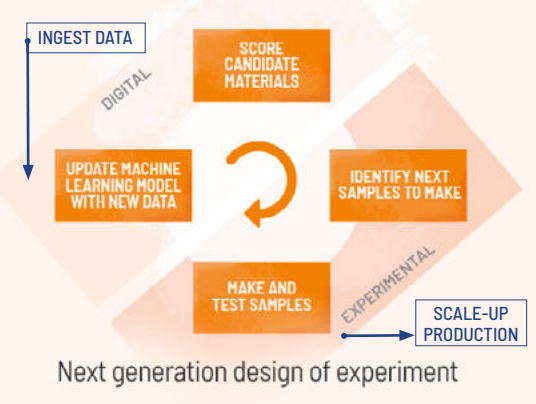
AI-GUIDED SEQUENTIAL LEARNING VS. TRADITIONAL DESIGN OF EXPERIMENT

CITRINE INFORMATICS
citrine.io • info@citrine.io

The Citrine Platform uses the next generation of design of experiment – *Sequential Learning*. This does not replace classical Design of Experiment (DOE), rather it is a new method that performs better in specific applications.

WHAT IS SEQUENTIAL LEARNING?

Sequential Learning (SL) is essentially an iterative form of DOE, where each step improves the AI model that is used to choose the next experiments to perform. Initial training data is used to create a first AI model to predict the target properties of all the candidate materials in the design space. Each prediction comes with a calculated uncertainty. The uncertainty will be higher in regions where there is less high-quality data in the training set.



The first small batch of experiments to carry out can now be chosen by Product Developers based on information delivered by the platform on candidates that are:

- Predicted to have the best target properties
- Have the highest probability of having good target properties
- Are in regions of the design space in which there is high uncertainty in the model

Results of the experimental or simulated batch of candidates are added to the platform and used to improve the AI model, which in turn generates the next candidates with better predictive accuracy.

HOW DOES IT REDUCE THE NUMBER OF EXPERIMENTS NEEDED?

In classical DOE without uncertainty calculations, each proposed experiment in the matrix of possible experiments is assumed to be equally valuable in getting to the final result. This is not the case. Some experiments will yield more useful information than others. **Each experiment is an investment**, and by carrying it out we are placing a bet that it will get us closer to our target. The uncertainty calculations provided by the Citrine Platform mean that each bet we are placing is consistently moving us further toward the goal.

WHEN SHOULD YOU USE WHICH METHOD?

	SEQUENTIAL LEARNING	DOE
How many dimensions does your design space have?	Sequential learning is ideal for multi-dimensional problems. The number of experiments required by AI-based Sequential Learning remains linear with the number of dimensions considered.	The number of experiments needed for DOE goes up exponentially with the number of dimensions to optimize.
Is your data varied, complex, and unstructured (e.g. micrographs and outputs from lab equipment)?	The Citrine Platform's data tools are designed to ingest complex materials and chemicals data from different sources.	Most DOE packages are generic and do well with simple, structured, tabular data.
Are you looking for global optimization over the whole design space?	Sequential Learning can survey vast, complex design spaces.	DOE is good at local optimization using linear models.
Do you have data from past projects that could be useful?	Sequential Learning requires a training data set from previous experiments or simulations. However, you can leverage a transfer learning approach to incorporate data from different sources or adjacent projects.	Without any data to create the initial AI model, Sequential Learning won't work and DOE would be better here.
Is there underlying science/domain knowledge that can guide your search?	You can incorporate this into the model on the Citrine Platform and use it to improve model performance and accelerate product development.	DOE is a purely statistical approach and does not use domain knowledge to guide experimental selection.