MATERIALS INFORMATICS FOR COATINGS FORMULATIONS

Applied Machine Learning Strategies for Rapid Reformulation

Will Erwin, PhD
Sr. Data Scientist
Citrine Informatics

AGENDA

- What is Materials Informatics?
- Materials Informatics in Paints and Coatings
  - Raw Material Replacement Case Study
- Summary + Q&A

WHAT IS MATERIALS INFORMATICS?

CITRINE INFORMATICS

- We built the first chemistry & materials aware AI platform
- We enable teams across North America, EU and Japan to scale materials informatics at their organizations

EASTMAN

synthomer

SOLVAY

lyondellbasell

Panasonic

SHOWA DENKO

MORROW

LANXESS Energizing Chemistry

AI 100

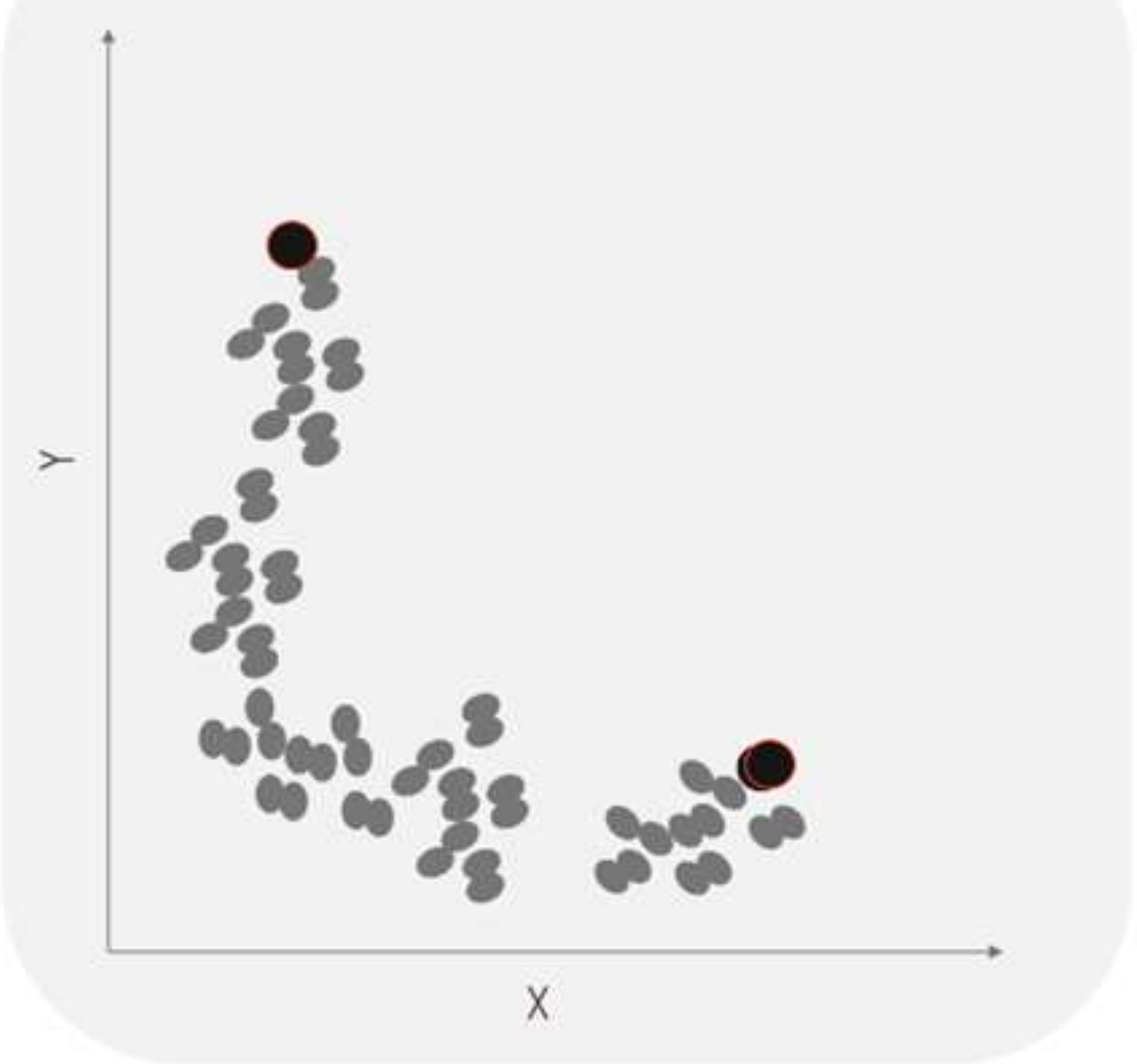
WORLD ECONOMIC FORUM

AI BREAKTHROUGH AWARDS

MATERIALS INFORMATICS ENABLES DATA DRIVEN PRODUCT DEVELOPMENT

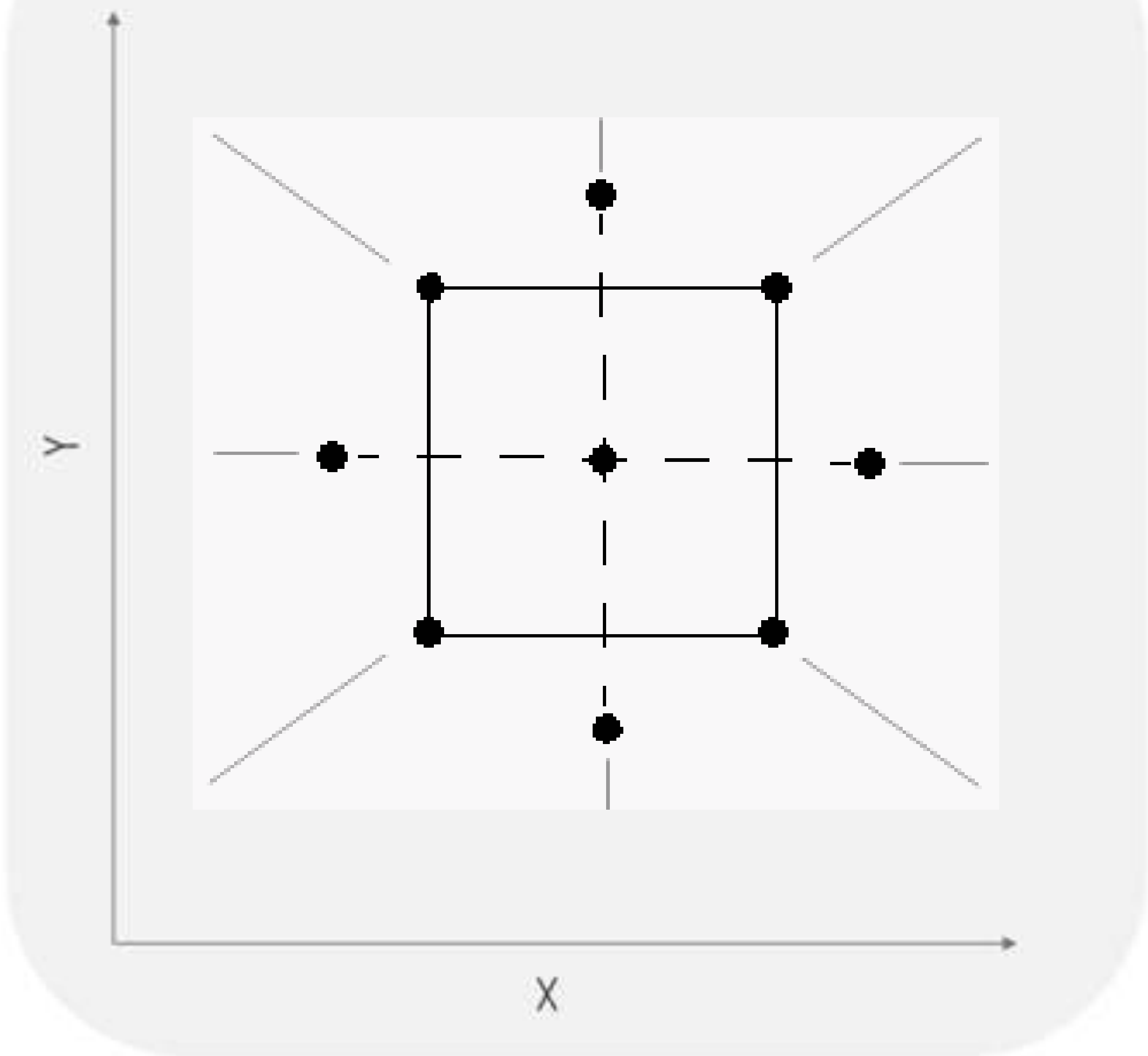
Reduce required experiments, improve efficacy, breach performance frontiers

Intuition Based



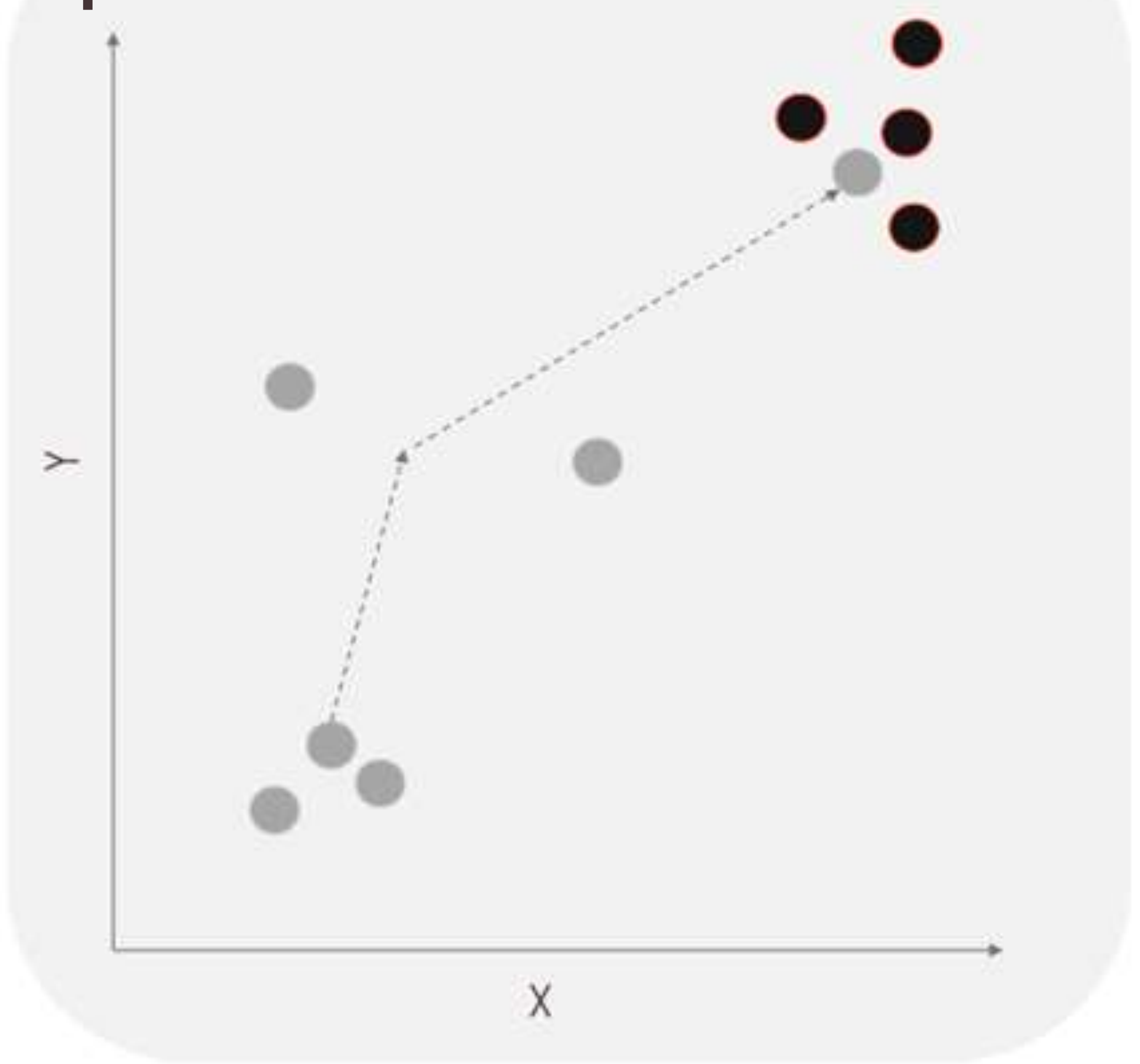
- Leverages experience

Design of Experiments



- Leverages data, but only what you give it
- Often inefficient in high dimensional spaces
- Emphasizes understanding feature effects

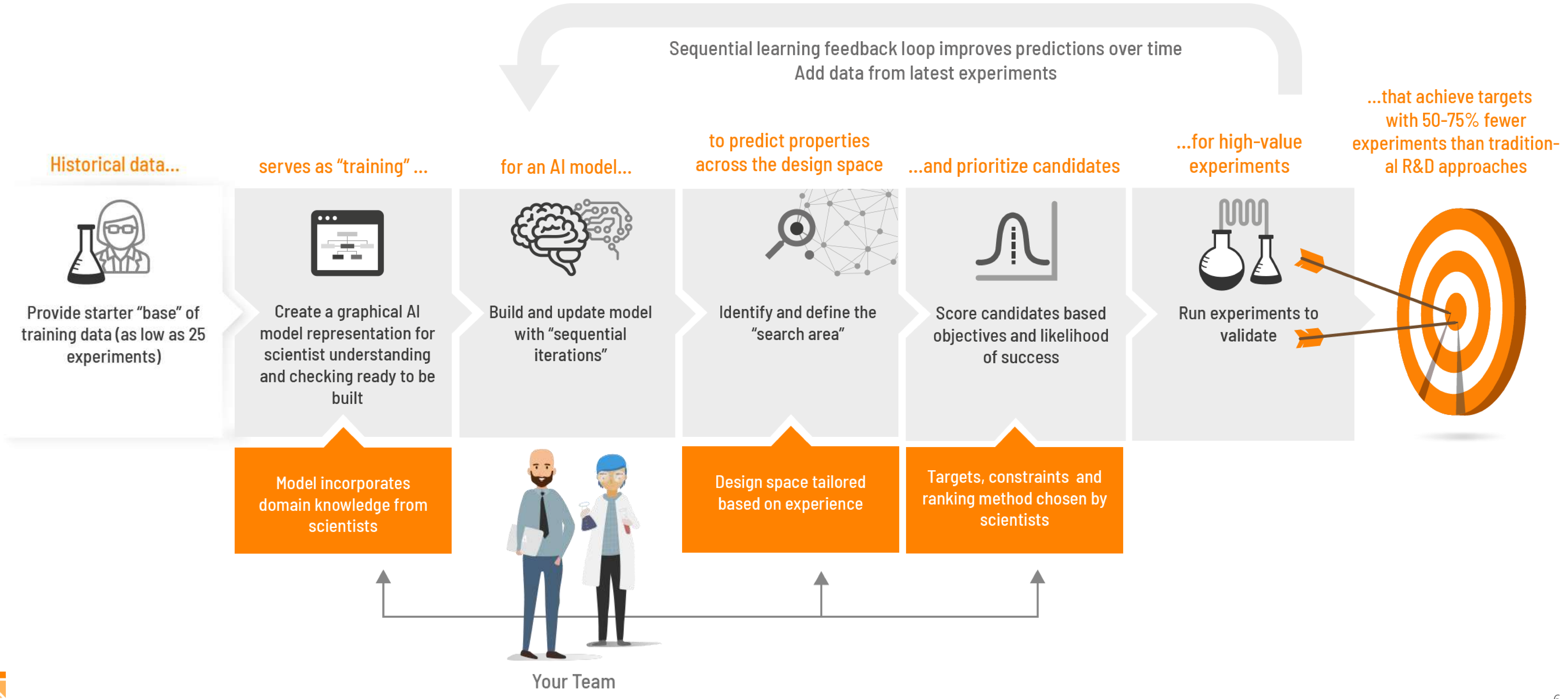
AI Guided Experimentation



- Leverages data *and* expertise
- High dimensional capability
- Emphasizes finding the answer

CITRINE AI-GUIDED SEQUENTIAL LEARNING

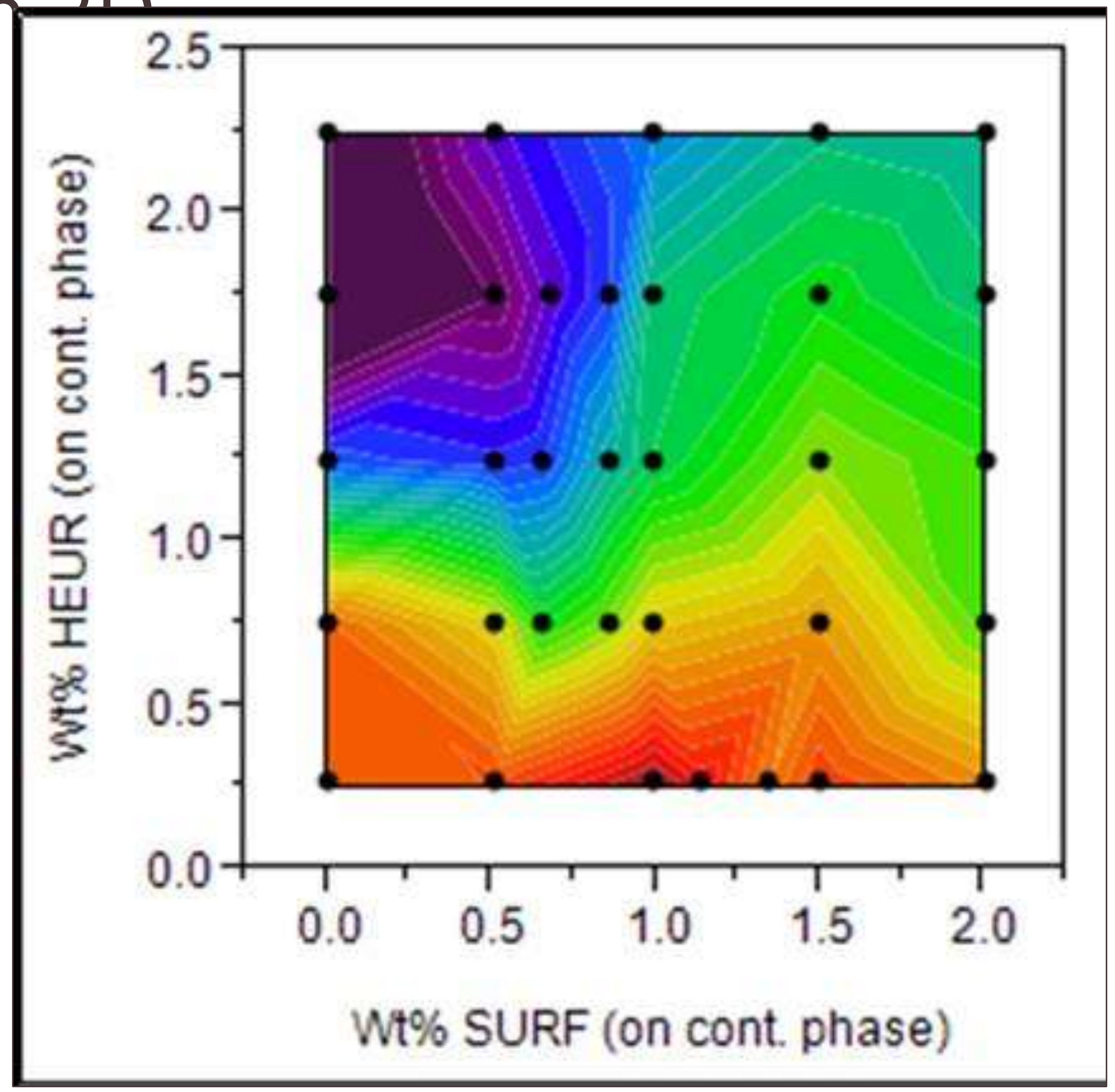
What is it?



BUT *HOW* DOES SEQUENTIAL LEARNING WORK

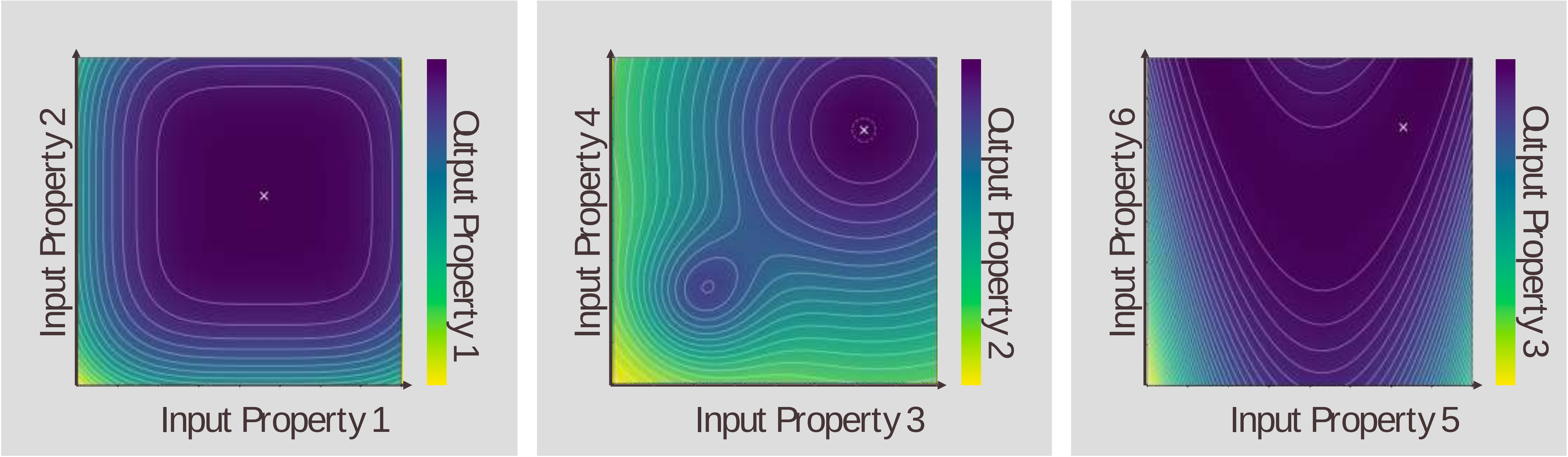
Design experiments to co-optimize multiple properties

Traditional DoE rigorously maps in 2D



JCT Coatings Tech · Feb 1, 2015

SL simultaneously optimizes over n-dimensions



Input Properties:
Formulation
Surfactant, Pigment, Solvent, Binder, Additives

Process Parameters
Grinding, Mixing, Application

Output Properties:
Rheology
Solution Stability
Applied Gloss

SEQUENTIAL LEARNING APPLIED

Sequential Learning enables efficient exploration over high dimensional spaces

Sequential Learning (SL) relies on machine learning and subsequent uncertainty estimates select the optimal experiments to conduct in pursuit of a specific set of goals.

SL 0: Pre-sequential learning

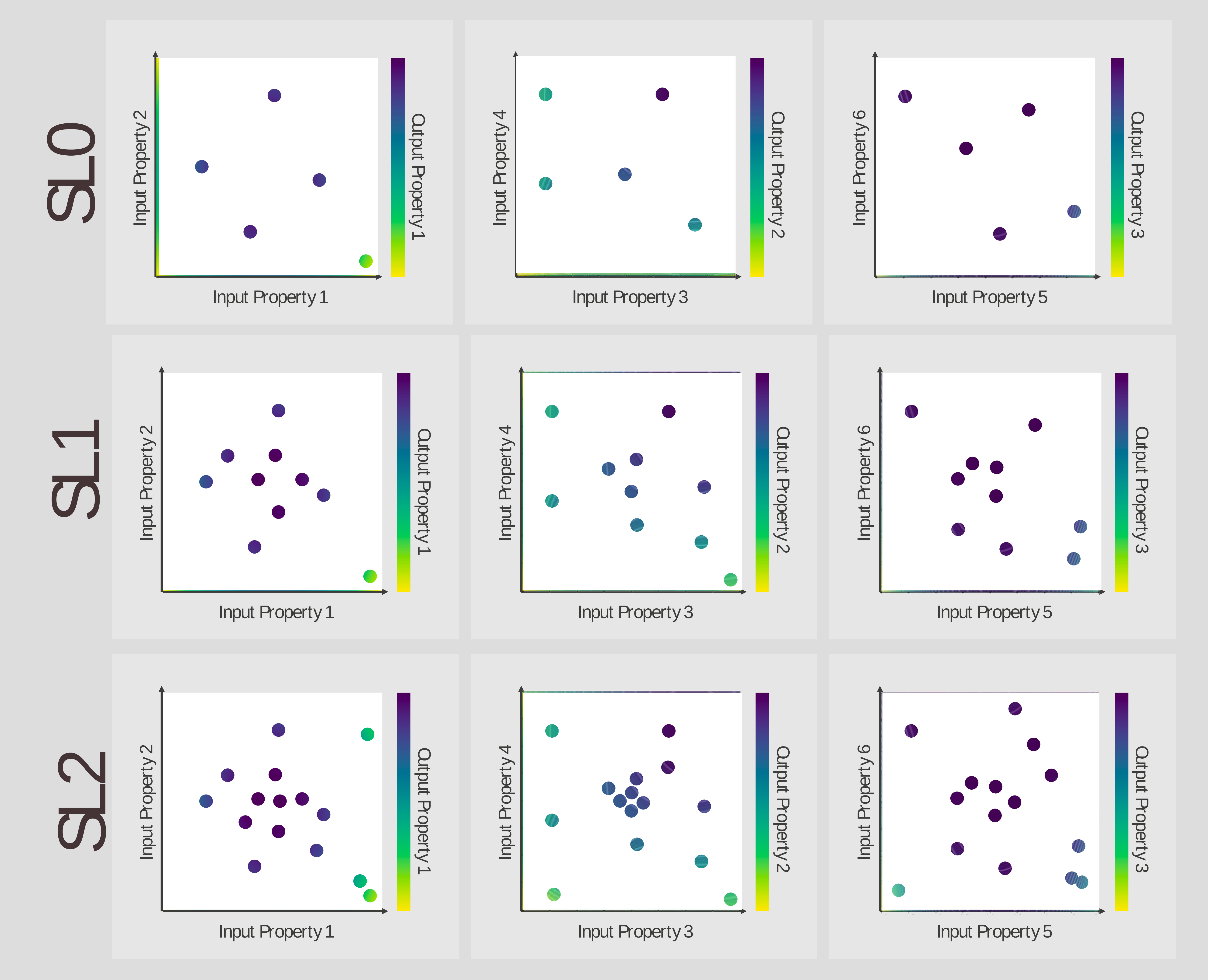
Existing data provides some insight into the system, with discrete samples according to historical experimental data.

SL 1: Sequential learning round one

High value experiments are selected, performed, and laboratory results are ingested back into the system.

SL 2: Sequential learning round two

With new information, a new set of experiments are selected, with increasing likelihood of meeting performance targets.



MATERIALS INFORMATICS IN PAINTS AND COATINGS

CASE STUDY: RAW MATERIAL REPLACEMENT

DEMO: REFORMULATE PAINT TO REMOVE APEO SURFACTANTS

TECHNICAL CHALLENGE

Reformulate an existing paint to remove critical materials (APEO based surfactant).

DEMANDING DESIGN TARGETS

Design a solution stable formulation that is semi-gloss on application with novel surfactant.

TRAINING ON PAST EXPERIMENTS

Leverage previous experiments to predict valid formulations without using APEO containing surfactants.

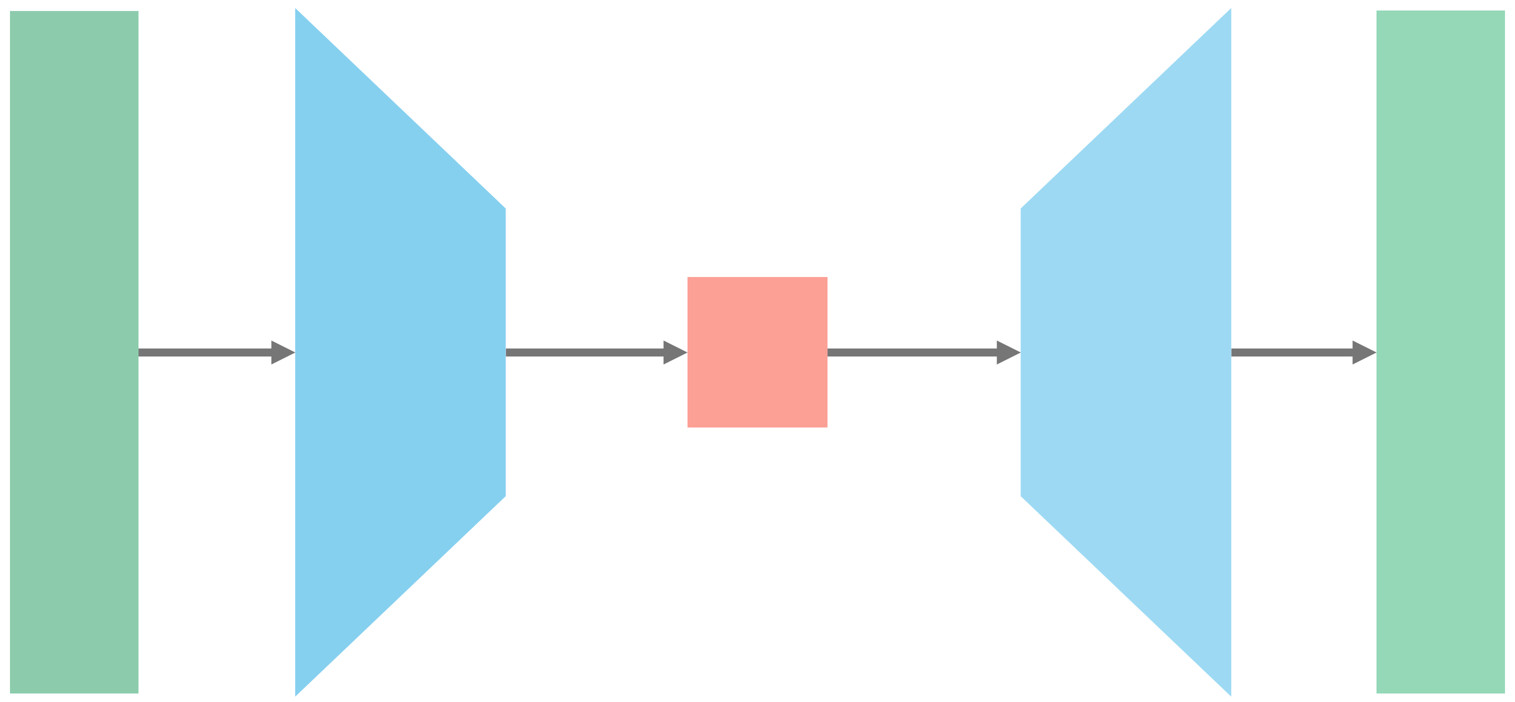
LEVERAGE & ENCODE DOMAIN KNOWLEDGE

Capture domain knowledge and leverage it through the sequential learning process

PREPARING DATA FOR SEQUENTIAL LEARNING

Domain aware data capture enables material specific modeling and design

NOT BIG DATA



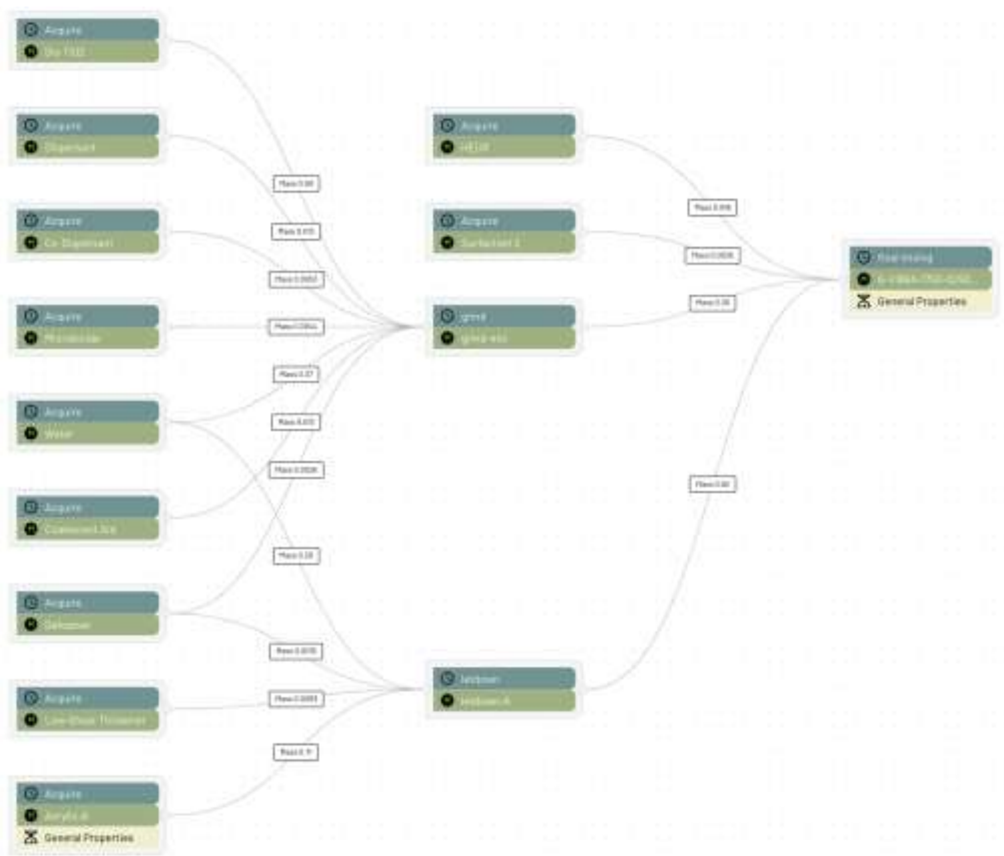
Large datasets enable machine learning *without* domain knowledge, we frequently have limited relevant data in R & D.

DOMAIN AWARE DATA

When few examples are available, it's critical to maximize the value of those data by capturing the context and associated domain knowledge.

- Which properties are impactful?
- Under which conditions was a product made?

A *material history* represents the entire history and process associated with producing a coating product.

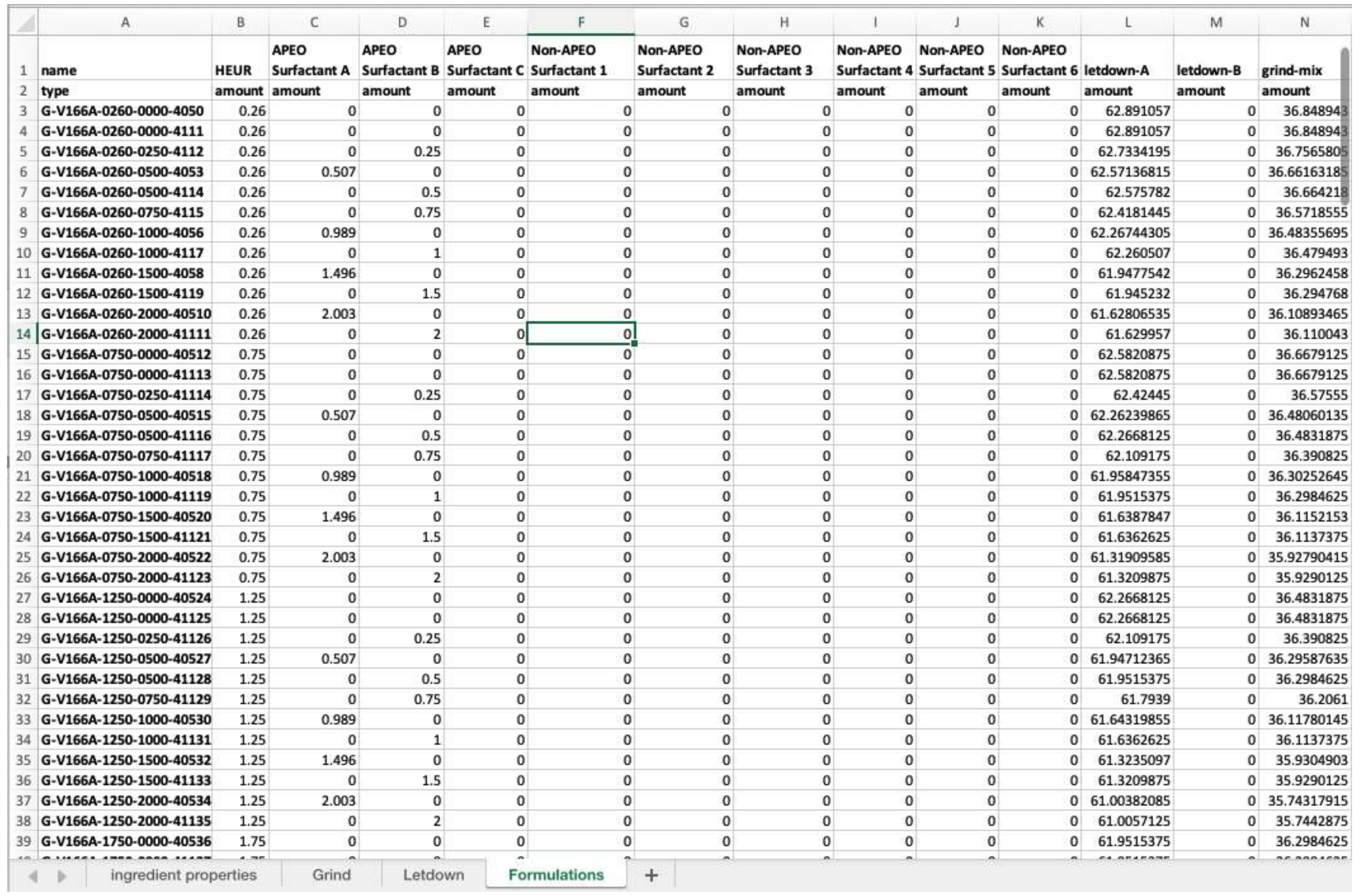


PREPARING DATA FOR SEQUENTIAL LEARNING

Leveraging a coatings aware data model enables analysis and machine learning

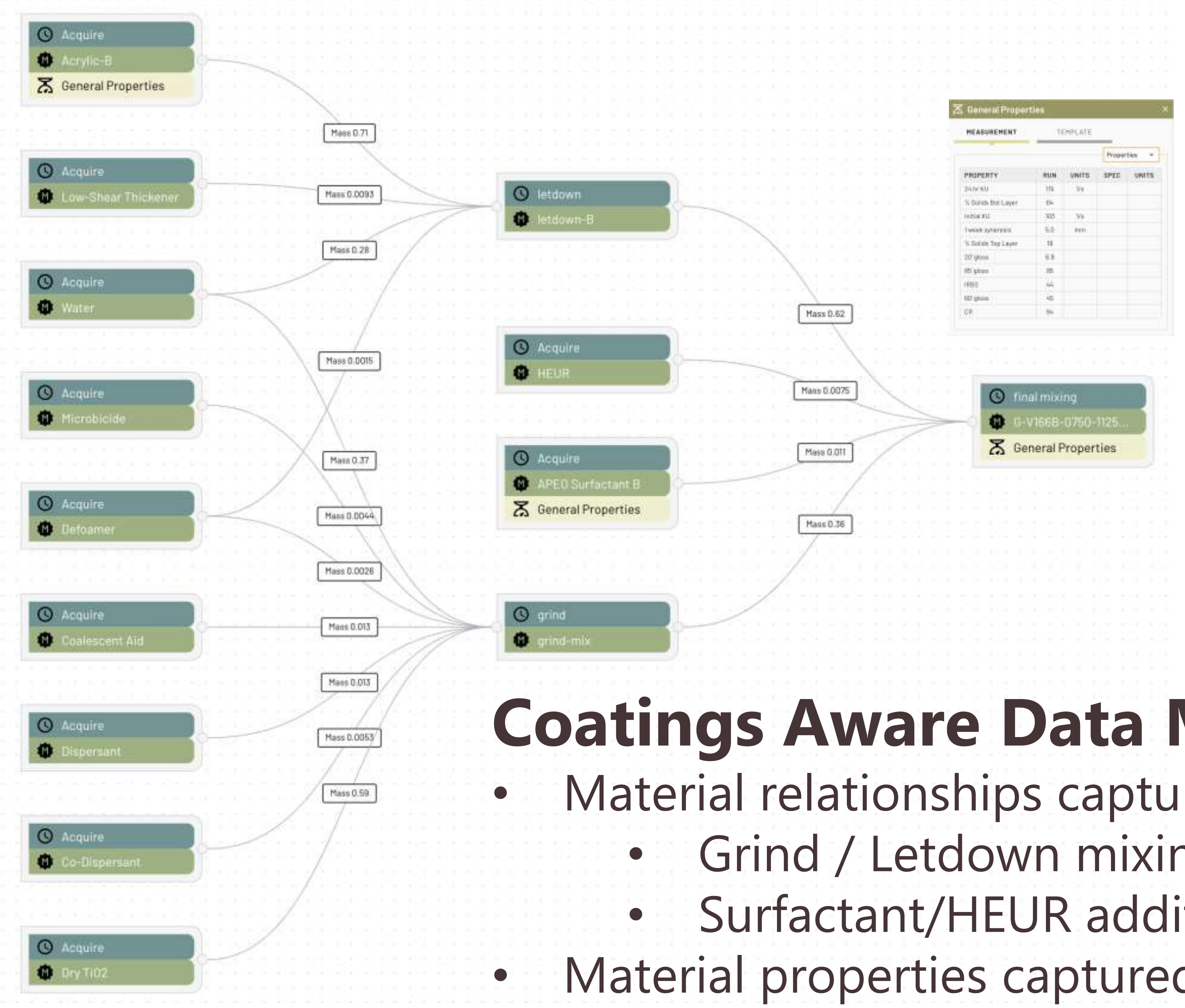
Structure data into tabular format

- Capture material properties that matter
  - inputs / outputs / raw material properties
- Represent all steps/processes



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	name	HEUR	APEO Surfactant A	APEO Surfactant B	APEO Surfactant C	Non-APEO Surfactant 1	Non-APEO Surfactant 2	Non-APEO Surfactant 3	Non-APEO Surfactant 4	Non-APEO Surfactant 5	Non-APEO Surfactant 6	letdown-A	letdown-B	grind-mix
2	type	amount	amount	amount	amount	amount	amount	amount	amount	amount	amount	amount	amount	amount
3	G-V166A-0260-0000-4050	0.26	0	0	0	0	0	0	0	0	0	62.891057	0	36.848943
4	G-V166A-0260-0000-4111	0.26	0	0	0	0	0	0	0	0	0	62.891057	0	36.848943
5	G-V166A-0260-0250-4112	0.26	0	0.25	0	0	0	0	0	0	0	62.7334195	0	36.7565805
6	G-V166A-0260-0500-4053	0.26	0.507	0	0	0	0	0	0	0	0	62.57136815	0	36.66163185
7	G-V166A-0260-0500-4114	0.26	0	0.5	0	0	0	0	0	0	0	62.575782	0	36.664218
8	G-V166A-0260-0750-4115	0.26	0	0.75	0	0	0	0	0	0	0	62.4181445	0	36.5718555
9	G-V166A-0260-1000-4056	0.26	0.989	0	0	0	0	0	0	0	0	62.26744305	0	36.48355695
10	G-V166A-0260-1000-4117	0.26	0	1	0	0	0	0	0	0	0	62.260507	0	36.479493
11	G-V166A-0260-1500-4058	0.26	1.496	0	0	0	0	0	0	0	0	61.9477542	0	36.2962458
12	G-V166A-0260-1500-4119	0.26	0	1.5	0	0	0	0	0	0	0	61.945232	0	36.294768
13	G-V166A-0260-2000-40510	0.26	2.003	0	0	0	0	0	0	0	0	61.62806535	0	36.10893465
14	G-V166A-0260-2000-41111	0.26	0	2	0	0	0	0	0	0	0	61.629957	0	36.110043
15	G-V166A-0750-0000-40512	0.75	0	0	0	0	0	0	0	0	0	62.5820875	0	36.6679125
16	G-V166A-0750-0000-41113	0.75	0	0	0	0	0	0	0	0	0	62.5820875	0	36.6679125
17	G-V166A-0750-0250-41114	0.75	0	0.25	0	0	0	0	0	0	0	62.42445	0	36.57555
18	G-V166A-0750-0500-40515	0.75	0.507	0	0	0	0	0	0	0	0	62.26239865	0	36.48060135
19	G-V166A-0750-0500-41116	0.75	0	0.5	0	0	0	0	0	0	0	62.2668125	0	36.4831875
20	G-V166A-0750-0750-41117	0.75	0	0.75	0	0	0	0	0	0	0	62.109175	0	36.390825
21	G-V166A-0750-1000-40518	0.75	0.989	0	0	0	0	0	0	0	0	61.95847355	0	36.30252645
22	G-V166A-0750-1000-41119	0.75	0	1	0	0	0	0	0	0	0	61.9515375	0	36.2984625
23	G-V166A-0750-1500-40520	0.75	1.496	0	0	0	0	0	0	0	0	61.6387847	0	36.1152153
24	G-V166A-0750-1500-41121	0.75	0	1.5	0	0	0	0	0	0	0	61.6362625	0	36.1137375
25	G-V166A-0750-2000-40522	0.75	2.003	0	0	0	0	0	0	0	0	61.31909585	0	35.92790415
26	G-V166A-0750-2000-41123	0.75	0	2	0	0	0	0	0	0	0	61.3209875	0	35.9290125
27	G-V166A-1250-0000-40524	1.25	0	0	0	0	0	0	0	0	0	62.2668125	0	36.4831875
28	G-V166A-1250-0000-41125	1.25	0	0	0	0	0	0	0	0	0	62.2668125	0	36.4831875
29	G-V166A-1250-0250-41126	1.25	0	0.25	0	0	0	0	0	0	0	62.109175	0	36.390825
30	G-V166A-1250-0500-40527	1.25	0.507	0	0	0	0	0	0	0	0	61.94712365	0	36.29587635
31	G-V166A-1250-0500-41128	1.25	0	0.5	0	0	0	0	0	0	0	61.9515375	0	36.2984625
32	G-V166A-1250-0750-41129	1.25	0	0.75	0	0	0	0	0	0	0	61.7939	0	36.2061
33	G-V166A-1250-1000-40530	1.25	0.989	0	0	0	0	0	0	0	0	61.64319855	0	36.11780145
34	G-V166A-1250-1000-41131	1.25	0	1	0	0	0	0	0	0	0	61.6362625	0	36.1137375
35	G-V166A-1250-1500-40532	1.25	1.496	0	0	0	0	0	0	0	0	61.3235097	0	35.9304903
36	G-V166A-1250-1500-41133	1.25	0	1.5	0	0	0	0	0	0	0	61.3209875	0	35.9290125
37	G-V166A-1250-2000-40534	1.25	2.003	0	0	0	0	0	0	0	0	61.00382085	0	35.74317915
38	G-V166A-1250-2000-41135	1.25	0	2	0	0	0	0	0	0	0	61.0057125	0	35.7442875
39	G-V166A-1750-0000-40536	1.75	0	0	0	0	0	0	0	0	0	61.9515375	0	36.2984625

ingredient properties | Grind | Letdown | Formulations | +



Coatings Aware Data Model

- Material relationships captured
  - Grind / Letdown mixing steps
  - Surfactant/HEUR addition
- Material properties captured
  - Molecular structures, particle sizes, HLB values

OVERVIEW OF THE DATA

Data modeling enables interoperable data

DATA OVERVIEW

- 200 formulations
- 20 ingredients
- Raw material properties …
- Ingredient roles
- Processing conditions

Ingredient Properties:

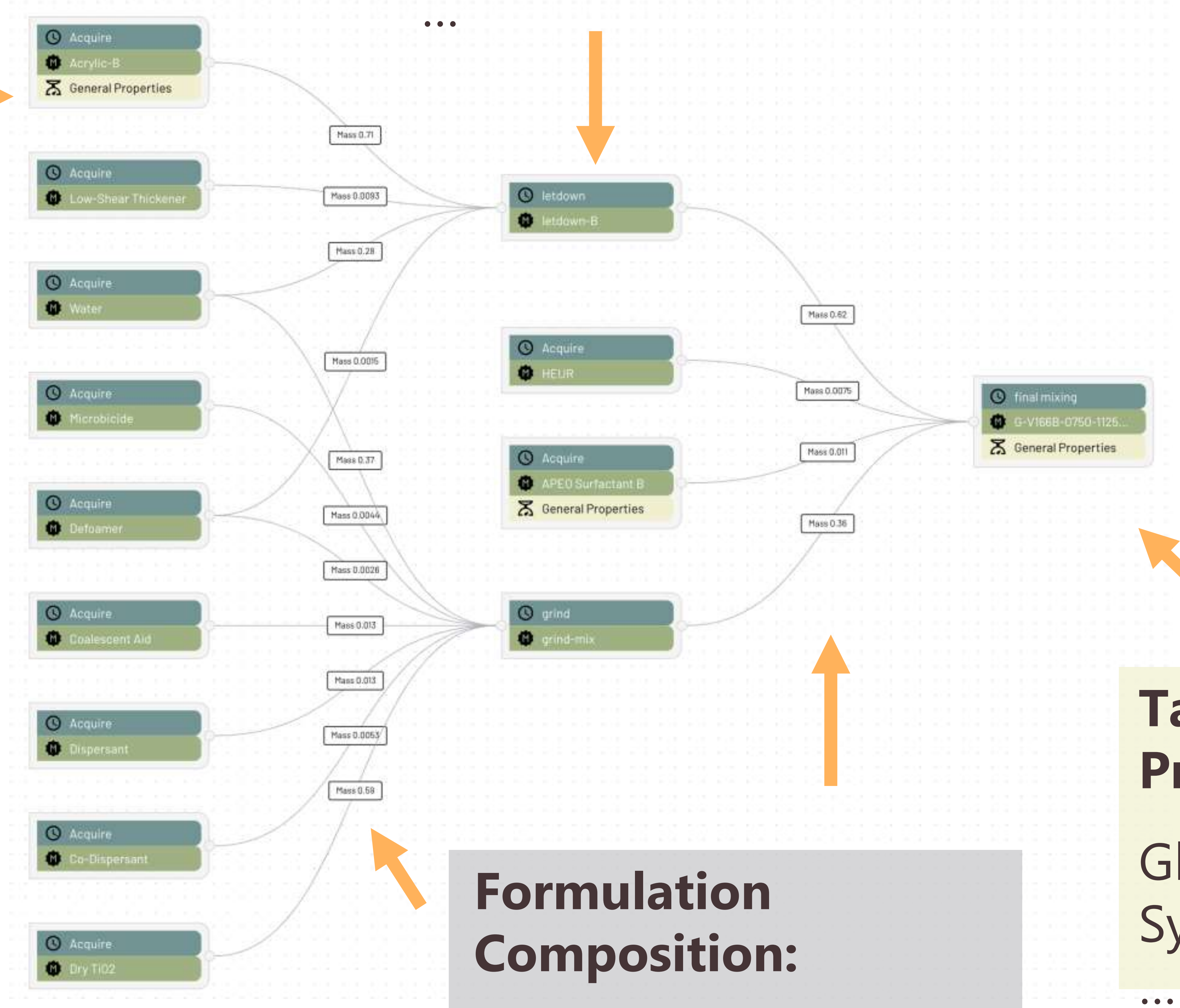
HLB value,
viscosity,
particle size,
molecular structure

Ingredient Roles:

Microbicide,
Surfactant,
Thickener
…

Processing Steps:

Grind Mixing
Letdown Mixing
…



Formulation Composition:

Mass composition,
Molar composition

Target Properties:

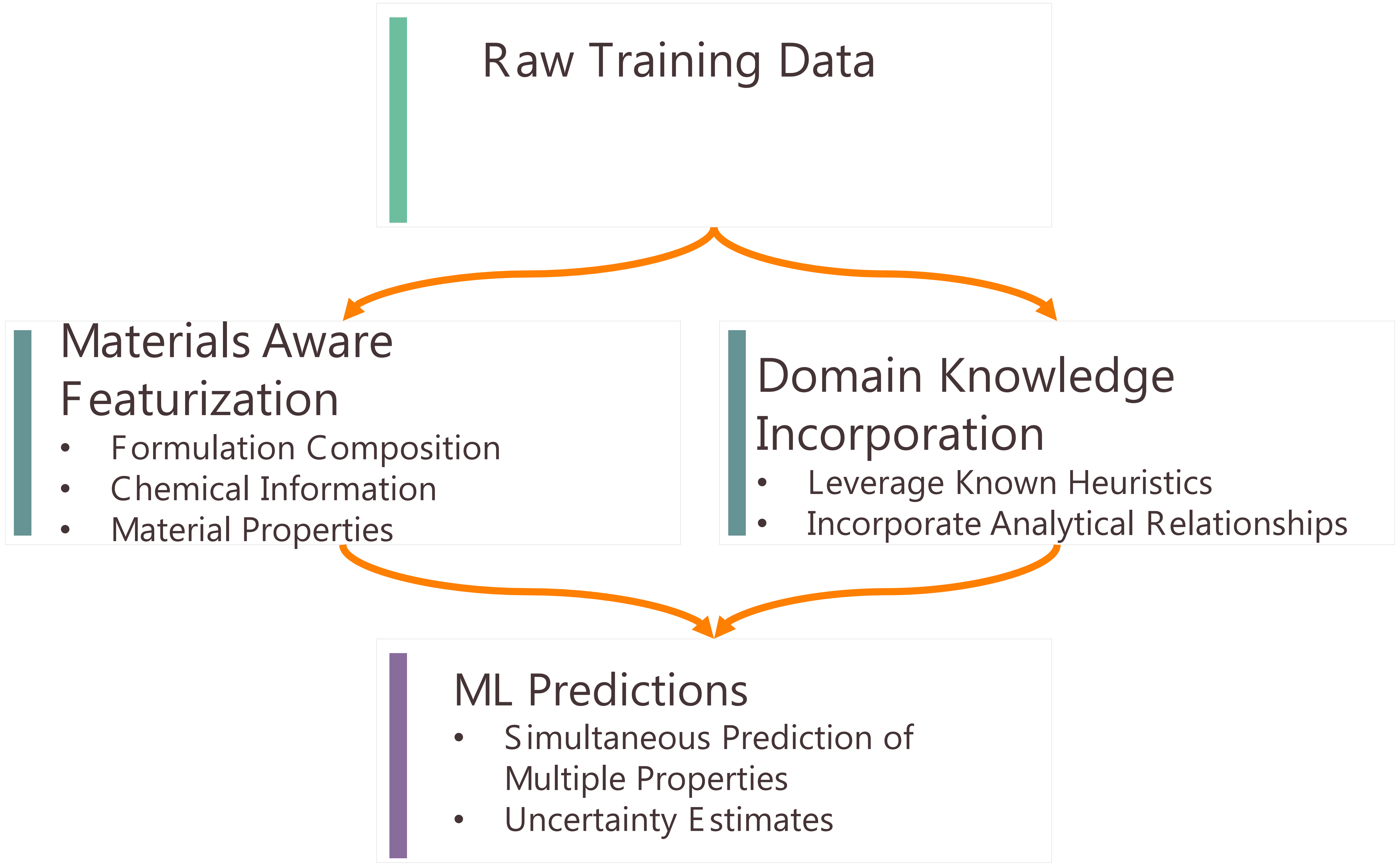
Gloss
Syneresis
…

AI MODEL

Incorporate Domain Knowledge into the Model Graph

Materials Aware Model Graph

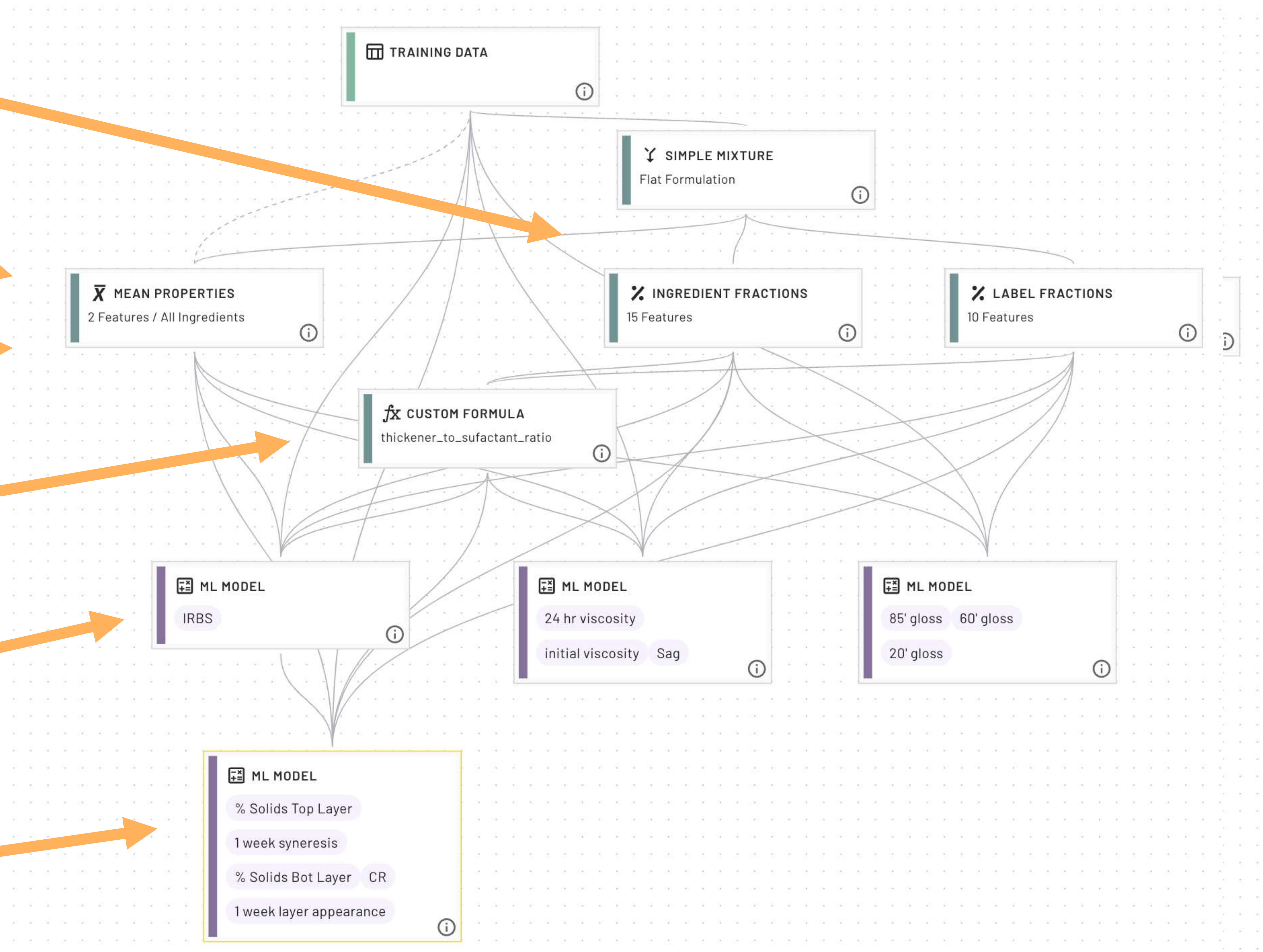
- Different graph nodes indicate different methods of “featurizing” the training data
- By imparting domain knowledge to the model, it can do more with less data.



BUILDING DOMAIN KNOWLEDGE INTO AN AI MODEL

What did we "tell" the model?

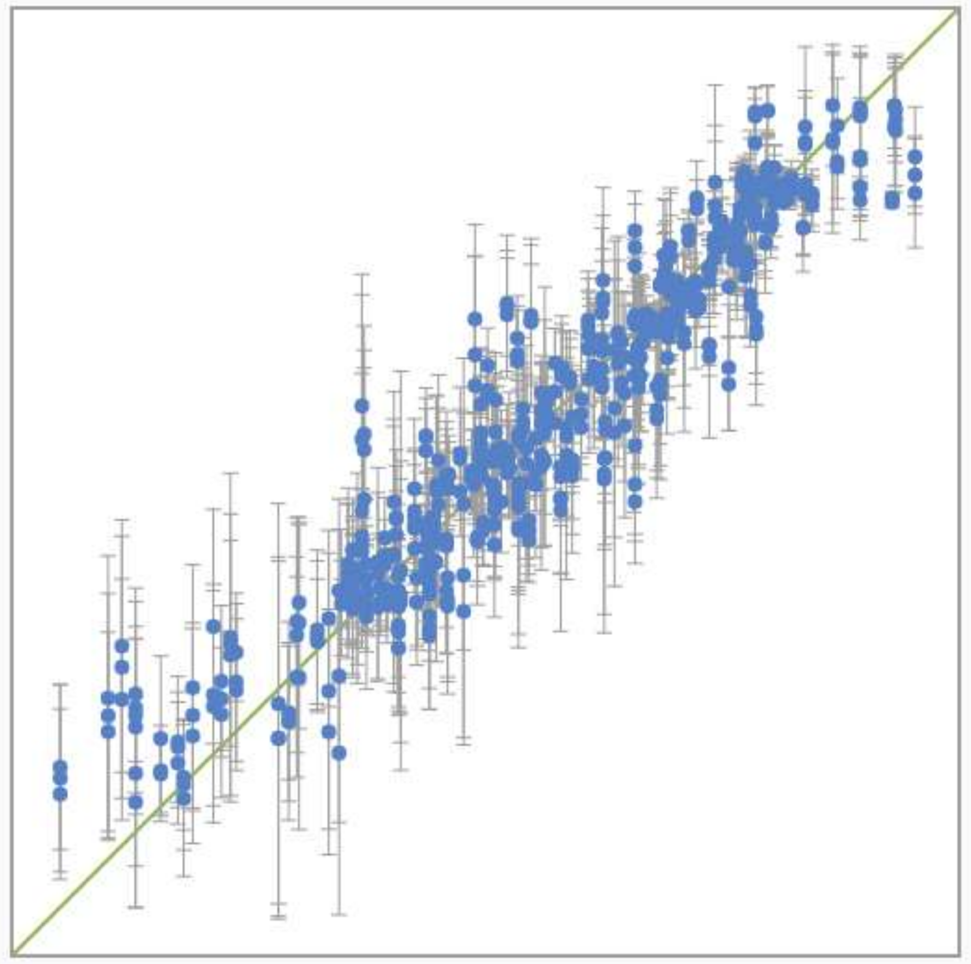
- The formulation composition dictates the system behavior
- The particle size of the acrylic impacts the gloss
- The HLB value of the surfactant impacts solution stability
- The thickener-to-surfactant ratio impacts the solution stability
- The IR Backscatter is a good leading indicator for syneresis
- Various measurements associated with solution stability exhibit correlated behavior



MODEL VALIDATION

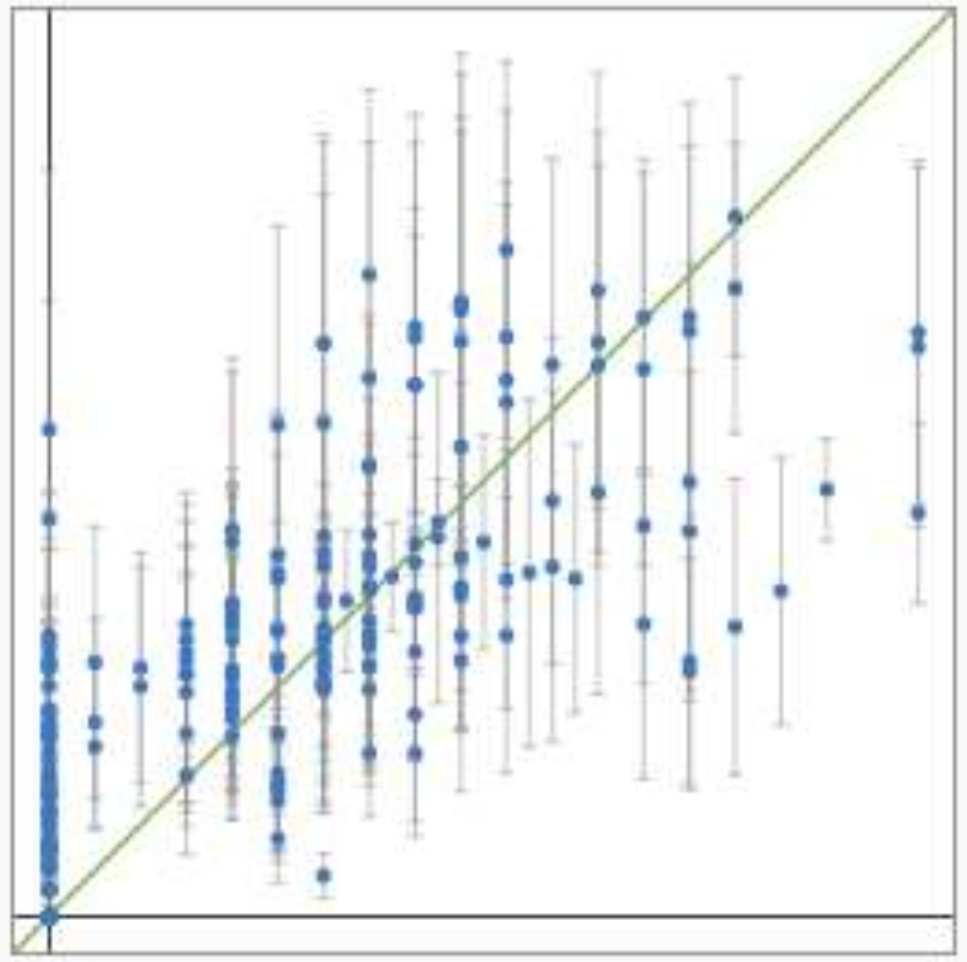
Cross-validation enables model evaluation over training data

24 hr viscosity



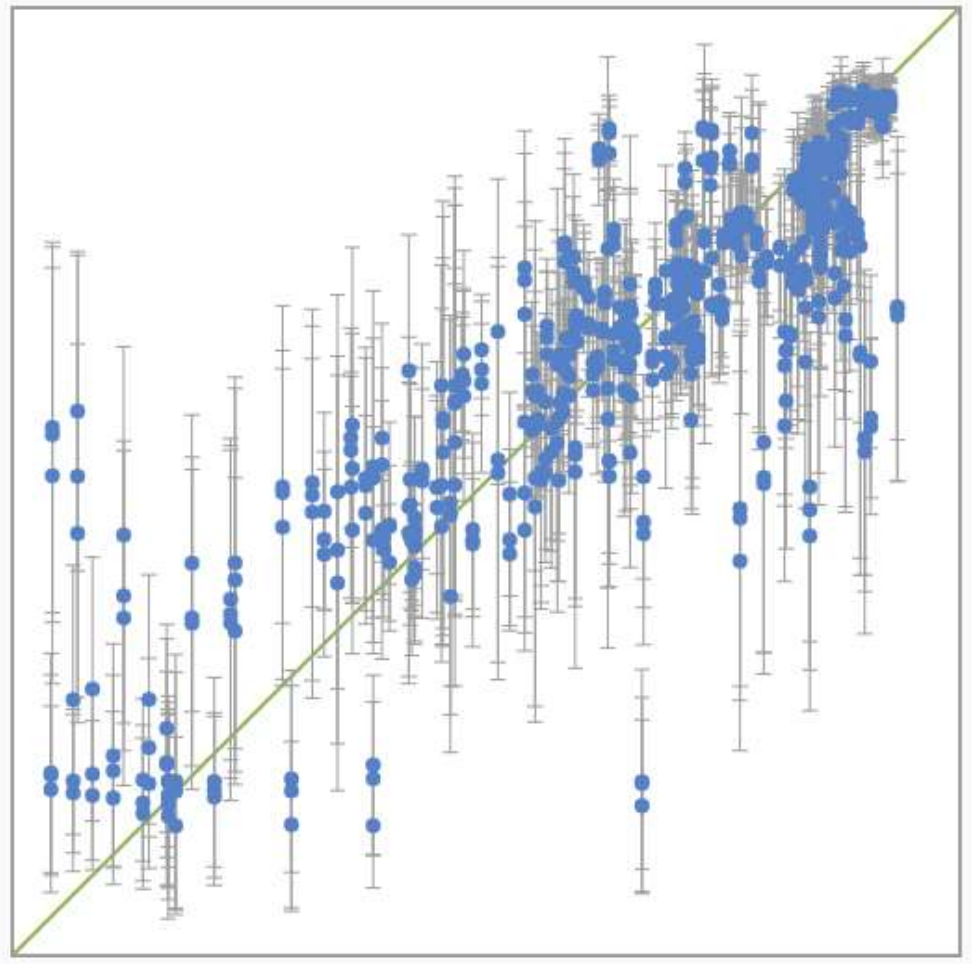
RMS ERROR	4.86 ± 0.760
NDM ERROR	0.306 ± 0.0478
R^2	0.904

1 week syneresis



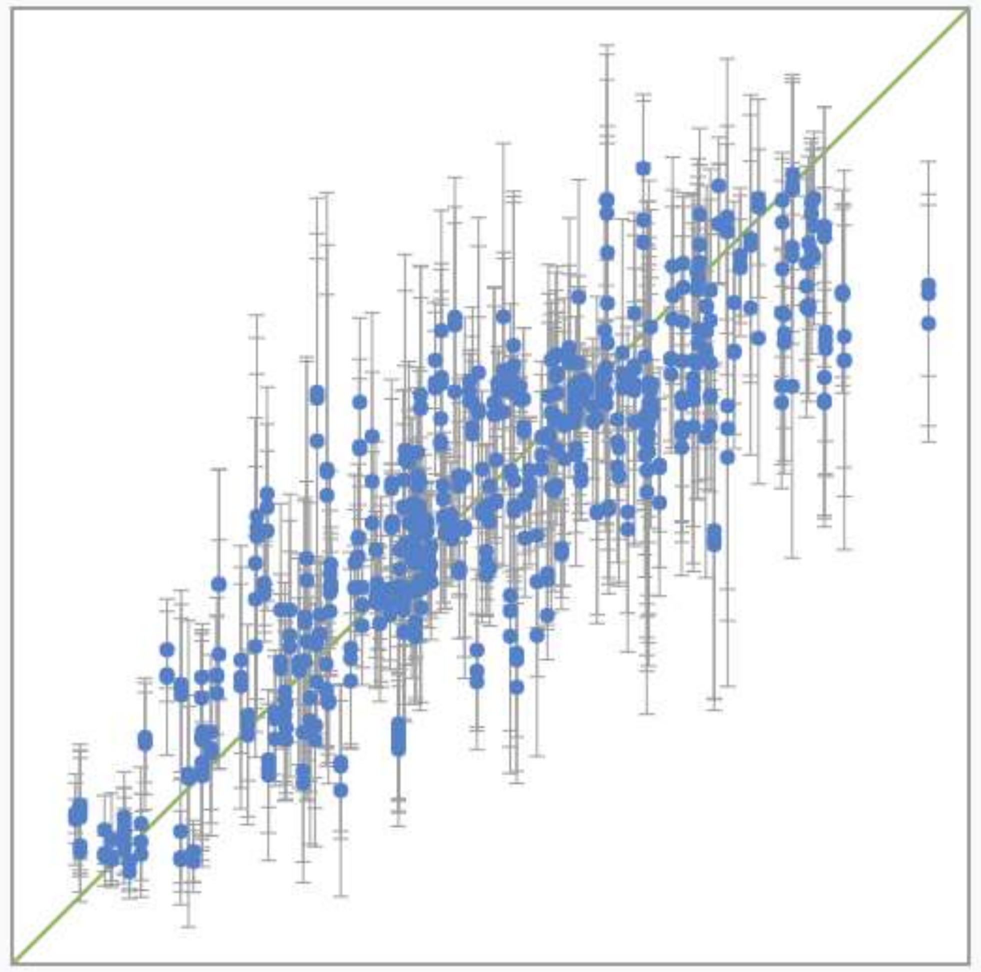
RMS ERROR	3.30 ± 0.112
NDM ERROR	0.709 ± 0.0240
R^2	0.496

60' gloss



RMS ERROR	7.45 ± 1.28
NDM ERROR	0.516 ± 0.0887
R^2	0.727

IRBS



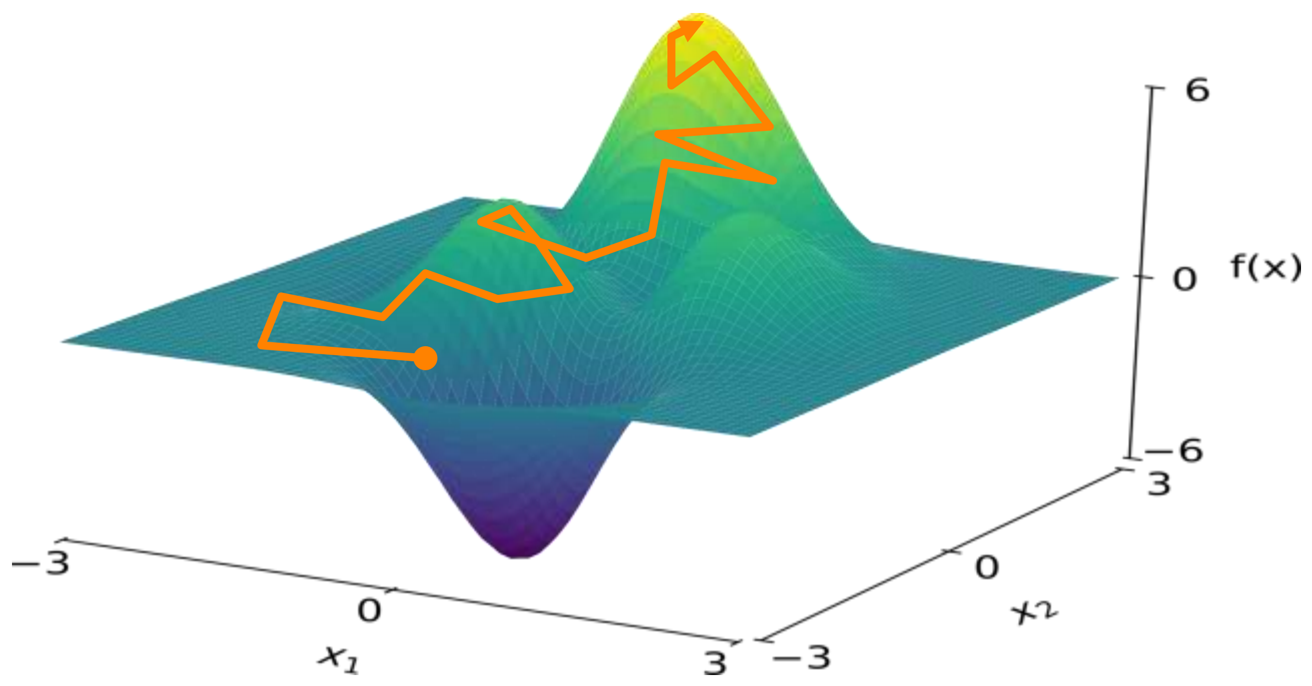
RMS ERROR	2.43 ± 0.311
NDM ERROR	0.498 ± 0.0638
R^2	0.749

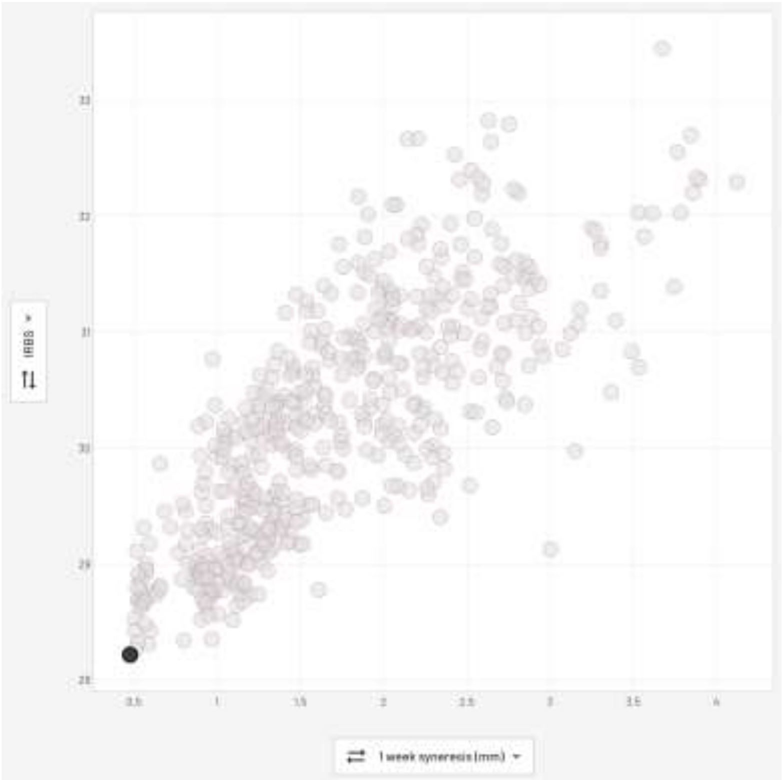
GENERATIVE DESIGN

Define your virtual lab bench, and set performance goals

- Constrain relevant degrees of freedom
  - Amount of an ingredient
  - Number of ingredients of a certain type
  - Ratio of mean properties
- Explore specific hypotheses
  - Swap ingredients
  - Adjust ingredient quantities
  - Adjust processing parameters

- View predicted performance along parameters of interest
  - Amount of an ingredient
  - Multiple predicted objectives
- Select promising candidates
  - Feasible based on domain knowledge
  - Test unintuitive combinations
  - Weigh uncertainty vs. performance





GENERATIVE DESIGN GUIDELINES AND GOALS

Define formulation heuristics, set design targets, and evaluate candidates

Define Generative Design "Search Space"

"Use all of the same ingredients as the training data, except for the APEO surfactants"

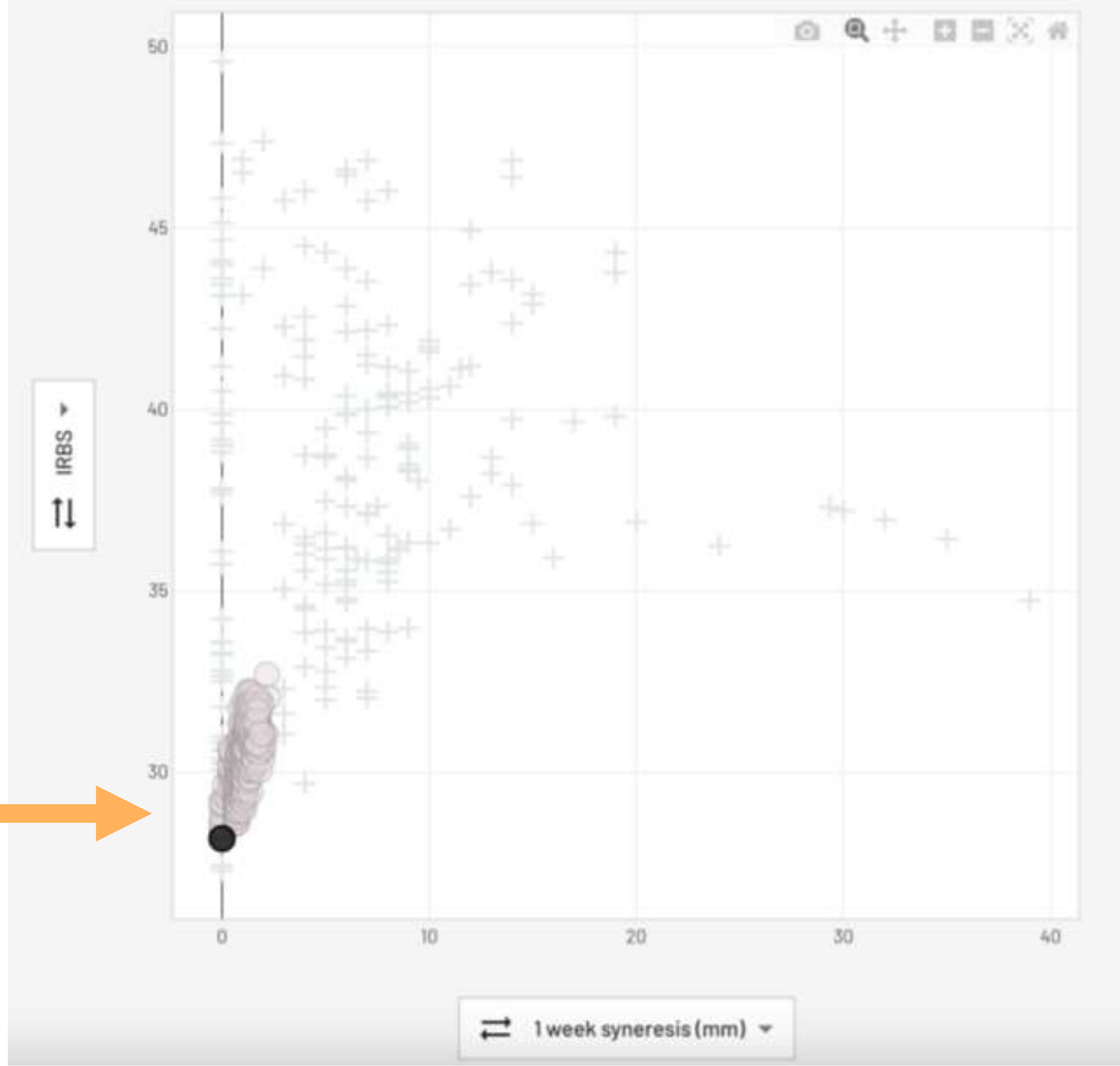
"Choose among six modern surfactants which are not APEO based, and propose formulations which will meet specific stability and appearance targets"

Set Generative Design Targets

Target 1: 0 syneresis after 1 week
Target 2: 60 degree gloss measurement of 60
Target 3: Viscosity of 100 s^{-1}

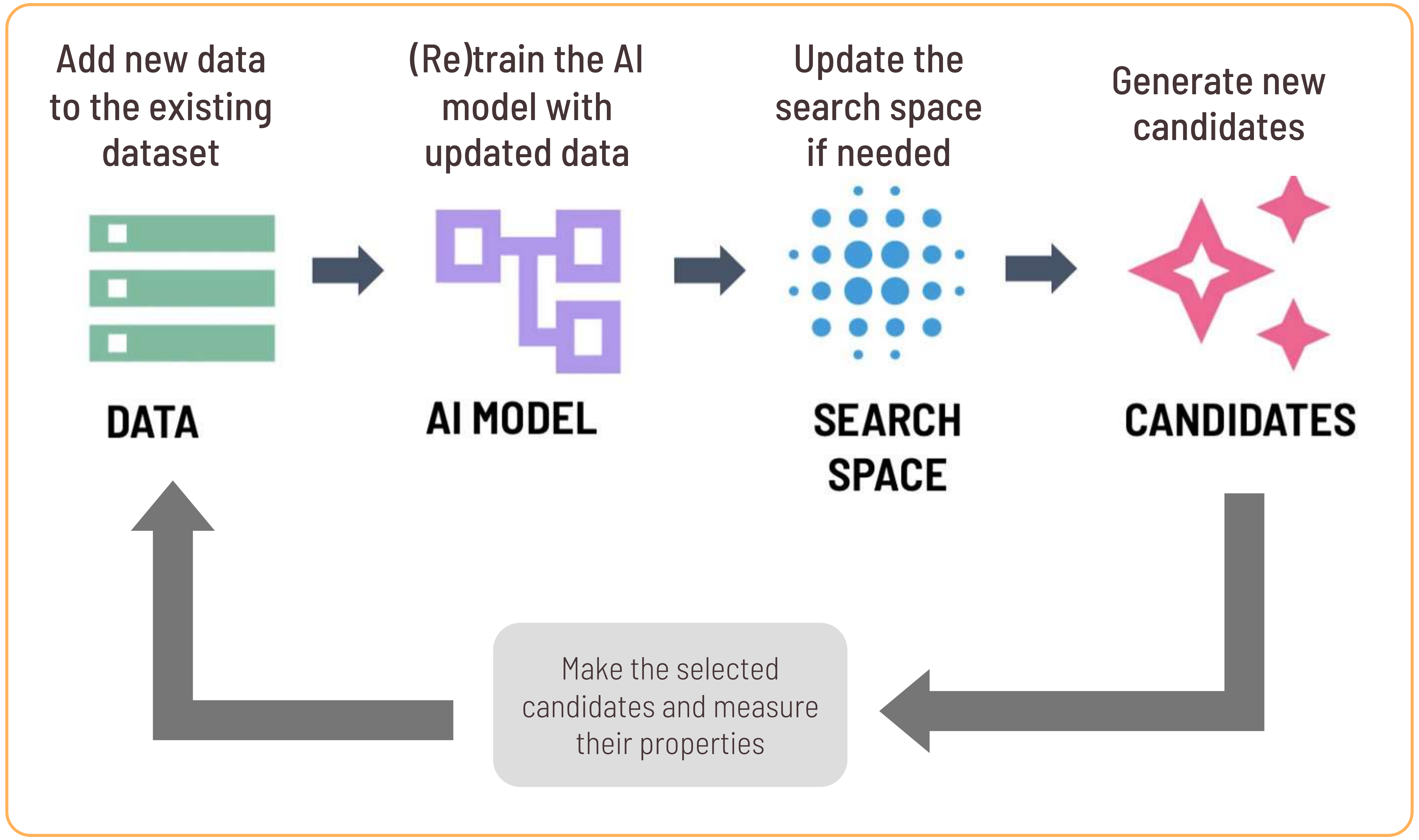
Use the machine learning model to "search" for optimal formulation candidates

Candidates near target values rank highly



CANDIDATE SELECTION & SEQUENTIAL LEARNING

SL enables efficient experimental design for multi-objective problems



Performance Targets:

- **Target 1:** 0 syneresis after 1 week
- **Target 2:** 60 degree gloss measurement of 60
- **Target 3:** Viscosity of 100 s^{-1}

Candidate selection criteria:

- High "Score"
- Realistic formulations
- Contains only non-APEO based surfactants
- Explore six *distinct* surfactants

SEQUENTIAL LEARNING RESULTS

SL was used to co-optimize gloss, stability, and rheology using novel surfactants

Sequential Learning Round 1

- Formulated 4 AI generated paints
- Re-trained model with new data
- Poor performance:
  - syneresis ✗
  - gloss ✗

Sequential Learning Round 2

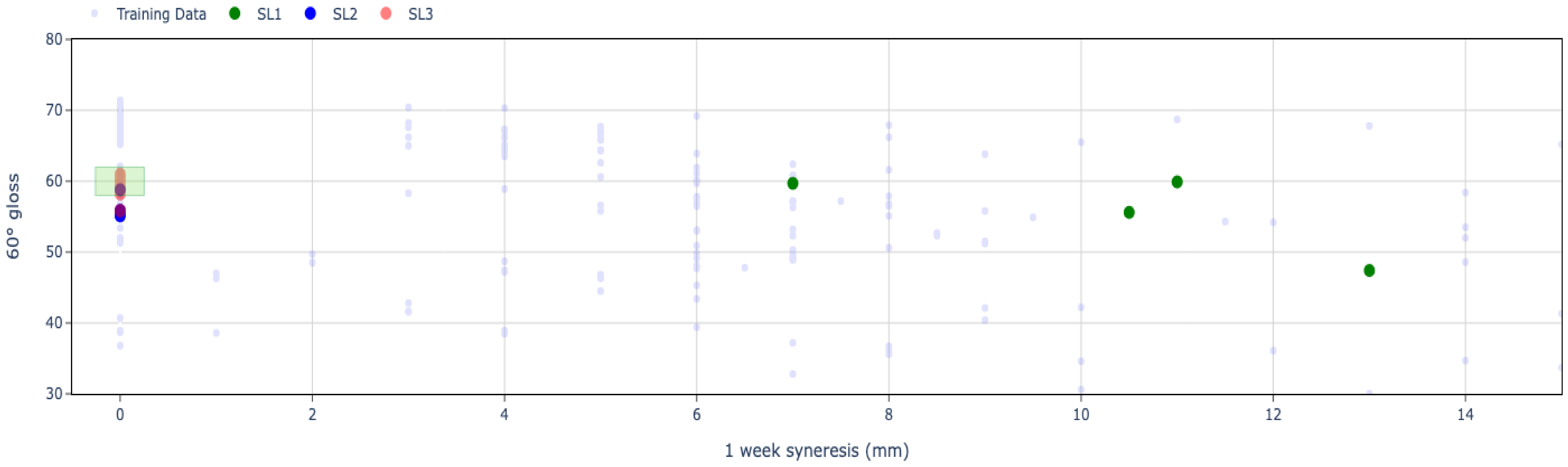
- Formulated 4 AI generated paints
- Re-trained model with new data
- Good syneresis, okay gloss
  - syneresis ✓
  - gloss ✗
  - scosity ✓

Sequential Learning Round 3

- Formulated 8 AI generated paints
- Re-trained model with new data
- Excellent Performance
  - syneresis ✓
  - gloss ✓

SEQUENTIAL LEARNING RESULTS

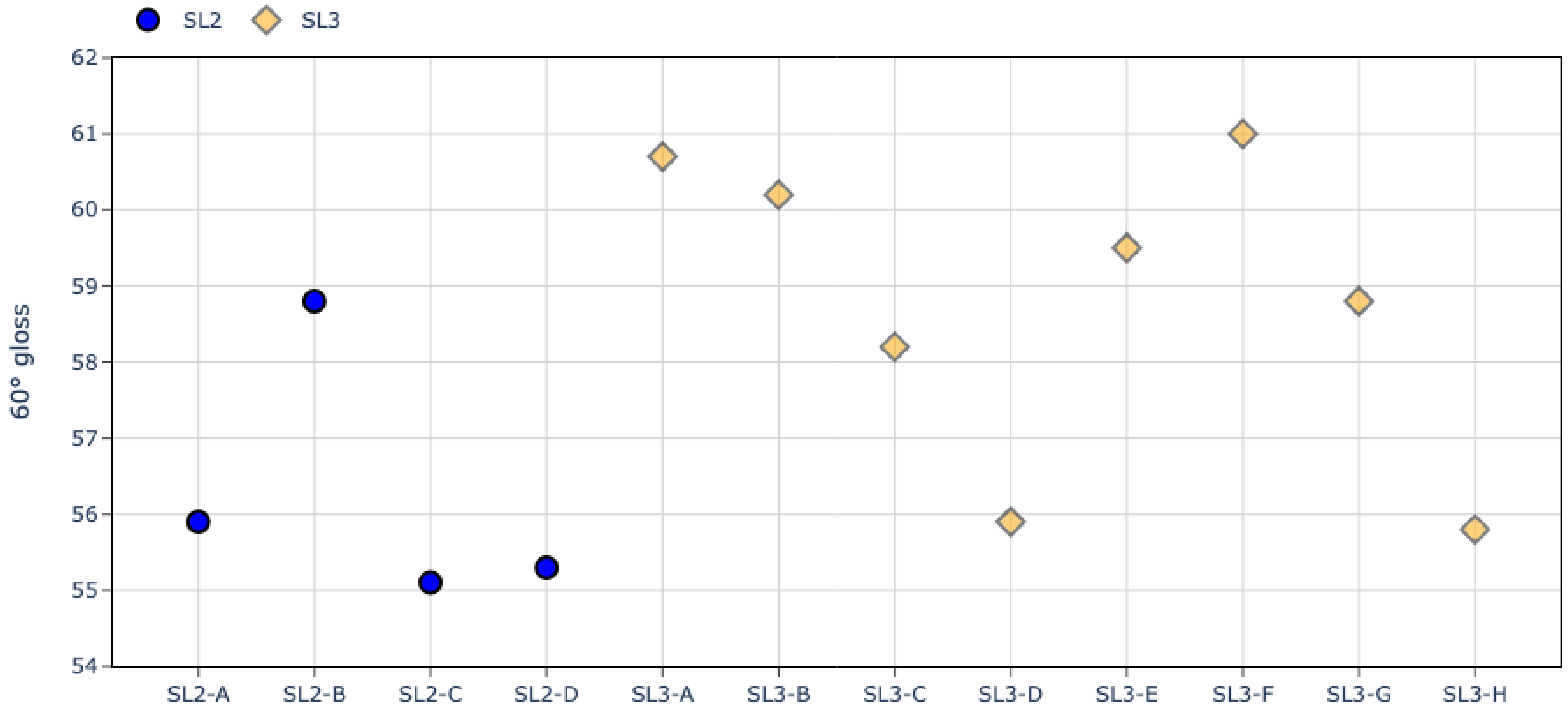
SL1 approached stability targets, but included failures



Performance optimization achieved with 90% less experiments

SEQUENTIAL LEARNING RESULTS

The focus of SL2 and SL3 was co-optimization of stability and gloss



SUMMARY + Q&A

SUMMARY

Sequential learning enables new material development

- Sequential learning enables co-optimization over multiple input and output dimensions
  - Structuring & annotating data
  - Incorporating domain knowledge
  - Defining realistic experiments
  - Executing sequential learning
- Sequential learning guided experiment enabled design of novel paint formulation
  - APEO free ingredients
  - Achieved target stability, gloss and rheology

Thanks to:

Tyler Bell, MS, *Account Executive*[*]
Erik Sapper, PhD, *Associate Professor*[†]
James Shannon, MS[†]

[*] Citrine Informatics
[†] California Polytechnic State University

Ready to get started?
Check out our webinar

How to prepare chemicals and materials data and teams for AI

QUESTIONS?

Will Erwin
werwin@citrine.io