

Applied machine learning

Introducing Intellegens

August 2024

Confidential

intellegens.com









Ben Pellegrini CEO Dr Gareth Conduit CSO

Dr Bogdan Nenchev Machine learning scientist

Agenda



10:00 Introductions

- **10:10** Machine learning, adaptive DOE, and Alchemite[™]
- 11:15 Break
- 11:30 Interactive demo
- **12.30** Lunch
- Afternoon Meetings about individual data sets

Confidential

Introducing Intellegens

- Innovative method extracts value from sparse, noisy data to solve complex, high-dimensional problems
- Strong focus on ease-of-deployment for immediate ROI
- Key use cases: Chemicals, Materials, Life Sciences, Manufacturing Processes, Food & Beverage

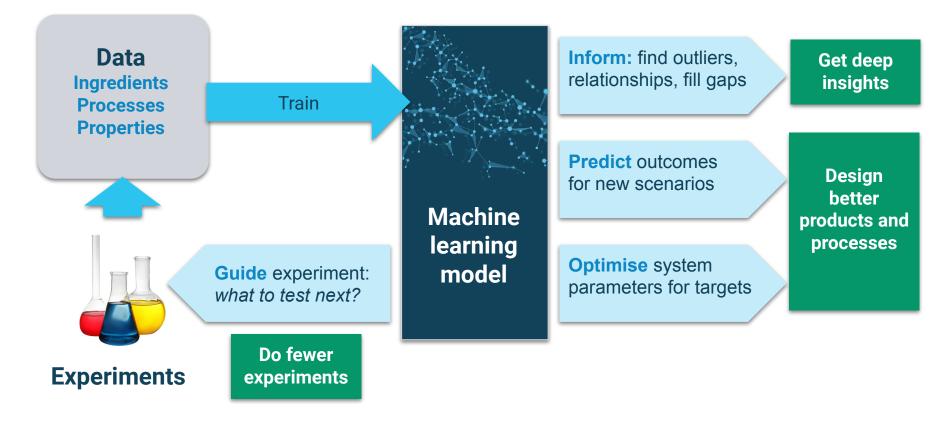








Experimental design accelerated by machine learning



Customers, collaborations, and partnerships

Selected case study examples



Other selected customers

Fast-moving consumer goods corporation **Biotech** Plastics, paints & coatings maker Global petrochemicals producers Construction chemicals provider Battery manufacturer Major food & beverages corporation Plant-based foods innovator Leading steelmaker Advanced materials organisation Mining and cement company Additive manufacturing specialist



Applied machine learning

Training Slides

Machine Learning Enhanced Adaptive Experimental Design

Introduction to Machine Learning, Adaptive Design of Experiments, and Alchemite™

Training Agenda

In this session, we'll be discussing:

• An overview of typical R&D business challenges.

• A traditional and adaptive R&D workflow.

• A discussion on how machine learning can enhance adaptive experimental design to expedite product development, with relevant case studies.



Confidential



Typical R&D Challenges

Typical R&D Business Challenges













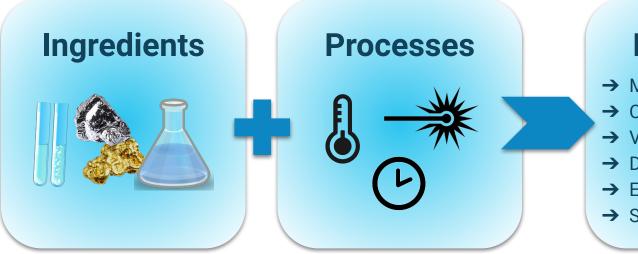
"We have to innovate, fast, to beat the competition" "Experimental programs cost \$millions and take years"

"We must meet
 net zero targets
 and grow market
 share"

"Price, supply, and regulatory issues are very disruptive" "We lose vital knowledge when people retire"

How to address these challenges while solving everyday R&D problems?

The trillion \$ R&D problem



Properties

→ Mechanical properties

→ Cost

- → Viscosity
- → Density
- → Environmental impact
- → Stability...

For chemicals, alloys, pharmaceuticals, plastics, foods, paints, cosmetics... (worth \$trillions!)

- High dimensional problem space
- Sparse, noisy data
- Costly, time-consuming experimental programs

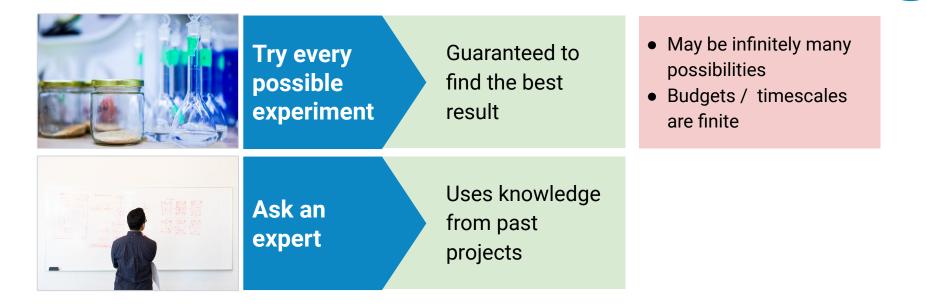


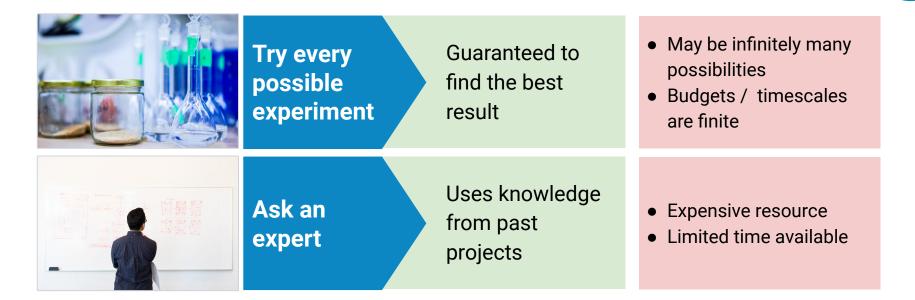
Confidential

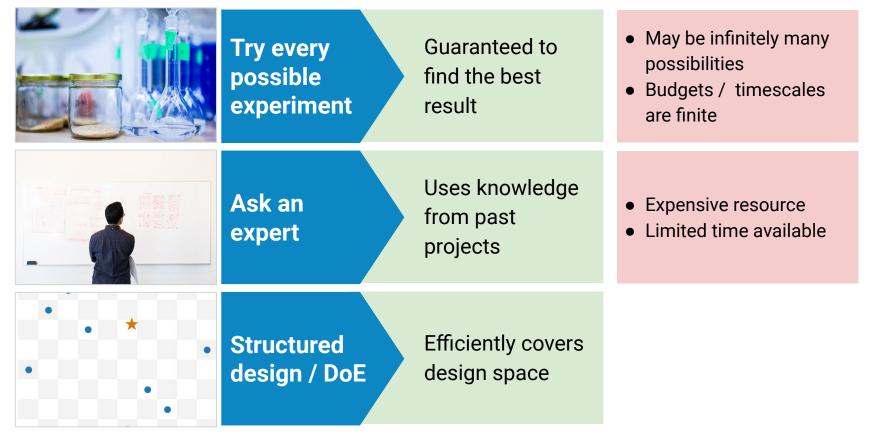
intellegens.com











Confidential

Try every possible experiment	Guaranteed to find the best result	 May be infinitely many possibilities Budgets / timescales are finite
Ask an expert	Uses knowledge from past projects	Expensive resourceLimited time available
Structured design / DoE	Efficiently covers design space	 May require a large number of experiments Requires statistical knowledge

Confidential



Section 1 Questions



What is Structured Design of Experiments?

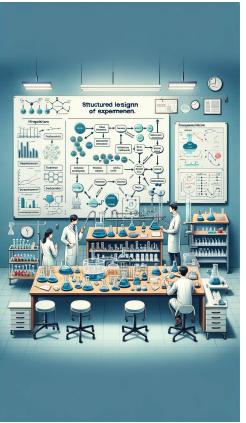
Structured Design of Experiments

• Structured DOE aims to cover design space as efficiently as possible

• Variety of structured designs available for different types of problem

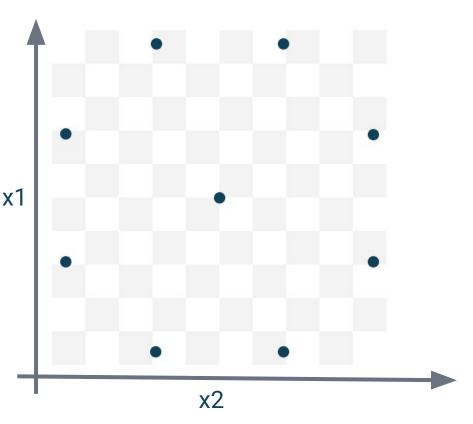
 Requires advanced statistical knowledge to understand which design to use and interpret results





Example Design: Central Composite Design (CCD)

- Used for detecting second order effects (e.g. x1 * x2)
- Effective for modelling quadratic relationships
- Helps understand the curvature of the surface
- Assumes no strong nonlinearity in the response variable



Structured Design of Experiments

- But real science is rarely limited to second-order effects
- With structured DoE, you are stuck with what the design tells you
- If that does not achieve your goals you often need to start again

Minimum of the function



What is Adaptive Experimental Design?

Confidential

intellegens.com

What is adaptive experimental design?

Start with a Plan

• Scientists begin with an initial experimental design.

Adapt Based on Results

- Adjust the plan as new data is collected.
- Increases efficiency and saves time and resources.

No Computers Needed

emerges.

- Simply review the data during the process.
- Modify the experimental plan as new information





Adaptive Experimental Design



Adaptive Experimental Design

Ĵ

It's good practice to format data into a structured format with rows for experiments and columns for inputs and outputs.

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1			

Adaptive Experimental Design



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1			



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1	Chemical 2	Yield
Experiment 1	99.1	0.27	



Try a formula



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1	99.1	0.27	149



Measure property



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1	99.1	0.27	149



Met requirements?



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1	Chemical 2	Yield
Experiment 1	99.1	0.27	149



Met requirements?



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	



Try a new formula



Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	170







Let's say we want a Yield of > 185

	INPUTS		PROPERTIES
	Chemical 1 Chemical 2		Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	170







	INP	PROPERTIES	
	Chemical 1	Chemical 2	Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	170







	INP	PROPERTIES	
	Chemical 1	Chemical 2	Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	170
Experiment 3	98.1	0.42	



	INP	PROPERTIES		
	Chemical 1	Chemical 2	Yield	
Experiment 1	99.1	0.27	149	
Experiment 2	98.6	0.35	170	
Experiment 3	98.1	0.42	179	



	INP	PROPERTIES		
	Chemical 1	Chemical 2	Yield	
Experiment 1	99.1	0.27	149	
Experiment 2	98.6	0.35	170	
Experiment 3	98.1	0.42	179	



	INP	PROPERTIES		
	Chemical 1	Chemical 2	Yield	
Experiment 1	99.1	0.27	149	
Experiment 2	98.6	0.35	170	
Experiment 3	98.1	0.42	179	
Experiment 4	98.4	0.55		



	INP	PROPERTIES		
	Chemical 1	Chemical 2	Yield	
Experiment 1	99.1	0.27	149	
Experiment 2	98.6	0.35	170	
Experiment 3	98.1	0.42	179	
Experiment 4	98.4	0.55	186	



	INP	PROPERTIES	
	Chemical 1	Chemical 2	Yield
Experiment 1	99.1	0.27	149
Experiment 2	98.6	0.35	170
Experiment 3	98.1	0.42	179
Experiment 4	98.4	0.55	186



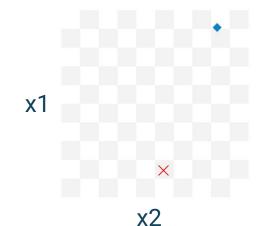


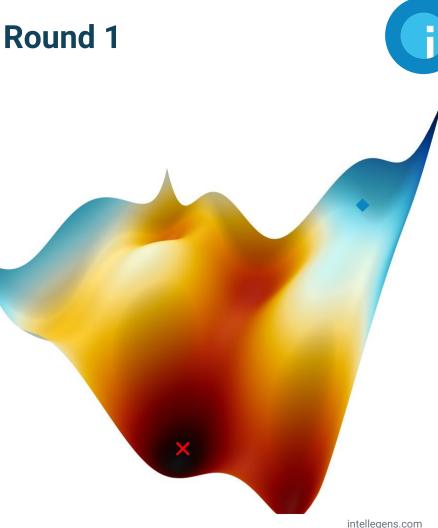
Now visualised

Confidential

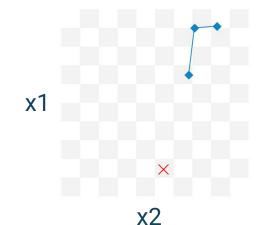
intellegens.com

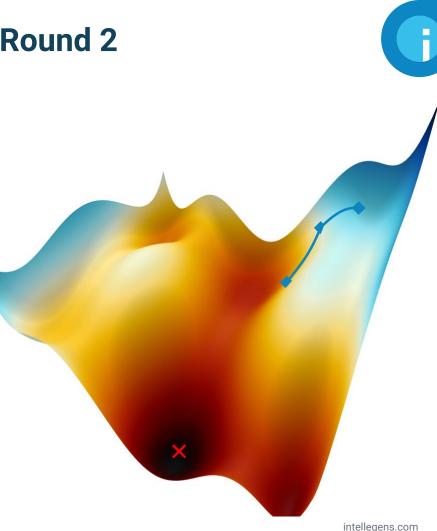
• Adaptive experimental design uses existing data to guide the best route to the goal



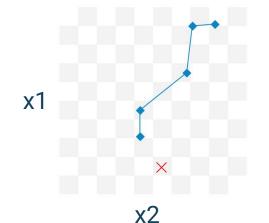


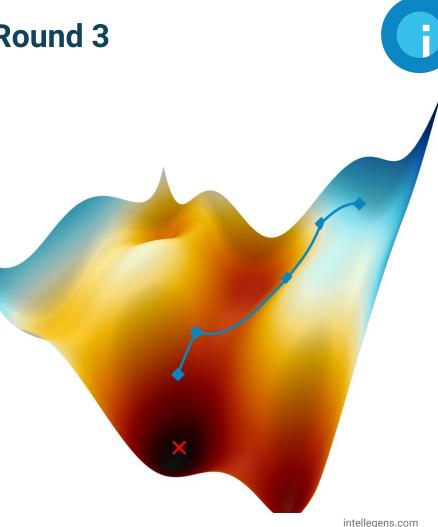
• Adaptive experimental design uses existing data to guide the best route to the goal



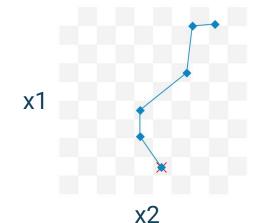


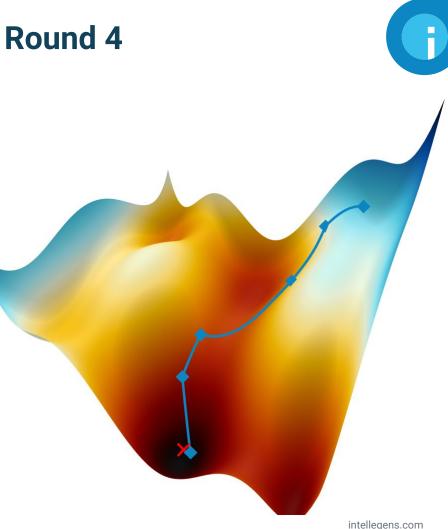
• Adaptive experimental design uses existing data to guide the best route to the goal





 Adaptive experimental design uses existing data to guide the best route to the goal







Sounds easy so what's the catch?

Confidential

intellegens.com



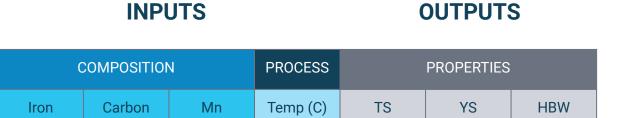


INPUTS

(PROCESS		
Iron	Carbon	Mn	Temp (C)

Confidential

How data actually is in R&D



OUTPUTS





INPUTS

OUTPUTS

	COMPOSITION			PROCESS	PROCESS PROPERTIES		
	Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
Experiment 1	99.1	0.27	0.6	842	76		149



INPUTS

OUTPUTS

	COMPOSITION			COMPOSITION PROCESS				PROPERTIES			
	Iron	Carbon	Mn	Temp (C)	TS	YS	HBW				
Experiment 1	99.1	0.27	0.6	842	76		149				
Experiment 2	98.6		0.9			80	170				



OUTPUTS INPUTS Tens of properties 🛌 Thousands of materials PROCESS COMPOSITION PROPERTIES Carbon Mn Temp (C) TS YS Iron HBW 99.1 0.27 0.6 842 76 **Experiment 1** 149 98.6 0.9 170 Experiment 2 80 0.42 1100 179 Experiment 3 98.4 0.55 0.8 118 70 Experiment 4

Confidential



This noisy, sparse, and high dimensional data is challenging for scientists to understand

		of properties		INPUTS			OUTPUTS		
materials			COMPOSITION		PROCESS	PROPERTIES			
of			Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
Thousands		Experiment 1	99.1	0.27	0.6	842	76		149
Thou	•	Experiment 2	98.6		0.9			80	170
		Experiment 3		0.42		1100			179
		Experiment 4	98.4	0.55	0.8		118	70	

j

Complex Inputs

• Problems often involve a lot more than a couple of input variables.

Multiple Targets

• Aim to optimize one output without compromising others.

Data Issues

• Data is often complicated, noisy, sparse, and high-dimensional.



Section 2 Questions

Confidential

intellegens.com



How can Machine Learning Help?

Confidential

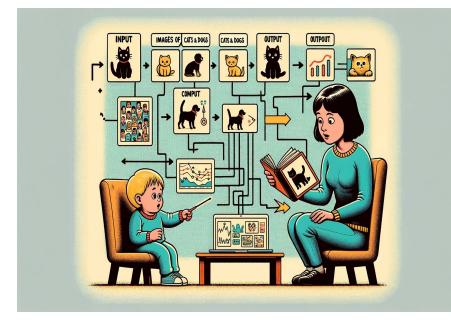
intellegens.com

Introduction to Machine Learning

Definition: Machine Learning (ML) is a branch of AI where computers learn from data to make predictions or decisions.

Core Idea: ML algorithms identify patterns in data and make decisions with minimal human input.

Applications: Used in image and speech recognition, recommendations, autonomous vehicles, etc.



Machine Learning - A Primer



STRUCTURED DATA



Inputs + Outputs

Machine Learning

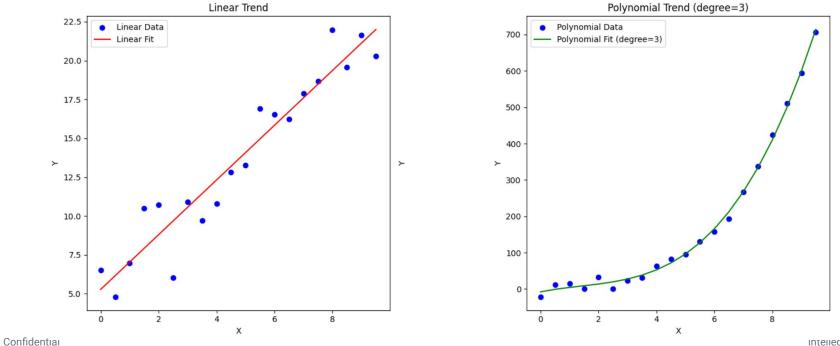


Extract **insights** Make **predictions**

Machine learning allows us to automatically create a 'model' to learn relationships between your inputs and outputs.

Simple Machine Learning Examples

We all use machine learning everyday without necessarily knowing. For instance:

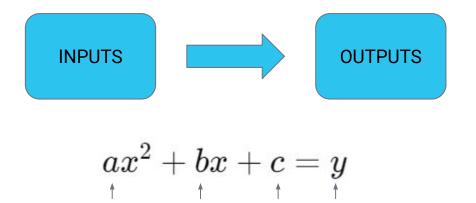


intellegens.com

Quadratic Equation



We can apply machine learning to determine the best values for *a*, *b*, and *c* to predict *y* most accurately across all training examples.



However, such a simple equation may struggle to capture complex relationships.

>30 418 18-30

High Risk

Weight

460

Low Risk

Relating age, weight, and smoker etc. (inputs) -> to the risk of heart attack (output).

Decisions trees allow us to model

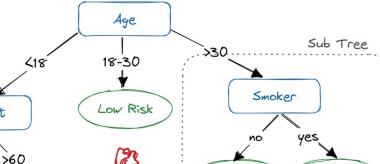
nonlinear relationships with many

variables.

Confidential

These models can capture more complicated relationships.

More Complicated Machine Learning Example



Understanding the risks to prevent a heart attack.

Low Risk

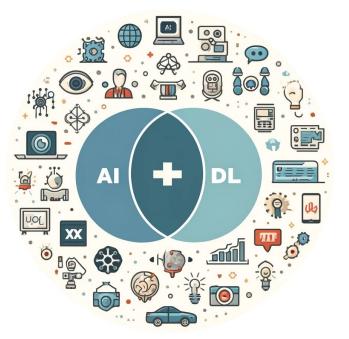


High Risk

The Steps to Train and Deploy a Machine Learning Model

Steps:

- Data Collection: Gathering data.
- **Preprocessing:** Cleaning and preparing data.
- Model Selection: Choosing the right algorithm
- **Training:** Teaching the model with data.
- **Evaluation:** Testing the model's accuracy.
- **Deployment:** Using the model for predictions.





How can machine learning enhance adaptive experimental design?





Collect an initial dataset to train a preliminary machine learning model.



Collect an initial dataset to train a preliminary machine learning model.

	••••••	COMPOSITION			PROCESS PROPERTIES			
		Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
	Experiment 1	99.1	0.27	0.6	842	76		149
V	Experiment 2	98.6		0.9			80	170
	Experiment 3		0.42		1100			179
	Experiment 4	98.4	0.55	0.8		118	70	

Thousands of materials

Tens of properties



With the initial model we can make predictions and fill in missing values.

••••								
		COMPOSITION			PROCESS	PROPERTIES		
		Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
V	Experiment 1	99.1	0.27	0.6	842	76	64±2	149
	Experiment 2	98.6	0.37±0.1	0.9	892±17	90±5	80	170
	Experiment 3	98.8±0.8	0.42	0.7±0.1	1100	91±9	77±3	179
	Experiment 4	98.4	0.55	0.8	980±38	118	70	241±6

Confidential

Thousands of materials

Tens of properties



We can apply constraints to our inputs to refine our design space and set requirements for outputs.

COMPOSITION PROCESS **PROPERTIES** Carbon Mn Temp (C) TS YS HBW Iron 99.1 0.27 0.6 842 76 64±2 149 **Experiment 1** 0.9 98.6 0.37±0.1 892±17 90±5 80 170 Experiment 2 98.8±0.8 0.42 0.7±0.1 1100 91±9 77±3 179 **Experiment 3** 98.4 0.8 0.55 980±38 118 70 241±6 Experiment 4 0.2-0.5 >115 >230 Requirement

Thousands of materials

Tens of properties

Confidential



Using the constraints and targets, the model makes a suggestion for a new experiment.

	COMPOSITION				PROPERTIES							
	Iron	Carbon	Mn	Temp (C)	TS	YS	HBW					
Experiment 1	99.1	0.27	0.6	842	76	64±2	149					
Experiment 2	98.6	0.37±0.1	0.9	892±17	90±5	80	170					
Experiment 3	98.8±0.8	0.42	0.7±0.1	1100	91±9	77±3	179					
Experiment 4	98.4	0.55	0.8	980±38	118	70	241±6					
Suggestion	98.7	0.45	0.8	1010								

Thousands of materials

Tens of properties

Confidential



The model predicts the performance of this virtual formulation.

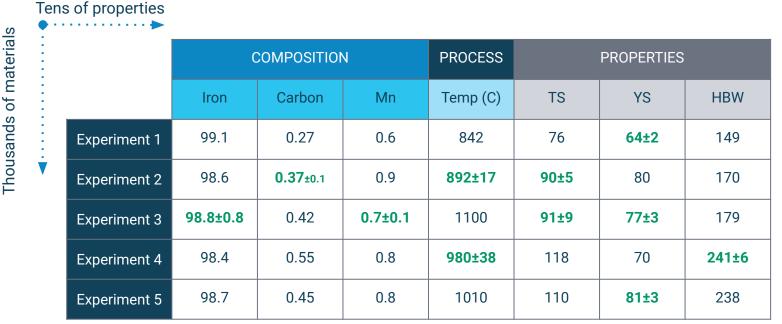
COMPOSITION PROCESS **PROPERTIES** Carbon Mn Temp (C) TS YS HBW Iron 99.1 0.27 0.6 842 76 64±2 149 **Experiment 1** 98.6 0.9 0.37±0.1 892±17 90±5 80 170 Experiment 2 98.8±0.8 0.42 0.7±0.1 1100 91±9 77±3 179 **Experiment 3** 98.4 0.8 118 0.55 980±38 70 241±6 Experiment 4 230±10 98.7 0.45 0.8 1010 105±4 81±3 Suggestion

Thousands of materials

Tens of properties



We measure the real composition, processing, and properties.





We check our requirements.

2		s of properties							
			C	COMPOSITION		PROCESS	PROPERTIES		
			Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
		Experiment 1	99.1	0.27	0.6	842	76	64±2	149
	V	Experiment 2	98.6	0.37±0.1	0.9	892±17	90±5	80	170
		Experiment 3	98.8±0.8	0.42	0.7±0.1	1100	91±9	77±3	179
		Experiment 4	98.4	0.55	0.8	980±38	118	70	241±6
		Experiment 5	98.7	0.45	0.8	1010	110	81±3	238

Thousands of materials



Is HBW > 230?

0	Tens	s of properties							
			C	COMPOSITION			PROPERTIES		
			Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
		Experiment 1	99.1	0.27	0.6	842	76	64±2	149
	V	Experiment 2	98.6	0.37±0.1	0.9	892±17	90±5	80	170
		Experiment 3	98.8±0.8	0.42	0.7±0.1	1100	91±9	77±3	179
		Experiment 4	98.4	0.55	0.8	980±38	118	70	241±6
		Experiment 5	98.7	0.45	0.8	1010	110	81±3	238

Thousands of materials



ls TS > 115?

• •	is of properties							
		COMPOSITION			PROCESS	PROPERTIES		
		Iron	Carbon	Mn	Temp (C)	TS	YS	HBW
	Experiment 1	99.1	0.27	0.6	842	76	64±2	149
	Experiment 2	98.6	0.37±0.1	0.9	892±17	90±5	80	170
	Experiment 3	98.8±0.8	0.42	0.7±0.1	1100	91±9	77±3	179
	Experiment 4	98.4	0.55	0.8	980±38	118	70	241±6
	Experiment 5	98.7	0.45	0.8	1010	110	81±3	238

Thousands of materials

Confidential



If goals aren't achieved, we can use this new data to update the model, and then repeat the process.

COMPOSITION PROCESS **PROPERTIES** Carbon Mn Temp (C) TS YS HBW Iron 76 99.1 0.27 0.6 842 149 **Experiment 1** 98.6 0.9 80 170 Experiment 2 1100 0.42 179 Experiment 3 98.4 0.55 0.8 118 70 Experiment 4 98.7 0.45 0.8 1010 110 238 Experiment 5

Confidential

Thousands of materials

Tens of properties



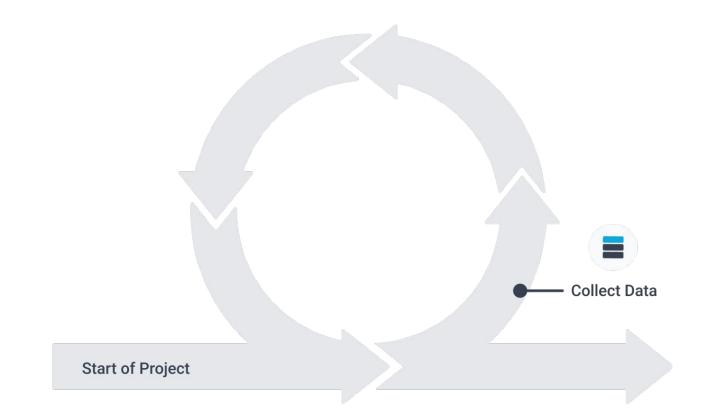
Now visualised

Confidential

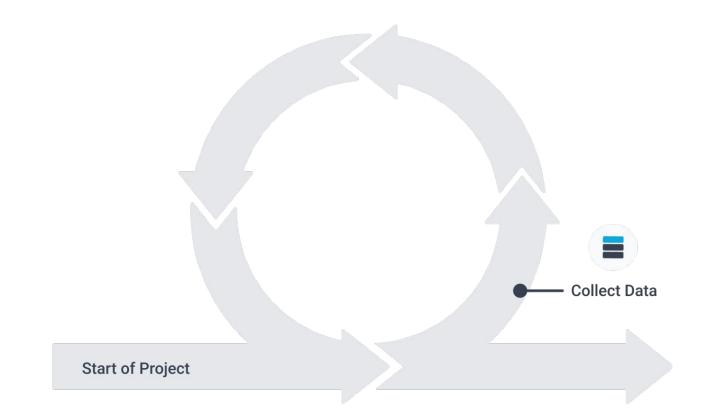
Start of Project

Start of Project

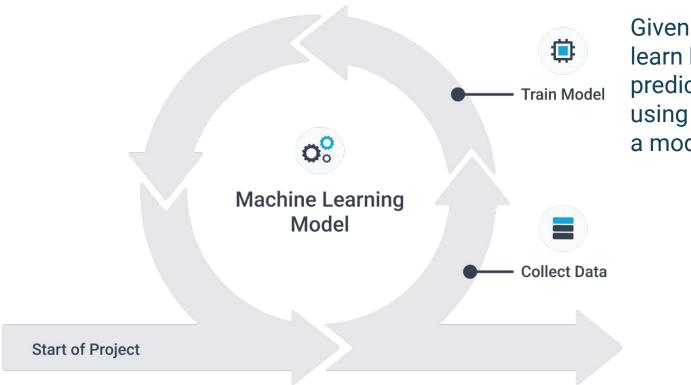
Set Requirements



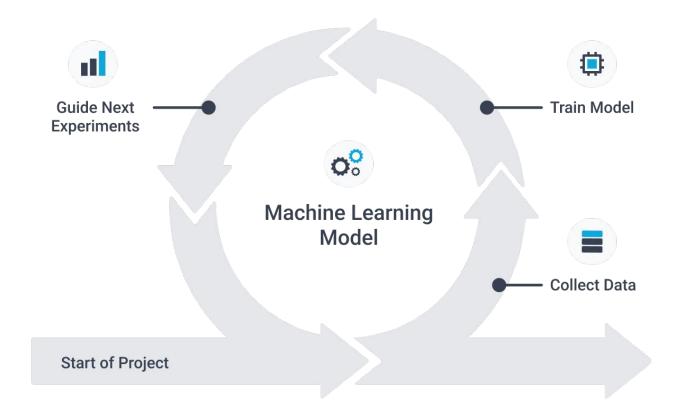


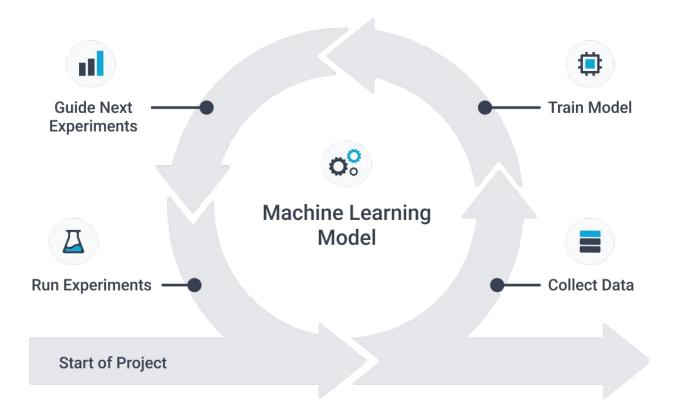


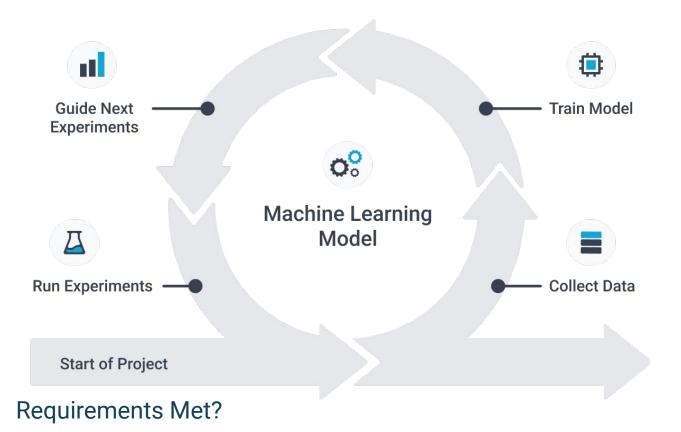


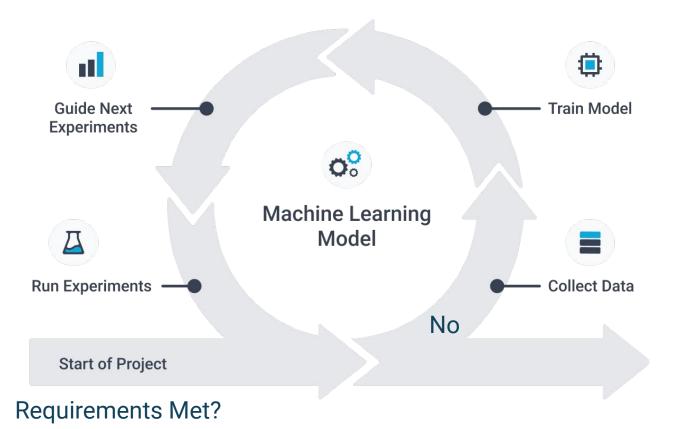


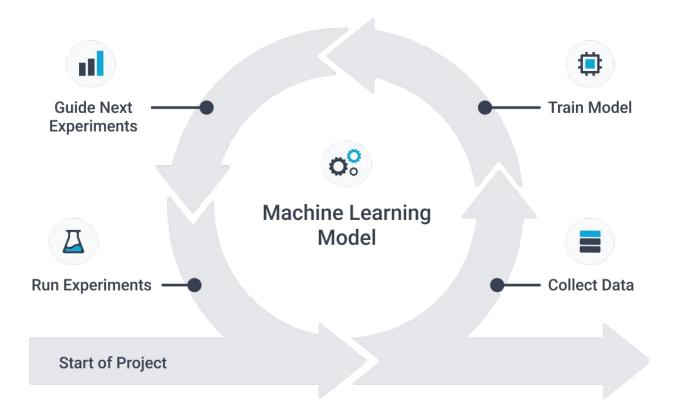
Given your data, learn how to predict outputs using inputs (train a model)

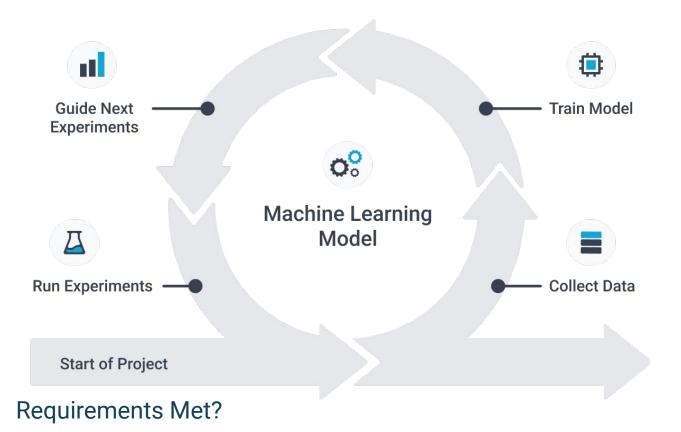


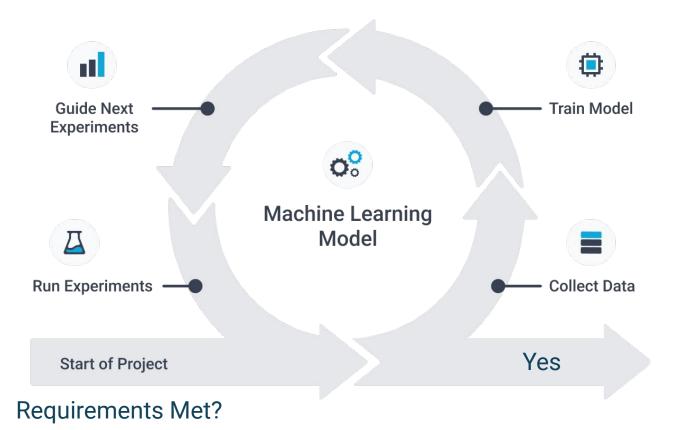


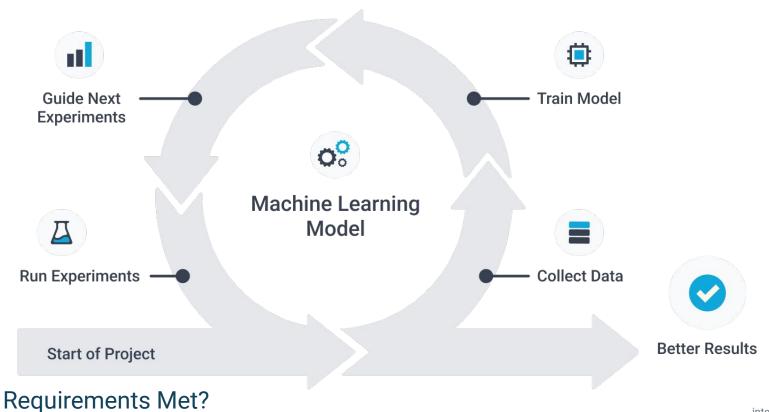












Why Machine Learning is Ideal for R&D



Learns Nonlinear Relationships: Understands complex patterns in data.

Handles High-Dimensional Data: Efficiently processes large and complex datasets.

Robust to Noise and Missing Data: Maintains performance despite data imperfections.

Suggests New Experiments: Recommends new experimental points to optimize outcomes.

Continuously Improves: Adapts and enhances models with new data.



Section 3 Questions

Confidential



Why Alchemite[™]?

Why is there not greater adoption of ML?





Why Alchemite[™]?



- 1. Setting up a machine learning framework for adaptive experimental design is hard.
- 2. Deploying models to teams is even harder.
- 3. The speed benefits of ML-AED are only realized with a platform.
- 4. If each chemist needed to know machine learning to set up the framework for ML-AED, they wouldn't be doing any R&D...
- That's why we have Alchemite[™] a tool to handle the machine learning for you.

What is Alchemite[™]



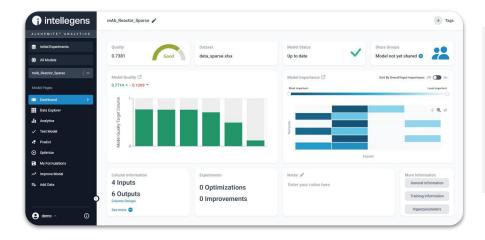
Alchemite provides the following:

- 1. Alchemite is a no code tool that allows adaptive experimental design to be applied to R&D data
- 2. It is built on top of the ML-AED framework
- 3. It is a platform that allows machine learning models to be built on noisy, sparse, and high dimensional structured data
- 4. It provides accurate uncertainty quantification to guide experimental design
- 5. It leverages the ML-AED framework to accelerate R&D



Products and Successful Applications of the Technology

Alchemite[™] – *ML* design platform



Real data

- Unique method for sparse, noisy data
- Accurate uncertainty quantification
- Complex, multi-dimensional problems

Fit to workflow

- Simple-to-use web user interface
- Designed for key analysis tasks
- Coding not required (optional API)

Trust

- Alchemite[™] Success Team
- Explainable AI tools
- Proven case studies, pilot projects

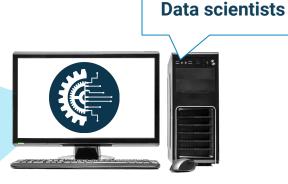
The Alchemite[™] Products

Scientists & engineers

Alchemite[™] Analytics

- Web UI insights on your desktop
- Optimise products, extract value from data, guide experiment





Alchemite[™] Engine

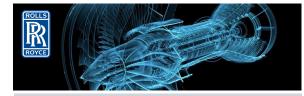
- Integrate into your workflows (API, Python)
- Advanced configuration, deploy models

Alchemite[™] Success

- Use our expertise in applying ML
- Ranging from 'getting started' advice to full project management



Successful applications





Design of an aero alloy

Multi-million \$ savings in discovery of new alloy

Ink reformulation

Cut experimental timescales from months to minutes



Drug discovery

Predict pharmacokinetics to improve compound selection



Component design

Validating Alchemite[™] for advanced engineering at NASA



Additive manufacturing

Reduce the number of tests by an order of magnitude



Flavours and fragrances

Predict the sensory properties of compounds

Successful applications





Optimising steels

Arcelor joint venture - insights from microstructure data

Drug discovery

Optimising kinase profiling programs despite sparse data

Virtual experimentation

Exploring new parts of material property space



Combatting wear

Better surface treatments with fewer alloying elements



Materials and processes

Drive efficiency in the use of simulation and experiment



Automotive catalysts

Reducing the amount of experimental work required

Customers, collaborations, and partnerships



Selected case study examples



Some other customers

Fast-moving consumer goods corporation **Biotech** Plastics, paints & coatings maker Global petrochemicals producers Construction chemicals provider Battery manufacturer Major food & beverages corporation Plant-based foods innovator Leading steelmaker Advanced materials organisation Mining and cement company Additive manufacturing specialist



BREAK

Confidential



Demo

Confidential

What's next?

info@intellegens.com

in /company/intellegensai

Further information intellegens.com/doe-made-easy/





SUBSCRIBE for monthly newsletter and webinar alerts

intellegens.com/subscribe