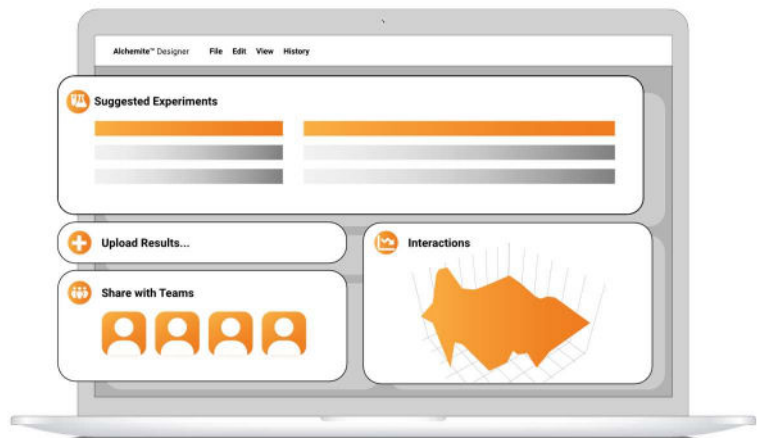


WHITE PAPER

# Why you should combine DOE and machine learning

How ML complements DOE methods such as  
JMP, Minitab, MODDE, and Design-Expert



© 2025 Intellegens Limited  
[intellegens.com](https://www.intellegens.com) | [info@intellegens.com](mailto:info@intellegens.com)  
Intellegens, The Studio, Chesterton Mill, Cambridge, CB4 3NP, UK

 intellegens



## Executive Summary

R&D teams can achieve faster innovation and better outcomes by combining the rigour of **Design of Experiments (DOE)** with the agility of **machine learning (ML)**. While conventional DOE ensures balanced coverage of the experimental space when data is scarce, ML builds on those results and any existing data to focus experiments on high-value targets in real time.

Key benefits of an ML-driven experimental design strategy include:

- **Reduced experiment** through adaptive selection of the most informative tests.
- **Shortened development cycles** by continuously updating recommendations as new data arrive.
- **Maximized ROI on existing tools and data** by importing DOE results and historical experiments into ML models.

Together, DOE and ML form a hybrid approach that lowers cost, accelerates decision-making, and drives more efficient discovery.



## DOE in R&D today

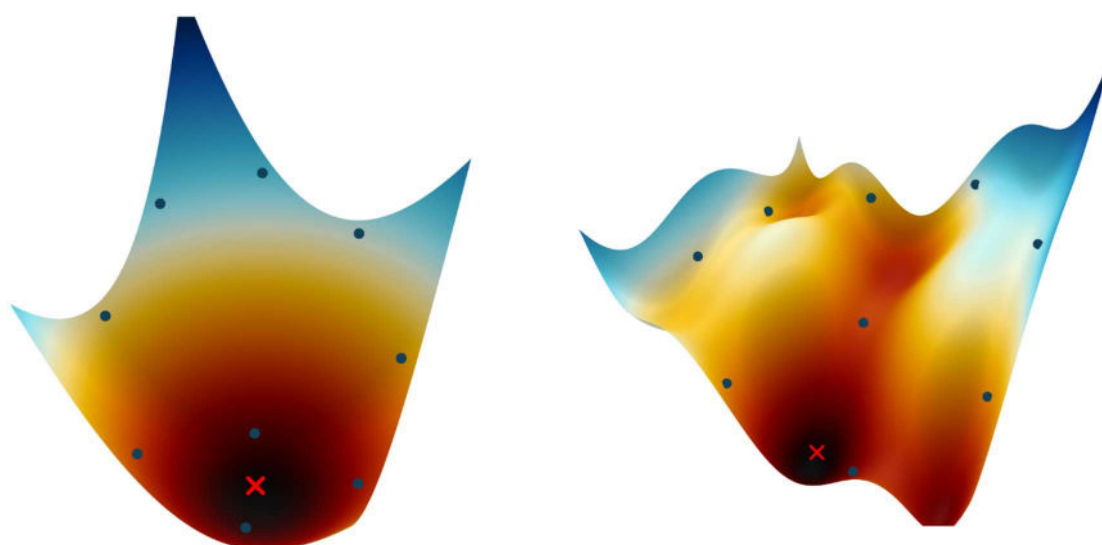
**Design of Experiments** has long been the foundation of structured experimental planning in industrial R&D. By selecting combinations of input factors according to statistical principles, DOE ensures that each experiment contributes to understanding the system under study. Common DOE workflows begin with domain expertise to identify key variables such as formulation components, process parameters, or environmental conditions. These approaches then apply a full factorial, fractional factorial, or response surface design to cover the space efficiently. When fewer assumptions can be made, options like Latin hypercube sampling provide flexible coverage with minimal runs.

Many experimental systems involve constraints, such as limits on mixture composition or total usage of certain ingredients. **Specialized DOE software packages** like [JMP](#), [Minitab](#), [MODDE](#), and [Design-Expert](#) include options like mixture designs to ensure feasibility under such restrictions while maintaining balanced statistical properties. These tools also provide guided interfaces for setting up designs, analyzing results, and generating interpretable polynomial response-surface models. The main strength of DOE lies in its rigour and clarity: researchers can see exactly how inputs affect outputs and build confidence in early-stage results.

However, these **conventional DOE designs** are static, not adapting as more information becomes available. Among other consequences, this means these designs still require many more experiments to cover the design space than could be possible through a more adaptive approach. They also struggle with complex non-linear relationships between variables and require expert knowledge to select the right design type and interpret the outcomes. Finally, such approaches cannot directly leverage unstructured or historical data that were not collected according to the chosen design. Once the initial DOE campaign is complete, teams must often switch methods or tools to take advantage of those additional data sources.

Conventional DOE designs are static, not adapting as more information becomes available

Intellegens



*Fig 1: where relationships between variables are simple and well-understood (left), DOE provides powerful sampling to understand the data. However, it struggles with complex non-linear relationships (right).*

## Enhancing experimental design with ML

**Machine learning** enables a dynamic, data-driven approach to experimental design once some initial data (whether from DOE or previous work) is available. Instead of defining all experiments up front, ML methods such as Bayesian optimization train surrogate models on initial results or legacy data and then suggest new experiments that either reduce uncertainty or improve performance. By balancing exploration of unknown regions with exploitation of high-performing areas, ML reduces wasted runs and achieves project targets more quickly.

At the heart of this approach is the ability of ML models to **learn complex, non-linear relationships** from data. As each new experiment is completed, the model updates its predictions and uncertainty estimates, guiding the next experiment toward regions where

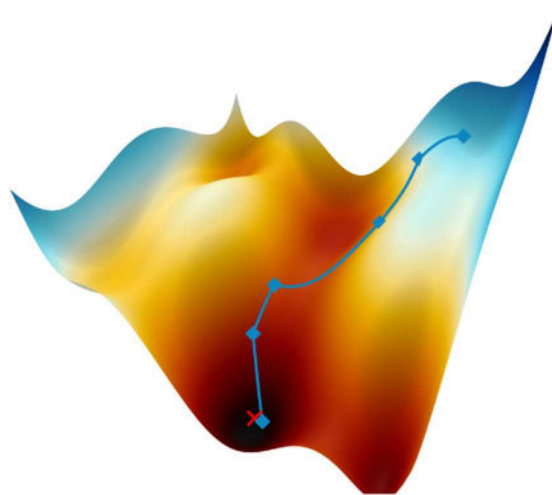


information gain or performance improvement is greatest. This **iterative loop** contrasts with the static nature of classical DOE and allows real-time adaptation as project goals evolve. The result is that targets are typically reached with 50-80% fewer experiments.

ML can also **incorporate all available data**, including unstructured or semi-structured records, so no useful information is left behind. Complex constraints such as ingredient limits or cost caps can be encoded directly into the optimization process, avoiding the need for manual filtering. As a newer approach, many ML-guided optimization tools still rely on coding to get started: [BoTorch](#), [Dragonfly](#), [ProcessOptimizer](#), and many others are Python packages that are free to use, but require a data science background to leverage. Web-based platforms, of which [Alchemite™](#) from Intellegens is an example, offer graphical interfaces that enable domain scientists to define objectives, import data and run optimization loops without writing code.



Of course, ML is subject to the “garbage in, garbage out” problem: its predictive power depends on data quality, representativeness, and adequate sampling. Careful data validation and an initial DOE planning phase help ensure that the model has a reliable foundation.



*Fig 2: ML-driven adaptive experimental design guides experimental projects from current knowledge towards specified project goals.*



## DOE and ML: A combined strategy

Rather than being seen as competing methods, **DOE and ML can serve as complementary phases** of a unified experimental workflow. A three-phase strategy leverages the strengths of each:



## Phase 1 – DOE kick-off

- Select initial experiments via fractional factorial, Box-Behnken, or optimal designs to explore broadly.
- Generate interpretable response-surface models that clarify the main effects and interactions.

## Phase 2 – ML-driven optimization

- Import DOE data and any historical experiments into an ML platform.
- Train a machine learning model and run Bayesian optimization loops, balancing exploration and exploitation.
- Adapt experiment suggestions in real time as the model's uncertainty and performance estimates evolve.

## Phase 3 – Scale-up & knowledge capture

- Validate optimized conditions at pilot or production scale.
- Feed scaled-up results back into ML data repository to accelerate future projects.

The Alchemite™ platform supports both DOE and ML within a single interface, allowing data to flow seamlessly from one phase to the next. Because the initial experiments serve as a starting point for rapid iteration guided by ML, there is less need to optimize this step and thus more scope to move quickly into experiment. This can **remove the need for detailed statistical knowledge and analysis**, for example, to select the right design type.

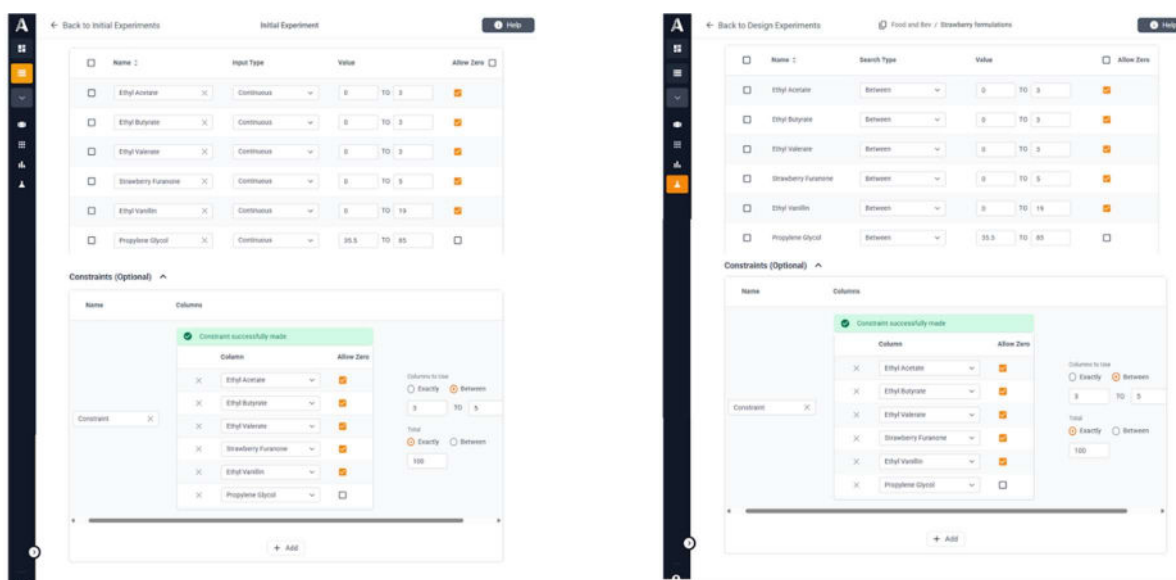


Fig 3: the workflow for creating experimental designs, including complex constraints, is directly comparable between DOE (left) and ML-driven Bayesian Optimization (right), shown here in the Alchemite™ Designer software.



This integrated workflow reduces friction, shortens learning curves, and helps ensure that every experiment builds on the last. Teams may, of course, prefer to retain their existing DOE software for Phase 1 and to bring in an ML tool only to support the later phases of the process. Whether using a combined solution such as Alchemite™ or applying ML alongside conventional DOE software, teams that adopt this hybrid approach will end up running fewer experiments, uncovering deeper insights, and moving more quickly from concept to commercial results.



## Business case for adoption

For R&D leaders, the key question is how to build on existing DOE investments to drive greater efficiency and speed. Introducing ML-driven design alongside current DOE approaches delivers clear business value:

- **Reduce time-to-result** by focusing on the most informative experiments, cutting experimental workload by 50-80%
- **Increase efficiency** by reusing prior data and adapting as new results arrive, lowering material and operator costs
- **Improve decision-making** with predictive models that highlight performance trade-offs and quantify uncertainty
- **Maximize ROI** on existing tools by exporting DOE results directly into ML platforms, avoiding data-migration projects

Introducing  
ML-driven design  
alongside current  
DOE approaches  
delivers clear  
business value

intellegens

Key metrics for tracking success include the percentage reduction in experimental runs, change in cycle time, cost per lead and overall payback period. With low-friction software integrations and user-friendly interfaces, deployment can begin with a high-visibility pilot campaign that demonstrates value in weeks rather than months.



## Conclusions

Design of Experiments and machine learning are not competing methods but **complementary tools** that support different stages of experimental research.

As table 1 (below) illustrates, DOE provides a statistically rigorous starting point when little is known, while ML offers a flexible, goal-oriented strategy once data becomes available.

By combining these approaches, researchers can design smarter experiments, extract deeper insights, and reach project goals more efficiently. As integrated DOE and ML platforms



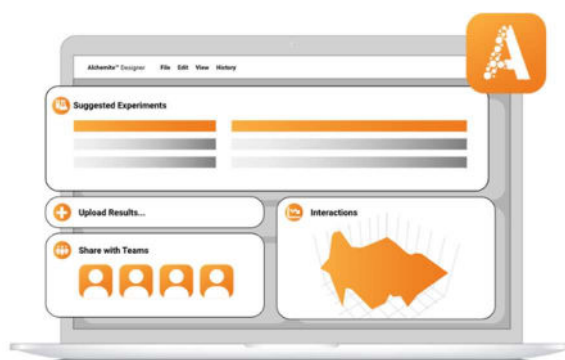
continue to evolve, this combined strategy will become even more accessible, helping R&D teams stay ahead in an increasingly competitive landscape.

Table 1. Pros and cons of conventional and ML-driven DOE. A hybrid approach may be best.

	Conventional DOE (E.g., JMP, Minitab)	ML-driven adaptive DOE (E.g., Alchemite™)
Pros	<ul style="list-style-type: none"><li>• Mature and familiar to many organizations as a means to focus experimental work</li><li>• Rigorous and clear approach</li><li>• Good starting point when little is known</li></ul>	<ul style="list-style-type: none"><li>• Adaptive approach can reduce number of experiments by 50-80%</li><li>• Captures complex, high-dimensional relationships</li><li>• Learns from data - minimal need for statistical knowledge to set up</li><li>• ML model can be analyzed to understand in detail what drives specific properties</li></ul>
Cons	<ul style="list-style-type: none"><li>• Static - doesn't adapt to new data</li><li>• Can still result in many experiments</li><li>• Struggles with non-linear relationships</li><li>• Requires statistical expertise to set up</li><li>• Maps out experiments but offers limited insight into the results</li></ul>	<ul style="list-style-type: none"><li>• Subject to quality, representativeness, and sampling of input data</li><li>• Requires excellent quantification of uncertainty</li></ul>

## Alchemite™ software

The **Alchemite™ Suite** is a range of easy-to-use R&D tools, each focused on a key challenge for R&D managers, scientists, experimentalists, or data scientists. Give the right app to the right team member, speeding and informing their work. Then share results and collaborate across your team, creating an integrated machine learning solution for your R&D organization.



**Alchemite™ Designer** is a simple web browser-based user interface enabling you to create experimental designs, whether you have data and a machine learning model to get started with or not – no lengthy training courses, no need for advanced statistics, no need for coding.

More at [intellegens.com/solutions/designer/](https://intellegens.com/solutions/designer/)



## About Intellegens

Our vision is that machine learning will drive innovation and deliver value wherever data is used in R&D. We aim for best-in-class easy-to-use machine learning software for data analysis in chemicals, materials, life science, and manufacturing. Our Alchemite™ technology originated at the University of Cambridge and development is on-going at Intellegens, in close collaboration with our growing community of customer organizations. These represent sectors including additive manufacturing, aerospace, alloys, batteries, biopharmaceuticals, ceramics, chemical processes, composites, consumer products, cosmetics, drug discovery, energy, food and beverage, formulated products, paints, plastics, and printing technology.

[www.intellegens.com](http://www.intellegens.com) | [info@intellegens.com](mailto:info@intellegens.com)