CITRINE INFORMATICS

WHITE PAPER:

TRUSTING THE EXPERIMENT

HOW AI ADOPTION, MODEL UNCERTAINTY, AND TRANSPARENT WORKFLOWS CAN IMPROVE MATERIALS AND CHEMICALS PRODUCT DEVELOPMENT

EXECUTIVE SUMMARY

Materials and chemicals companies have spent the last decade proving that AI can improve product development. AI is already being used to guide formulation, prioritize experiments, improve material performance, reduce cost, and respond faster to customer, regulatory, and supply chain pressure.

The next challenge is adoption.

Many product development teams already believe AI can help. The harder problem is getting more scientists to use AI consistently, understand its recommendations, and trust those recommendations enough to act. This is especially important when AI suggests an experiment that looks unintuitive, carries uncertainty, or does not resemble the safest variation on what worked before.

The full value of AI is not realized when a model simply predicts a better candidate. It is realized when scientists can use AI to make better experimental decisions, including decisions that help the model learn.

This white paper argues that the next phase of AI in materials development depends on five shifts:

- **MOVING FROM ISOLATED AI PROJECTS TO EVERYDAY SCIENTIFIC WORKFLOWS.**
- **HELPING SCIENTISTS UNDERSTAND PREDICTION UNCERTAINTY, NOT JUST PREDICTED PERFORMANCE.**
- **RECOGNIZING THAT SOME FAILED EXPERIMENTS ARE PRODUCTIVE BECAUSE THEY TEACH THE MODEL.**
- **CAPTURING AND REUSING KNOWLEDGE ACROSS TEAMS, DOCUMENTS, DATASETS, AND EXPERTS.**
- **COMBINING STRONGER MODELING ARCHITECTURES WITH TRANSPARENT WORKFLOWS THAT KEEP SCIENTIFIC JUDGMENT CENTRAL.**

Catalyst, Citrine's AI workflow accelerator, and Apex, Citrine's new suite of advanced modeling architectures, represent this shift. Catalyst lowers the barrier to creating and acting on AI workflows. Apex raises the quality of the models behind them. Together, they help materials and chemicals teams move from product goals to better experimental decisions faster.

1. PRODUCT DEVELOPMENT TEAMS ARE BEING ASKED TO SOLVE HARDER PROBLEMS FASTER

Materials and chemicals product development has always required trade-off management. Teams balance performance, manufacturability, cost, raw material availability, customer requirements, regulatory pressure, and sustainability goals.

Those trade-offs are becoming harder.

A polymer may need to maintain mechanical performance while reducing density or cost. A coating may need to improve durability while reducing a restricted ingredient. A lubricant may need to maintain performance after a base oil becomes unavailable. A specialty formulation may need to meet multiple customer-critical properties while using different suppliers or lower-cost inputs.

The target keeps moving. Raw materials become unavailable or prohibitively expensive. Regulations change. Customers ask for more specific performance. Experienced scientists retire. Lab capacity remains limited.

AI can help teams explore more options and make better decisions under these constraints. But broad adoption requires AI to fit into the daily work of product development teams, not sit apart from it.

2. THE NEXT BARRIER IS ADOPTION

The early question for AI in materials development was whether models could make useful predictions. That question has been answered in many industrial settings.

The next barrier is whether more scientists can use AI consistently.

Adoption is not just a training issue. It is a trust issue.

Scientists are accountable for the experiments they run. A model recommendation can consume scarce lab time, testing capacity, raw materials, and weeks of attention. Product experts need to understand why an experiment is being suggested before they commit resources to it.

This is especially true when a recommendation is unintuitive.

AI can search more broadly than a human expert can. It can identify combinations of ingredients, process conditions, and trade-offs that may not be obvious from prior experience. That is one of its greatest strengths. It is also one of the reasons scientists may hesitate to act.

Experienced product experts are often risk averse for practical reasons. They know what has worked before. They understand which experiments are easy or hard to run. They may have limited access to the equipment required to make samples. They may be under pressure to show progress quickly.

As a result, teams may favor safer variations on previous successes.

That instinct is understandable, but it can limit the value of AI. If organizations only run the recommendations that already look familiar, they may never get the full benefit of exploring the design space.

The goal is not to persuade scientists to ignore their judgment. The goal is to give them enough context to decide when a less obvious experiment is worth testing.

3. INNOVATION REQUIRES MORE THAN SAFER VARIATIONS

Companies often adopt AI because they want to boost innovation. They want to evaluate more options, discover unexpected paths to performance, and move beyond the limits of intuition and manual experimentation.

AI is well suited to that challenge because it can explore a large design space systematically.

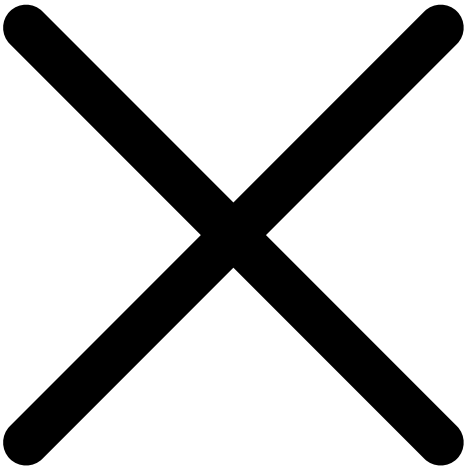
In formulation, that design space might include many possible ingredient combinations, ratios, process settings, and performance targets. In materials discovery, it might include molecular structures, synthesis conditions, processing variables, and property trade-offs.

Human experts bring essential judgment. They know which constraints matter, which recommendations are practical, and which trade-offs will be acceptable to customers or manufacturing teams. But human intuition is shaped by previous experience. That experience is valuable, but it can also pull teams toward familiar regions of the design space.

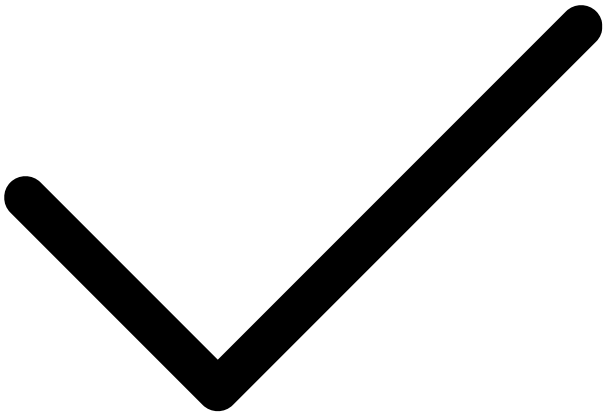
AI can push teams to consider options they might not otherwise test.

That is where trust becomes essential. If a recommendation looks unusual, the scientist needs to know why it was proposed. They need to see whether the model is confident in the prediction, uncertain but curious, or balancing competing objectives in a way that deserves closer review.

Without that context, unintuitive recommendations are easy to reject.



With that context, they become part of a disciplined exploration strategy.



4. MACHINE LEARNING HAS A DIFFERENT RELATIONSHIP WITH UNCERTAINTY

Traditional design of experiments can be powerful, particularly when the design space is well understood and the relevant variables are known. It helps teams plan experiments systematically and evaluate factor effects.

Machine learning adds a different advantage: it can estimate uncertainty.

A model does not only predict which candidates may perform well. It can also identify where its own understanding is weak and where new data could improve future predictions.

This changes the logic of experiment selection.

Models often suggest two broad types of experiments.

The first type are high-confidence candidates. These have strong predicted properties and relatively low uncertainty. They are attractive because the model expects them to perform well and has more confidence in the prediction.

The second type are learning candidates. These may have higher uncertainty, but they sit in regions of the design space that could contain strong performance or important information. Some of these experiments will fail to hit the target properties. But they can still teach the model something valuable.

For product development teams, this distinction is critical.

If every experiment is judged only by whether it immediately hits all targets, teams may undervalue experiments that reduce uncertainty and improve later recommendations. In an active learning workflow, some experiments are selected because they are likely to teach the model, not because they are guaranteed to succeed immediately.

Models are a little like PhD students. They need to be taught, but they learn quickly when given the right experiments. Early failed experiments can help the model understand boundaries, reduce uncertainty, and suggest better candidates in the next round.

The fastest-looking route may be to run only the safest candidates. The faster route overall may include a few experiments that miss the target but accelerate learning.

5. A FORMULATION EXAMPLE: LEARNING FROM THE FIRST ROUND

A SPECIALTY CHEMICAL COMPANY WAS TRYING TO OPTIMIZE NINE PROPERTIES IN A FORMULATION.

The team had already tested 70+ similar formulations, but none had hit all nine targets. They trained a model on that historical data and ran a batch of 10 AI-recommended experiments.

None of the first 10 passed all the targets.

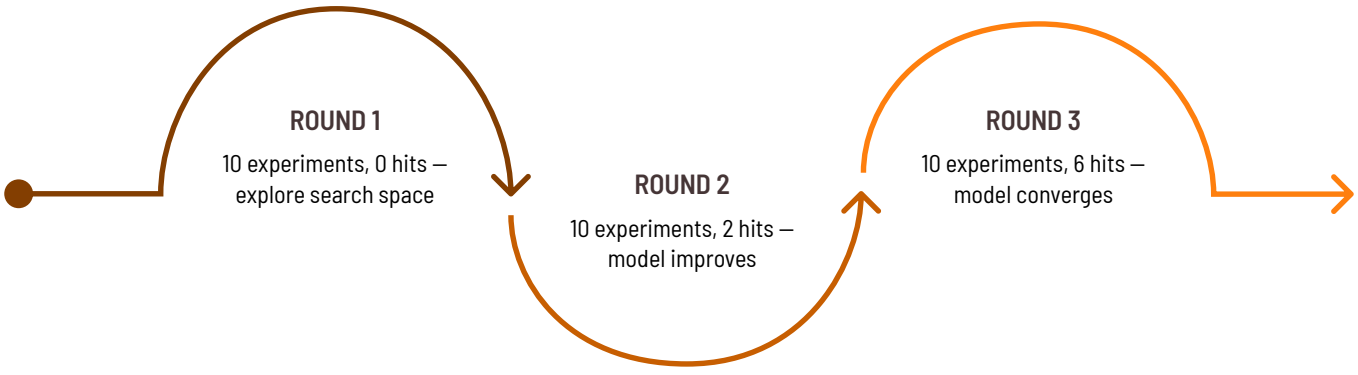
That outcome can be uncomfortable for researchers. On the surface, it can look like the model failed. But those 10 experiments gave the model new information about the design space.

The team retrained the model using the additional data. In the next batch of 10 recommended experiments, two suggested formulations hit all nine targets. In a third round, six suggested formulations hit all targets.

The first round did not produce the winning formulation. It produced learning.

That learning changed the trajectory of the project.

This example illustrates why AI adoption is as much about scientific confidence as model performance. If the team had abandoned the model after the first 10 experiments, they would not have reached the later rounds. They needed to understand that early misses could still be useful because they reduced uncertainty and improved the next set of recommendations.



6. BETTER WORKFLOWS MAKE BETTER DECISIONS POSSIBLE

A model recommendation is only one part of an experimental decision.

A scientist still needs to know:

- Which data was used
- Which properties were optimized
- Which constraints were applied
- What the model is uncertain about
- Why one experiment was suggested over another
- Whether the recommendation is practical in the lab
- Whether the trade-offs are acceptable for the product and customer

This is where workflow design matters.

AI systems used in materials development need to show their work. They need to make the workflow inspectable and editable. They need to help scientists distinguish between high-confidence recommendations and higher-uncertainty learning experiments.

Natural-language interaction can help, but only if it is connected to the actual product development workflow.

The value of a natural-language prompt is not novelty. The value is reducing the distance between scientific intent and AI-guided action.

A product expert should be able to describe a goal in ordinary language, then review the dataset, model, search space, target properties, and suggested next experiments created from that goal. The scientist should be able to adjust assumptions, apply product judgment, and decide what to run.

That is how AI becomes more usable without reducing the role of expertise.

7. FROM PRODUCT GOAL TO NEXT EXPERIMENT

Consider a scientist working on a lubricant reformulation.

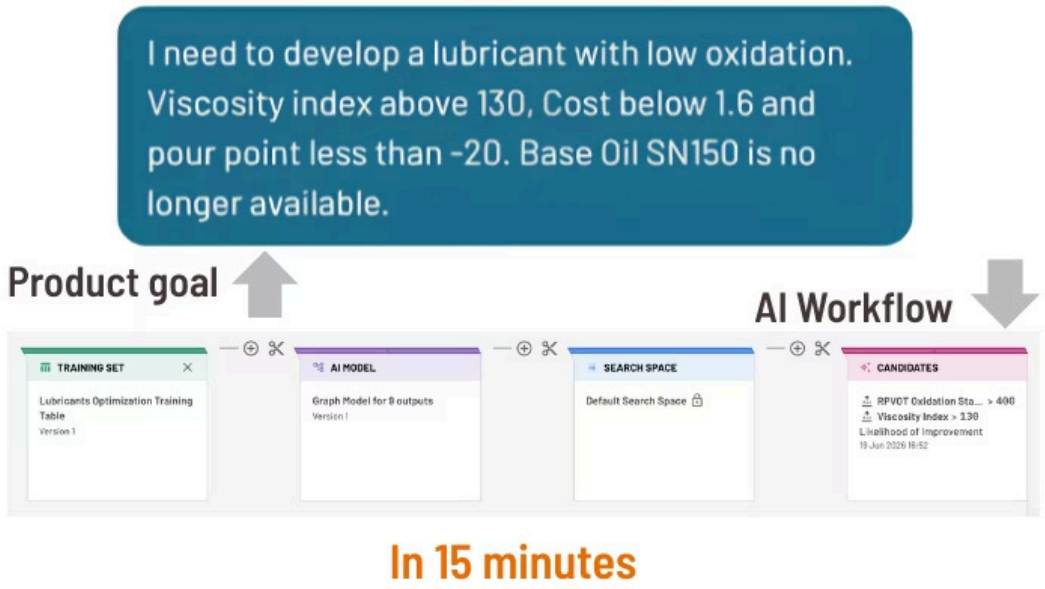
The product needs low oxidation, a viscosity index above a defined threshold, a cost below a target level, and a pour point below a specified limit. A base oil is no longer available.

Traditionally, turning that product goal into an AI workflow would involve multiple steps: identifying relevant historical data, preparing a dataset, setting target properties, defining the search space, training or selecting a model, generating candidates, and reviewing suggestions.

Each step is manageable. Together, they create friction.

That friction matters because product teams are already busy. If the path from goal to workflow takes too long, AI becomes something teams use occasionally rather than a standard part of product development.

Catalyst, Citrine's AI workflow accelerator, is designed to shorten that path.



A scientist describes the product goal. Catalyst reviews available data, creates the dataset, builds the AI model, sets the search space and target properties, and suggests next experiments in minutes. It also provides a rationale for the workflow and the recommendations.

The scientist remains in control. They review the workflow, adjust it, and decide which experiments are worth running.

8. STRONGER MODELS RAISE THE CEILING

Lowering the barrier to AI adoption does not reduce the need for strong models. It increases the importance of model quality.

Materials and chemicals problems vary widely. A high-dimensional formulation challenge may require a different model type than a molecular design problem, a process-property relationship, or a tabular prediction task. No single model architecture is best for every project.

This is where Apex, Citrine's new suite of advanced modeling architectures, comes in.

Apex evaluates different model approaches, including graph neural networks, transformers, and other advanced architectures, then returns results from the model that performs best for the project.

It also uses pre-trained chemical embeddings based on scientific literature and molecular data, giving models a stronger chemistry-aware starting point.

That matters because materials problems often depend on relationships among composition, processing, structure, and properties. A stronger modeling layer can help teams prioritize experiments more effectively and learn faster from each round.

In internal Citrine benchmarks, raw Apex models, before tuning or added expert knowledge, performed as well as or better than expertly tuned best-in-class models 95% of the time.

On complex, high-dimensional optimization problems, raw Apex models get to target properties up to **twice as fast** as the current best-in-class models.

The practical value is not the architecture itself. The practical value is better experimental prioritization.

X2

FASTER

Product development

95%

BETTER

Without expert tuning

9. CROSS-TEAM LEARNING IS WHERE PRODUCTIVITY COMPOUNDS

AI adoption also depends on the organization's ability to reuse knowledge.

Many materials and chemicals companies already have valuable information spread across datasets, technical reports, formulation notes, technical data sheets, safety data sheets, spreadsheets, and historical project records. They also have expert judgment that lives with experienced scientists.

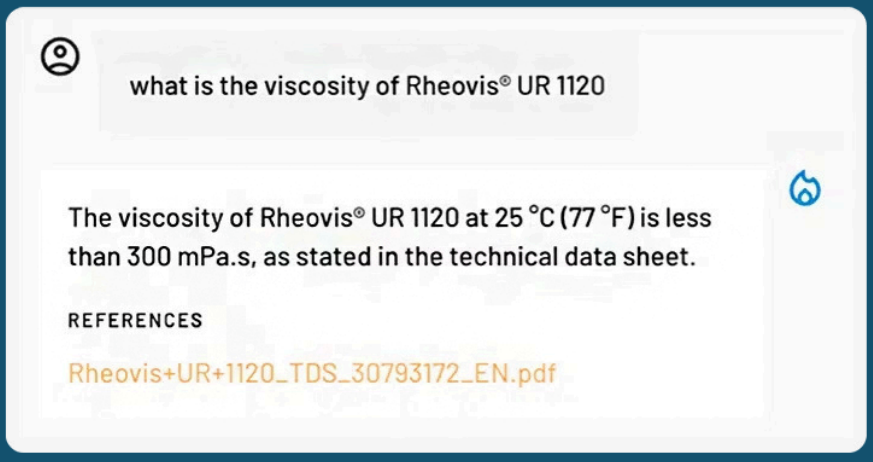
When that knowledge is hard to find, teams repeat work. They miss relevant prior experiments. Newer scientists take longer to ramp up. Senior experts spend time answering questions that the organization has already answered before.

AI can help, but only if knowledge is connected to the workflow.

Knowledge assistant can help scientists search approved documents and literature, returning answers with references. Model assistant can help product experts add scientific relationships to models through plain-language interaction. Workflow accelerators can bring prior data, modeling, search spaces, and candidate recommendations together faster.

Together, these capabilities help organizations move from isolated expertise to cross-team learning.

This is especially valuable as experienced scientists retire. AI will not replace their judgment, but it can help capture and reuse more of what they know.



QUERY YOUR TEAM'S DATA WITH NATURAL LANGUAGE.

10. SCIENTIFIC JUDGMENT STAYS CENTRAL

The future of AI in materials development is not autonomous decision-making that removes scientists from the process.

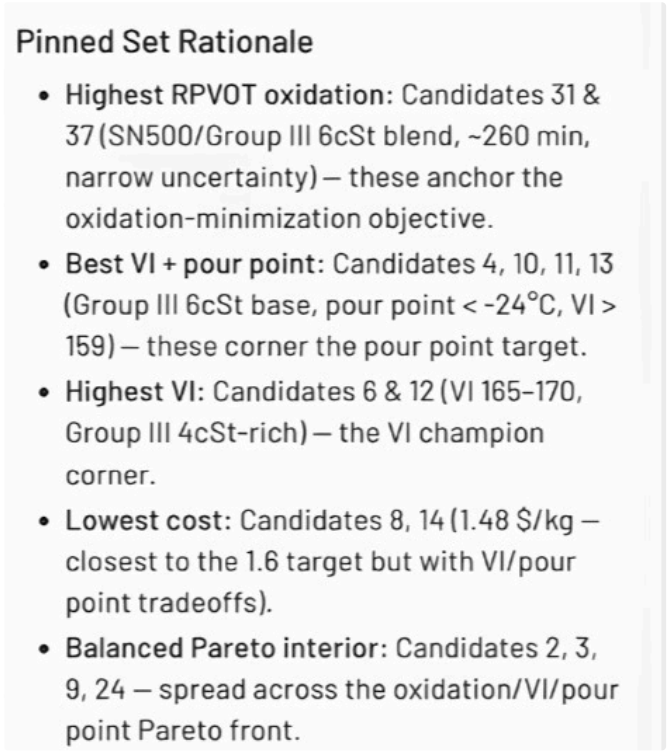
The more useful direction is AI that gives scientists a stronger starting point.

AI can assemble relevant data, suggest candidate experiments, identify likely drivers of performance, and highlight trade-offs. Scientists then decide what to trust, what to change, and what to run.

This distinction matters for adoption.

Scientists are more likely to use AI when they can see how the workflow was built and why recommendations were made. They need to understand the model's reasoning, but they also need the ability to intervene.

Strong AI workflows do not ask scientists to surrender judgment. They give scientists a better basis for exercising it.



Pinned Set Rationale

- **Highest RPVOT oxidation:** Candidates 31 & 37 (SN500/Group III 6cSt blend, ~260 min, narrow uncertainty) — these anchor the oxidation-minimization objective.
- **Best VI + pour point:** Candidates 4, 10, 11, 13 (Group III 6cSt base, pour point < -24°C, VI > 159) — these corner the pour point target.
- **Highest VI:** Candidates 6 & 12 (VI 165–170, Group III 4cSt-rich) — the VI champion corner.
- **Lowest cost:** Candidates 8, 14 (1.48 $/kg — closest to the 1.6 target but with VI/pour point tradeoffs).
- **Balanced Pareto interior:** Candidates 2, 3, 9, 24 — spread across the oxidation/VI/pour point Pareto front.

CATALYST SUGGESTS EXPERIMENTS AND EXPLAINS WHY IT RECOMMENDS THEM.

11. RECOMMENDATIONS FOR MATERIALS AND CHEMICALS TEAMS

Organizations seeking broader AI adoption in product development should focus on five practical priorities.

1

TREAT UNCERTAINTY AS USEFUL INFORMATION

A high-uncertainty recommendation is not automatically a bad recommendation. It may be a learning opportunity. Teams should distinguish between experiments chosen to produce near-term candidates and experiments chosen to improve the model.

2

MAKE THE WORKFLOW INSPECTABLE

Scientists need to see which data, targets, constraints, and assumptions shaped the recommendation. A workflow that cannot be inspected will be harder to trust.

3

PRESERVE EXPERT CONTROL

AI should support scientific decision-making, not replace it. Product experts need the ability to adjust assumptions, apply judgment, and decide what to run.

4

REUSE PRIOR KNOWLEDGE

Historical data, technical reports, formulation notes, and expert knowledge should feed future decisions. Organizations that reuse knowledge across teams will learn faster.

5

MEASURE PROGRESS ACROSS ROUNDS, NOT ONLY EXPERIMENT BY EXPERIMENT

Some early experiments may fail to hit all targets but still improve the model. Teams should evaluate whether each round improves learning and accelerates the path to target properties.

12. THE NEXT PHASE OF MATERIALS AI

AI in materials and chemicals has moved beyond the question of whether models can make useful predictions. The next phase is broader adoption.

That requires stronger models, but it also requires workflows scientists can create, understand, and trust. It requires systems that can surface prior knowledge, explain recommendations, capture expert judgment, and shorten the path from product goal to experimental action.

It also requires a cultural shift. Teams need to become comfortable with experiments that teach the model, even when those experiments do not immediately hit every target.

Catalyst and Apex represent that shift.

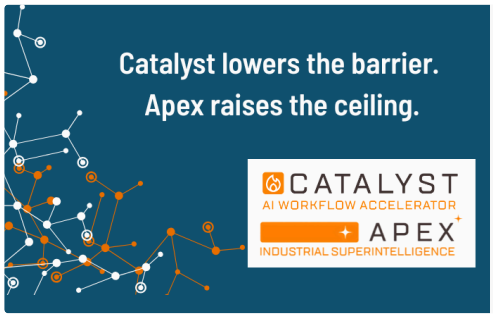
Catalyst lowers the barrier to AI workflows. Apex raises the quality of the models behind them. Together, they help materials and chemicals teams move from product goals to better experimental decisions faster.

The companies that benefit most from AI will not simply be the ones with the most advanced models. They will be the ones that make those models usable by more scientists, across more projects, with more confidence in the next experiment.

ABOUT CITRINE

Citrine Informatics is the proven AI platform for materials and chemistry product development. Citrine customers use the platform to accelerate product development, optimize processes, and deliver results across complex product development workflows. The Citrine Platform is trusted by 10 of the top 20 specialty chemical companies and has generated hundreds of millions in documented customer value.

From lab to launch, faster.



info.citrine.io

From Lab to Launch Faster

Meet Apex and Catalyst. A step change in Materials and Chemicals AI and how to interact with it.