PREPARING YOUR PRODUCT DEVELOPMENT TEAM FOR AI

THREE MINDSET SHIFTS THAT MOVE MATERIALS AND CHEMICALS PRODUCT DEVELOPMENT FROM PREPARATION TO MEASURABLE VALUE.

START WITH SMALL DATA

GET TO THE LAB QUICKLY

LEARN ABOUT FAILURE

CITRINE INFORMATICS

EXECUTIVE SUMMARY

Many materials and chemicals companies understand the potential value of AI in product development. Their challenge is converting that interest into an experimental operating model that scientists trust and the business can measure.

Programs often delay action for three reasons. Teams wait for a comprehensive, perfectly curated data foundation. They treat model accuracy as a gate that must be passed before new laboratory work begins. They also preserve successful experiments more consistently than unsuccessful ones, leaving the model without the evidence needed to understand feasibility boundaries.

Experience across Citrine's current platform activity, including 380 projects, 550 AI models, and 70,000 experiment suggestions each month, points to a more practical sequence.

1 BEGIN WITH A SMALL, RELEVANT, DIVERSE DATASET

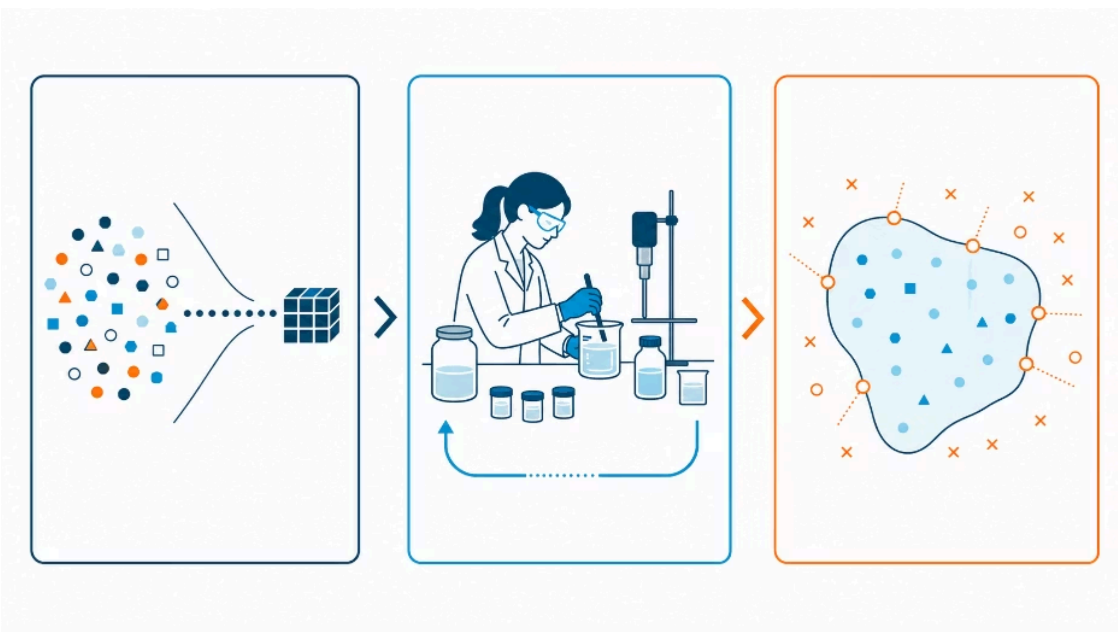
As few as 30 well-chosen data points can be a useful starting point.

2 RETURN TO THE LAB EARLY

Use uncertainty-driven iterative experimentation to improve each experimental round.

3 CAPTURE FAILED EXPERIMENTS

With the same discipline as successful ones so the model learns both performance opportunity and practical constraints.



These shifts change how leaders evaluate progress. The primary question becomes whether the team is reducing the number of iterations required to reach a commercially viable target profile. Model metrics, data architecture, and curation effort remain important, but they serve that product-development outcome.

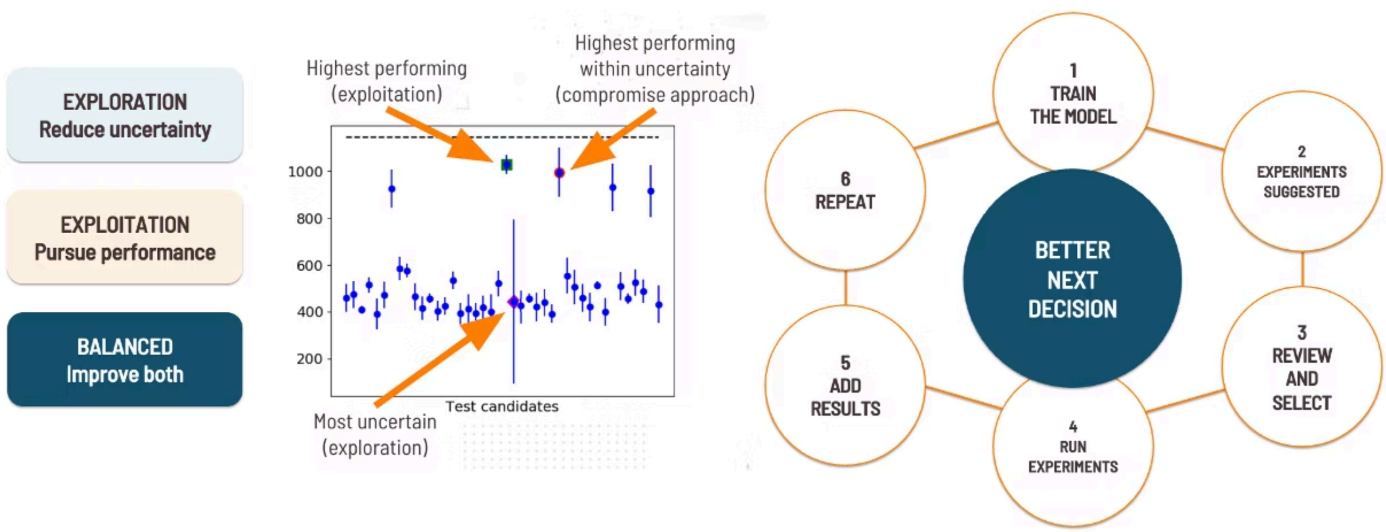
The resulting business case is built on shorter development cycles, fewer wasted experiments, faster customer response, better retention of experimental knowledge, and more focused investment in data. Citrine customers reach first experiment suggestions in a median of 6.5 days and a first technical win in a median of 25 weeks, based on customers since 2021.

THE ADOPTION PROBLEM IS AN EXPERIMENTAL-STRATEGY PROBLEM

AI adoption is often framed as a technology decision. Product-development organizations select a platform and begin a data-readiness program. Those elements matter, but they do not determine whether a scientist's next experiment becomes more informative.

Materials and chemicals product development is constrained by physical experimentation. A formulation or material has to be mixed, processed, and tested. Multiple properties must often be met at the same time. Raw material availability, processing windows, manufacturability, cost, safety requirements, and regulatory pressure shape the feasible solution.

> The economic value of AI emerges through a learning loop. Existing evidence informs a model. The model identifies promising candidates and important uncertainties. Scientists select and run experiments. The results update the evidence base and the model is retrained. Each cycle should improve the next decision.



This iterative learning loop uses model uncertainty to navigate the most efficient path to target values. Use uncertainty to push beyond current property frontiers.

When organizations delay that loop, they delay the first measurable return. The most common delays come from beliefs that are sensible in isolation or when applied to traditional trial-and-error routes but expensive when treated as universal requirements.

MINDSET SHIFT 1:
BEGIN WITH USEFUL DATA

THE COST OF THE COMPREHENSIVE-DATA-FIRST SEQUENCE

A comprehensive data initiative promises consistency across projects and business units. This approach often draws on lessons from data-rich industries such as banking and retail, where records are generated at scale through relatively standardized transactions. Laboratory data has a different economics. Each experiment consumes time, materials, equipment capacity, and scientific effort.

HISTORICAL DATA ISN'T ALWAYS USEFUL

Historical experimental records may span decades, teams, test methods, information systems, and product priorities. Even when each experiment was carefully documented according to the practices and requirements of its time, combining those records can reveal inconsistent terminology, missing contextual information, obsolete methods, and product contexts that no longer match the active development challenge.

THE TWO-YEAR DATA-LAKE TRAP

An organization can spend substantial time harmonizing historical data before it knows whether those records will improve a development decision. A two-year data-lake effort also carries an opportunity cost. During that period, the team may not yet have demonstrated faster reformulation, fewer experiments, faster responses to customer requirements, or a better method for selecting the next set of laboratory experiments.

THE BETTER APPROACH

Materials and chemicals organizations should select AI designed to work with the small, sparse, and heterogeneous experimental datasets they already have. They should not need to recreate the big-data conditions found in other industries before they can begin.

The initial model can extract useful guidance from limited evidence, identify where information is missing, and help the team design new experiments whose conditions and results are consistently documented from the outset.

A PROBLEM-LED STARTING POINT

A stronger sequence begins with one decision that matters to the business. Leaders and scientists define the scope, targets, constraints, and outcome measures before gathering data.

1

FORMULATION OR PROCESS SPACE

Define the scope clearly — what is in and out of bounds for this project.

2

PROPERTY TARGETS & TRADE-OFFS

Specify the performance targets and acceptable trade-offs between competing properties.

3

COMMERCIAL FEASIBILITY CONSTRAINTS

Identify the constraints that determine whether a candidate can be manufactured and sold.

4

LABORATORY RESOURCES

Confirm which experiments can be run with current equipment, materials, and capacity.

As few as **30 well-chosen data points** can provide a practical starting point for an initial model. This number should be treated as a starting benchmark, not a guarantee — the amount of evidence required depends on the dimensionality, noise, chemistry, constraints, and target properties of the project.

> "The thing that I liked about Citrine is you don't need a huge amount of data to get started."
> **— David Cranfill, Technology & Innovation Director, Huntsman Polyurethane**

Using Citrine, Huntsman Building Solutions later reduced starter formulation creation time from a few months to less than one day.

DATA STRATEGY THROUGH APPLICATION

The first model and experiment round make data priorities concrete. They can reveal that a missing process condition explains apparent inconsistency, that one property lacks coverage in the target region, or that a large historical collection adds little because it was generated under different methods.

The organization can then focus curation, standardization, and integration resources on information with demonstrated decision value. A small application becomes a way to shape the wider data strategy with evidence. Data management can be driven by value to the business and occur in parallel to revenue generating product development instead of delaying it.

This matters especially in specialty chemicals, where value often depends on responding rapidly to a customer requirement, achieving a new combination of properties, managing ingredient-cost pressure or supply disruption, or supporting a regulatory substitution program.

AI-GUIDED DATA ACQUISITION

MINDSET SHIFT 2:

JUDGE MODELS BY THE DECISIONS THEY IMPROVE

ACCURACY IS NECESSARY CONTEXT, NOT THE BUSINESS OUTCOME

Development teams are right to question model quality. Poor predictions can waste materials, lab capacity, and scientific effort. The challenge is choosing a measure of model quality that reflects how the model is likely to perform when it encounters new experimental data.

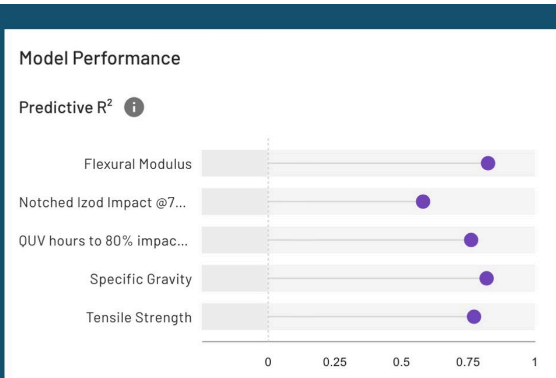
TRADITIONAL GOODNESS-OF-FIT R^2

Measures how closely a model fits the data used to train it. A high value can be reassuring, but it may overstate how well the model will generalize beyond those observations.

PREDICTIVE R^2 (CITRINE PLATFORM)

Estimates how well a model is expected to perform on new, unseen data. Calculated using leave-one-cluster-out cross-validation (LOCO-CV). The model is repeatedly trained while a cluster of related observations is withheld and then evaluated against that withheld evidence — a stricter and more realistic test.

In the Citrine Platform, Predictive R^2 ranges from $-\infty$ to 1. A value close to 1 indicates strong generalization, while a positive value indicates that the model has begun to capture useful predictive relationships. Because Predictive R^2 tests performance on withheld data, it will often appear lower than a conventional goodness-of-fit R^2 calculated from the training dataset.

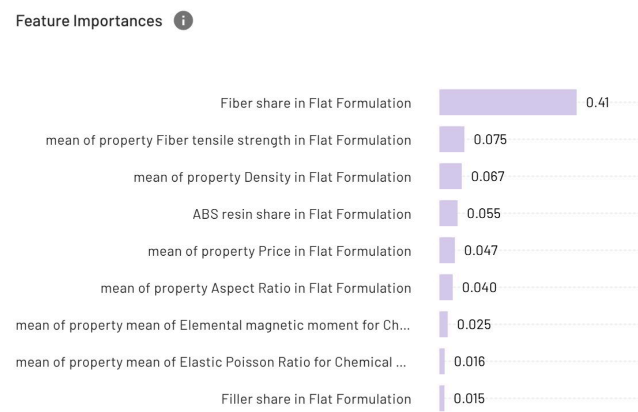


We advise customers to run experiments once the Predictive R^2 is above 0

Citrine generally recommends beginning the next experimental round once **Predictive R^2 is above 0**. This is a practical indication that the model has captured enough signal to guide a useful group of experiments — not a claim that every prediction will be accurate.

HOW ELSE CAN I ASSESS THE USEFULNESS OF THE MODEL?

Feature importance plots in the platform show which raw materials, processing parameters, or molecular features the model is relying on most to make predictions. Product experts can check whether the model is consistent with their scientific intuition. This is another way to validate the usefulness of the model.



WHAT THE MODEL MUST DO AT PREDICTIVE $R^2 > 0.3$

IDENTIFY PROMISING CANDIDATES

Surface candidates within the project constraints that are most likely to meet performance targets.

SHOW HIGH UNCERTAINTY

Reveal where its uncertainty is high so scientists can design experiments that reduce it.

REVEAL TRADE-OFFS

Expose potentially valuable trade-offs between competing properties or constraints.

BALANCE EXPLORATION

Balance near-term performance with exploration of less-tested regions of the design space.

SELECT INFORMATIVE EXPERIMENTS

Choose experiments that will improve the evidence available for the next decision round.

ITERATIVE AI-DRIVEN EXPERIMENTAL LEARNING

The initial round may not produce a formulation that meets every target. Its value also lies in the information it creates. An exploratory experiment can eliminate an infeasible region, expose a missing constraint, identify a key property driver, or clarify a difficult trade-off. When those results are incorporated into the model, subsequent recommendations become better informed and more focused.

> A Citrine user reported reaching all target properties after **three rounds of sequential learning** while also capturing precise knowledge about how different components affected final properties. The result was both a target-hitting formulation and a stronger basis for future product-development decisions.

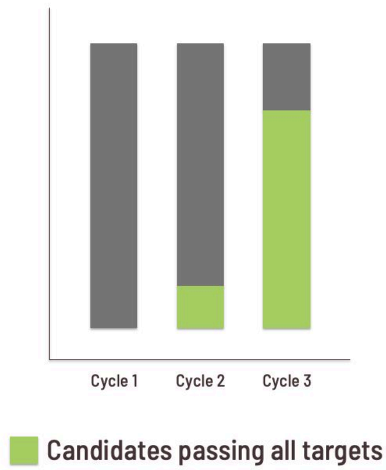
Real example formulation program

9 properties to optimize

6 treated as constraints

74 historical formulations (none of which passed all targets)

Experiments run in batches of 10



MEASURE LEARNING ACROSS THE FULL LOOP

Leaders should retain appropriate technical model review while also measuring whether the iterative learning workflow is improving product development performance.

- Predictive R^2 before each experimental round
- Calendar time from project framing to first experiment run
- Number of experimental rounds to reach target profile
- Total experiments vs. previous workflow
- Percentage of recommended experiments that are feasible
- Knowledge captured about key drivers, trade-offs, and failure boundaries

These measures connect model development to the outcomes that matter to the organization: product development throughput, lab capacity, customer responsiveness, and time to value. The purpose of the model is not to achieve the highest possible accuracy score in isolation. It is to make each experimental round more informative and move the team toward a successful product with fewer iterations.

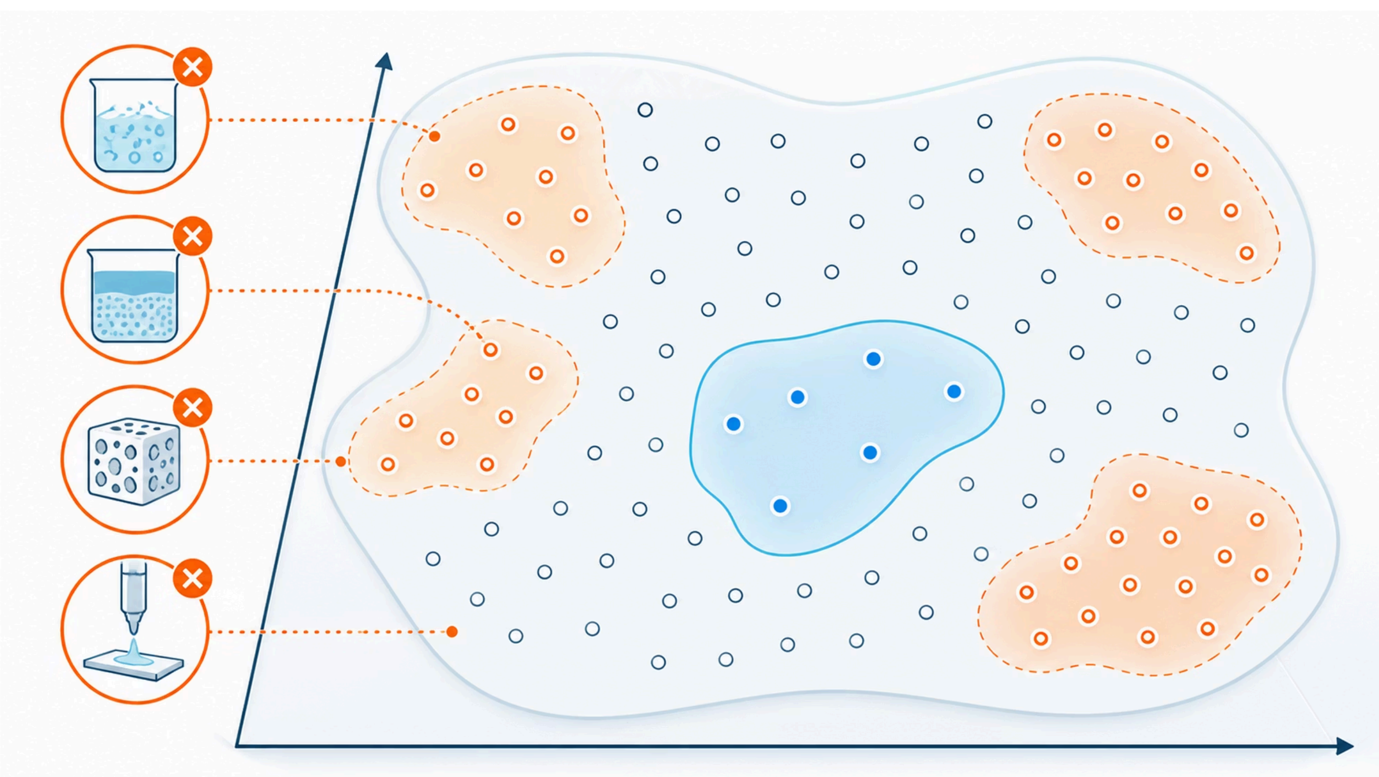
MINDSET SHIFT 3: TREAT FAILURE AS BOUNDARY INFORMATION

SUCCESS-ONLY DATASETS CREATE BLIND SPOTS

Successful formulations naturally receive the most attention. They progress into reports, presentations, product specifications, and databases. Unsuccessful outcomes may be recorded in systems that are less accessible to other teams or are not structured for reuse in future product-development work.

This creates a blind spot for AI. If a model sees mostly successful outcomes, the feasible region can appear larger and safer than it is. The model cannot learn the practical significance of phase separation, porosity, impurities, mechanical failure, excessive viscosity, unstable aging, infeasible processing, waste, or prohibitive cost unless those outcomes are represented in the data.

> Capturing failure makes future recommendations more realistic. Evidence of infeasible combinations allows the model to avoid known failure regions and recommend experiments that clarify uncertain boundaries.



ACCELERATE PRODUCT QUALIFICATION

AI can compress the qualification timeline by predicting the outcomes of expensive, time-consuming end-stage tests using data from earlier, lower-cost experiments. Qualification tests that are run externally, cost a great deal or take a long time such as stability, flammability and animal toxicity tests can be targeted. Teams that build this capability reduce the number of confirmatory tests required, avoid late-stage failures, and bring products to market faster. By surfacing likely failures earlier, this approach also de-risks the product development pipeline — reducing the cost and disruption of late-stage surprises.

CASE STUDY 1

HUNTSMAN BUILDING SOLUTIONS

Huntsman used AI to predict the outcomes of large-scale flammability tests using data collected from bench-top experiments. By identifying which formulations were likely to pass or fail before committing to costly full-scale testing, the team was able to prioritize candidates more effectively and reduce the number of large-scale tests required.

Citrine Informatics

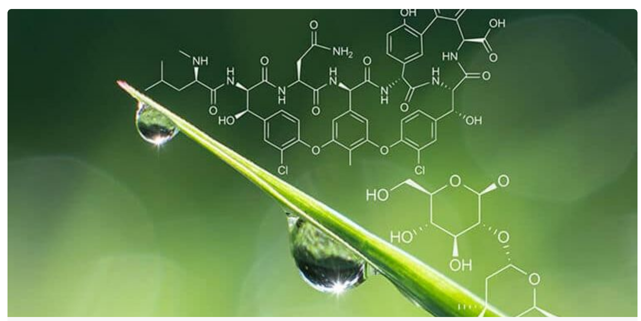
Revolutionizing The Use Of Fire Tests

Find out how Huntsman Building Solutions reduces reliance on full-scale fire tests...

CASE STUDY 2

AGROCHEMICAL PRODUCER

An agrochemical producer applied AI to predict toxicity outcomes, reducing the number of end-stage animal tests needed during product qualification. By training models on earlier-stage experimental data, the team could screen out high-risk candidates earlier in the development process, improving both efficiency and animal welfare outcomes.



Citrine Informatics

Screen Out Toxic Formulations

Learn how the Citrine Platform was used to predict toxicity in complex, multi-phase...

> In both cases, the AI model did not replace the qualification test — it changed which candidates reached it. Fewer tests, better candidates, and faster decisions are the compounding benefit of predictive qualification.

DOCUMENTATION DETERMINES WHETHER FAILURE BECOMES AN ASSET

A generic "failed" label has limited learning value. Whether a failure occurs in the laboratory, during scale-up, or in production, teams should capture structured information that supports scientific comparability and model learning.

OBSERVABLE FAILURE MODE

The measurement or criterion used to classify the failure.

FORMULATION CONTEXT

The formulation and raw-material context at the time of failure.

PROCESSING CONDITIONS

The processing conditions and operating window applied.

TEST METHOD

The test method and measurement conditions used.

EQUIPMENT ANOMALIES

Relevant equipment or execution anomalies that may have contributed.

STAGE OF FAILURE

The stage at which the failure occurred — from laboratory development through production.

When development and production data use compatible definitions, a failure observed at manufacturing scale becomes reusable product-development knowledge. Scientists can incorporate that evidence into subsequent models and design new formulations or processes that avoid the same limitation.

A BROADER DEFINITION OF EXPERIMENTAL SUCCESS

An experiment creates value when it improves the product, improves the manufacturing process, or improves the next decision. This definition does not excuse poor preparation or uncontrolled work. It recognizes well-designed experiments that test an important uncertainty, expose a feasibility limit, or reveal a scale-up risk.

FOR LEADERS, THIS REQUIRES BOTH A CULTURAL AND AN OPERATIONAL CHANGE

Scientists should not be penalized for a controlled experiment that identifies a genuine failure boundary. Production teams should have a clear route for returning relevant outcomes to product development. Project reviews should ask:

- What did the result reveal about technical performance?
- What did it reveal about manufacturability?
- Does the development model represent this failure mode?
- What changed in the model or the feasible design space?
- How should the next experimental round differ?

The objective is not simply to find a formulation that works once in the laboratory. It is to develop a product that meets its technical targets and can be manufactured reliably at commercial scale. By connecting development and production evidence, AI can help teams address both objectives earlier and accelerate scale-up.

THE RETURN-ON-INVESTMENT LOGIC

These three mindset shifts prepare product-development teams to work effectively with AI. Leadership must also demonstrate how the new experimental workflow creates measurable business value. Tracking that value is essential for sustaining adoption, justifying continued investment, and deciding where AI-guided experimentation should be scaled next.

The relevant return is not captured by model performance alone. It appears in faster decisions, fewer experiments, more productive use of lab capacity, quicker scale-up, and improved customer responsiveness. A practical ROI model includes five value pools.

1. EXPERIMENTAL EFFICIENCY

Estimate the reduction in experiments and external tests required to reach the target profile. Apply the fully loaded cost of materials, equipment time, technician time, scientist time, and outsourced testing.

Avoided experimental cost = experiments avoided × fully loaded cost per experiment

2. DEVELOPMENT-CYCLE VALUE

Estimate the commercial value of reaching a viable formulation earlier — earlier customer qualification, faster response to a tender, reduced delay under a raw material disruption, or earlier entry into a new market.

Time value = months gained × expected monthly contribution associated with the opportunity

3. RAW MATERIAL AND PROCESSING-COST IMPROVEMENT

AI-guided formulation work can explore technical performance, energy usage (e.g., processing temperatures), and raw material cost together. One consumer goods company reduced their raw material costs by 20% by reformulating to use locally sourced ingredients.

4. RESILIENCE AND RESPONSE VALUE

Quantify the exposure associated with a supplier loss, price spike, or need for ingredient substitution. The value may appear as avoided downtime, protected revenue, reduced qualification delay, or maintenance of a multiple-supplier strategy.

5. KNOWLEDGE REUSE

Estimate the time senior scientists spend recreating prior work, helping new team members find starter formulations, or resolving repeated failures. Structured experimental evidence and models can preserve relationships between raw materials, processes, properties, and failure modes for future projects.

6. MARGIN IMPROVEMENT

AI prompts product experts to try novel experiments and pushes technical performance beyond the status quo. By developing more differentiated products, companies can justify larger margins.

1600%

CONFIRMED ROI

Citrine has a confirmed example of 1,600% ROI in the first year. Each organization should build its own finance-reviewed model using project-specific assumptions.

6.5

DAYS TO FIRST SUGGESTIONS

Citrine customers reach first experiment suggestions in a median of 6.5 days (based on customers since 2021).

25

WEEKS TO FIRST WIN

Customers achieve a first technical win in a median of 25 weeks (based on customers since 2021).

A LEADERSHIP AGENDA FOR THE FIRST THREE PROJECTS

CHOOSE LIGHTHOUSE PROJECTS, NOT TECHNOLOGY SHOWCASES

The first AI projects carry significance beyond their immediate technical objectives. Their success or failure will inevitably become a bellwether for how the wider organization judges AI. A poorly chosen project can undermine confidence even when its difficulties arise from project selection rather than the technology itself.

Successful early projects create the opposite effect. They provide visible evidence that AI improves day-to-day product development, justify continued investment, and give other teams a credible example to follow. The scientists behind these projects can become internal advocates and mentors, helping colleagues adopt the new workflow more quickly. In this sense, the first projects and teams become lighthouses for wider adoption.

As AI becomes normalized across the chemicals industry, competitive advantage will not come simply from using it. Advantage will come from using AI effectively and scaling that capability across projects, teams, and business units. The first three projects should establish the foundation for that scale.

Each initial project should have four characteristics:

1

A COMMERCIALLY VALUABLE OUTCOME

The project should address an objective that matters to the business — customer responsiveness, margin, supply resilience, technical performance, or regulatory substitution. A successful outcome should be valuable enough to justify continued investment.

2

A RESPECTED AND ENTHUSIASTIC PRODUCT DEVELOPER

The team should include a credible scientist who is respected by peers, open to experimentation, and motivated to try a new way of working. Their experience will influence how other scientists perceive the value and practicality of AI.

3

A VIABLE STARTING DATASET

The project should begin with approximately 30 relevant, diverse data points. The data do not need to be perfect, but they should provide enough coverage for an initial model and a useful first set of experimental recommendations.

4

SHORT, WELL-SUPPORTED EXPERIMENTAL CYCLES

Experiments and testing should be completed quickly enough to sustain momentum. The team should have ring-fenced laboratory capacity and resources to act on recommendations and begin the next learning cycle without avoidable delays.

The first projects should not be low-value pet projects that cannot demonstrate meaningful business impact. Nor should they be exceptionally complex problems that experienced teams have repeatedly failed to solve, projects in which testing takes several months, or initiatives imposed on teams that have little interest in adopting a new workflow.

Organizations can progress to more complex, slower, and higher-risk applications once scientists and leaders have seen the day-to-day value of AI-guided experimentation. Early lighthouse projects establish confidence, working practices, and internal expertise, making that broader expansion faster and more effective.

PROTECTING THE LEARNING LOOP & PLANNING FOR SCALE

PROTECT THE LEARNING LOOP

Scientists need time to frame the problem, learn the workflow, run the experiments, and interpret results. David Thomas, Director of Research Development at Syensqo, has described the importance of making space for accomplished scientists to learn a new skill and rewarding them for doing so.

ESTABLISH A MINIMUM DATA CONTRACT

Define the conditions, units, methods, outcome categories, and failure details that every new experiment will capture. Focus the initial standard on the active project, then extend it based on demonstrated reuse.

REVIEW DECISIONS AND EVIDENCE TOGETHER

Governance should examine both technical and business progress: what did the model recommend, why was that experimental group selected, what uncertainty was reduced, which target or failure boundary became clearer, what will change in the next round, and how many experiments and how much time has the workflow required.

PLAN FOR SCALE AFTER PROOF

A focused success can create the evidence needed for wider adoption. The enterprise question then becomes how to repeat the workflow across scientists, projects, business units, and geographies while preserving scientific ownership.

180

USERS

Citrine's largest single deployment has 180 AI users — not just data entry and analysis.

34

PROJECTS

Saint-Gobain onboarded 34 projects across 22 global entities in 18 months.

FROM LAB TO LAUNCH, FASTER.

AI creates value in materials and chemicals product development when it changes the quality of experimental decisions. That change can begin before an enterprise data foundation is complete and before a model is globally predictive across every value.

START SMALL

Begin with a small, relevant, diverse body of evidence — as few as 30 data points.

LEARN FAST

Use the model's predictions and uncertainty to return to the lab early and iterate.

CAPTURE EVERYTHING

Capture successful and unsuccessful outcomes with equal scientific discipline.

Each round makes the next round more informed. Over time, the team spends less effort on low-information experiments, builds a clearer map of feasible performance, and creates reusable experimental knowledge.

> Citrine Informatics is the proven AI platform for materials and chemistry product development. Ten of the top 20 specialty chemical companies have trusted Citrine. Customers have created hundreds of millions in documented value using the platform.

Contact Citrine Informatics to learn more from our experience supporting AI-guided product-development programs.

Request a meeting

CITRINE.IO