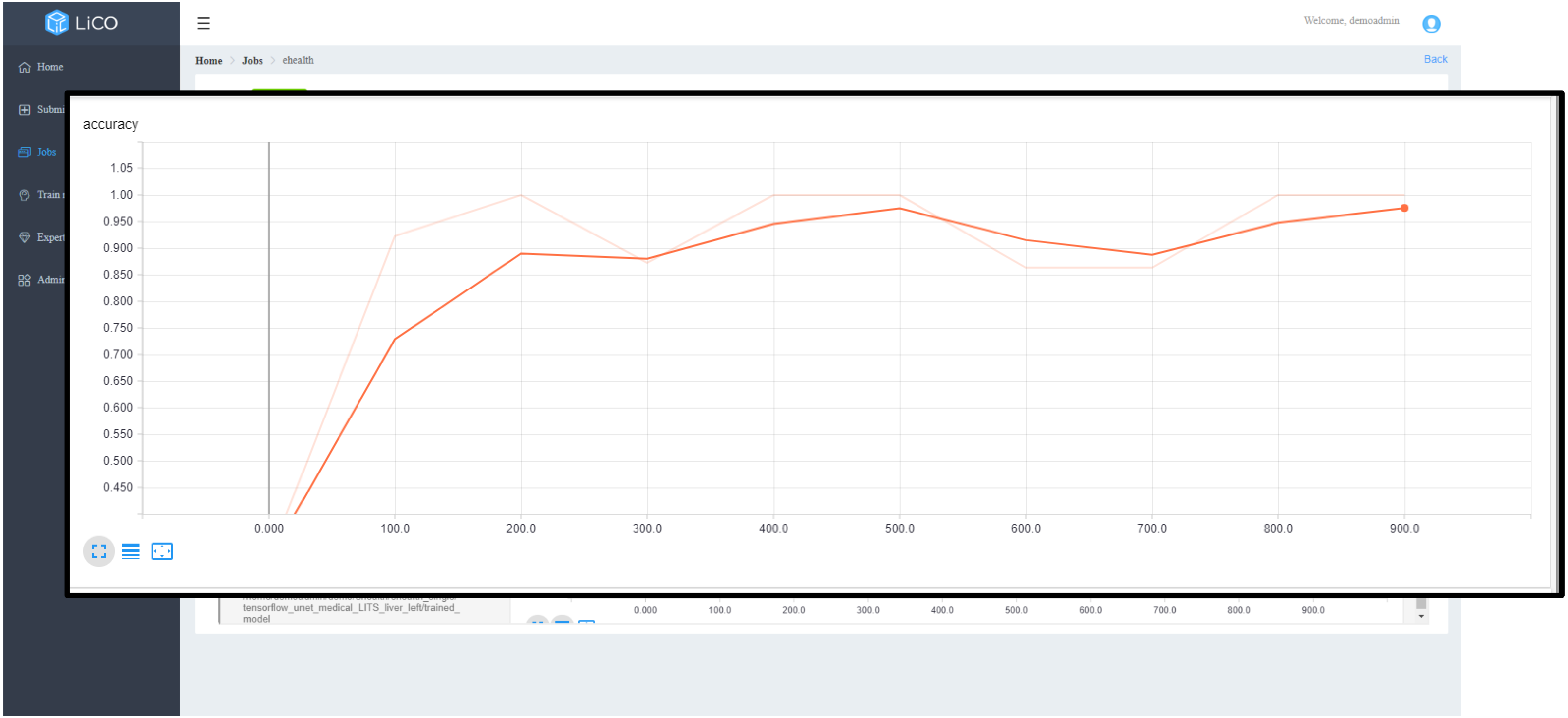
Principle: Record and Learn from Poor Models

- Mitigates survival bias
- Data needed for later robustness analysis

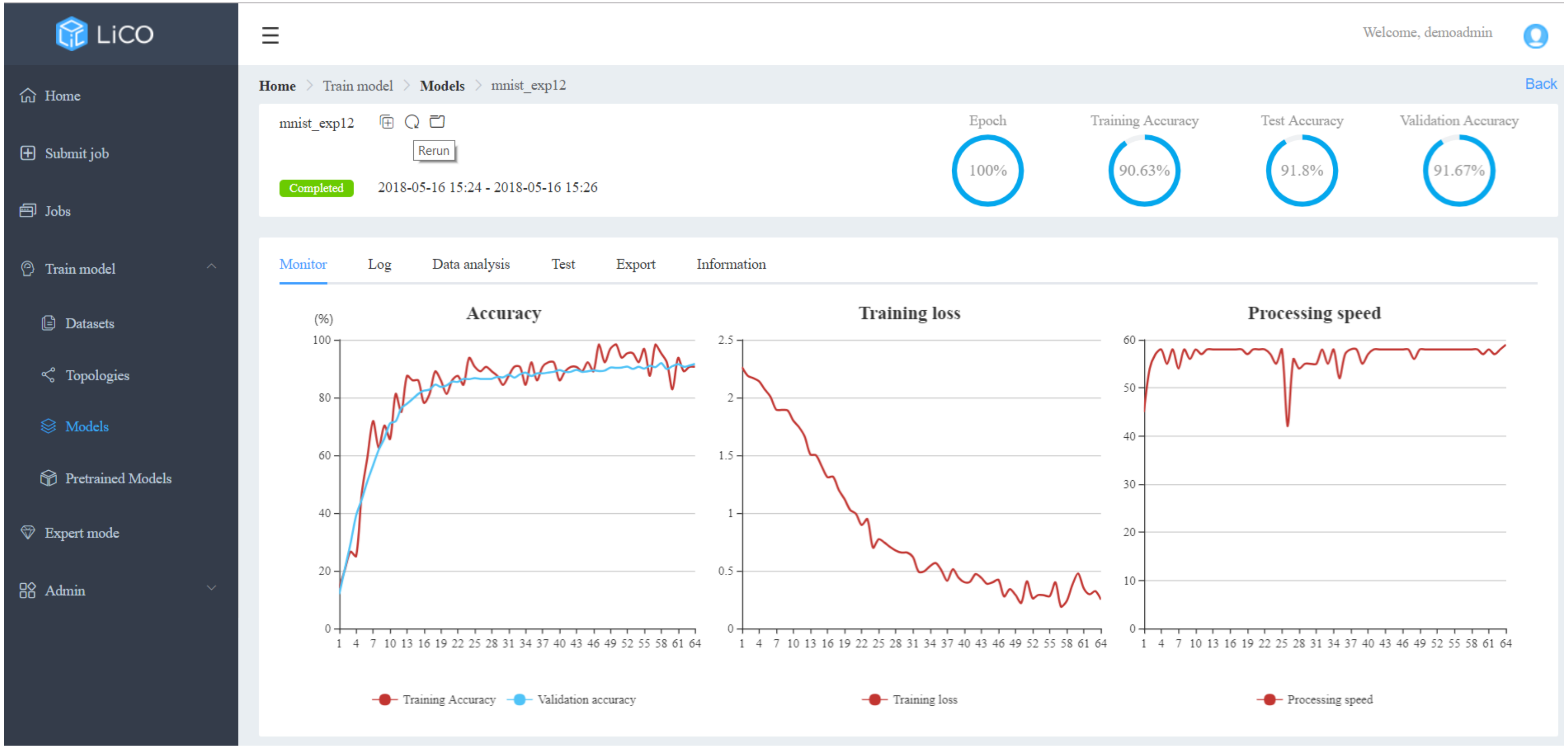
Controlled: Flexibility Across Frameworks



Controlled: Same Metrics Across Frameworks

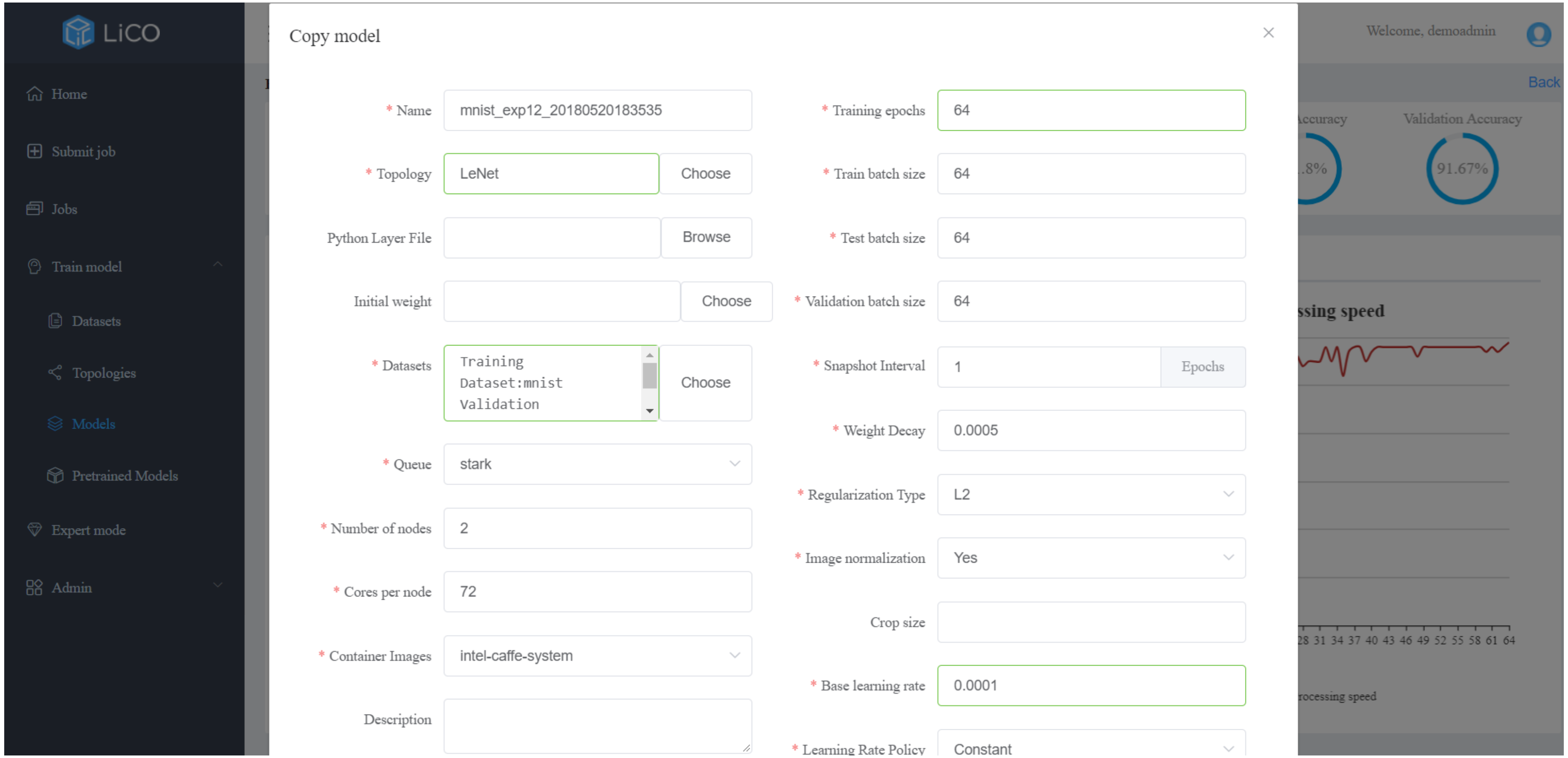


Reproducible: Others Can Validate Your Model



Principle: If Others Cannot Easily Reproduce Your Model, No One Will

Reproducible: Others Can Make Adjustments



Principle: Others Should be Able to Test the Boundaries/Robustness of Your Model