

Why

chemistry-native AI

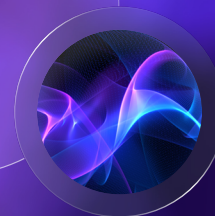
is

45x more
accurate

than

general-purpose AI

AI is on the agenda of every R&D leadership meeting across the chemical industry.



The rationale makes sense: decades of experimental data, siloed insights across different teams, institutional knowledge that walks out the door every time someone moves on. If any technology can help surface and apply that knowledge faster, it should be AI.

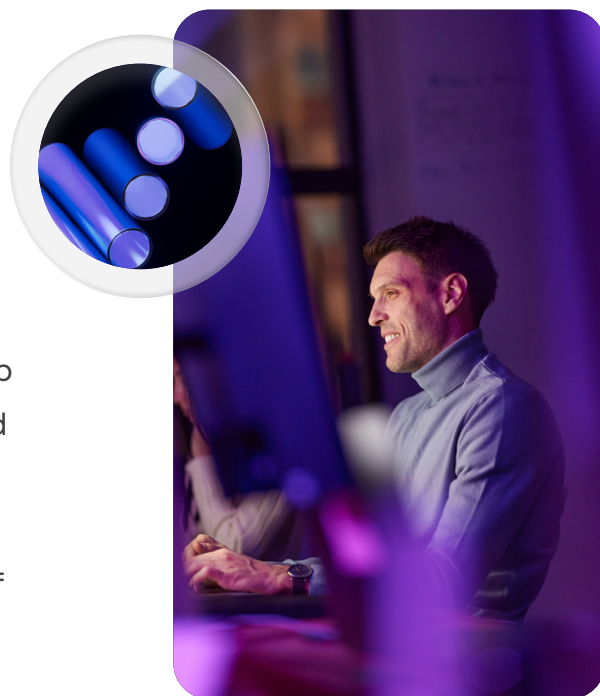
Many organizations have already made the investment, especially with so many commercial, general-purpose tools available that promise natural language search across their document libraries. Early results look encouraging, but in practice, these solutions break down.

Sometimes, chemists discover the gaps quickly and stop using the tool after a few hallucinations. Other times, it takes scale to expose the problem, where the tool that works reasonably well on a few hundred documents starts returning plausible but wrong answers once the corpus grows to the enterprise scale of thousands, or tens of thousands, of documents. The results feel close, but in a complex R&D environment where success may hinge on a tenth of a percent composition change, complete accuracy is a requirement.

The pressure to show ROI from AI is real, which only heightens the frustration when value doesn't materialize. If you've been in that position, it's worth understanding the root cause: **the tools your team deployed are working as designed, but they weren't designed for chemistry.**

General models can classify, but can't discriminate

Think about the senior scientists in your R&D organization. They have years of bench work, deep familiarity with their domain, and the ability to read a passage from a patent or historical report and immediately understand whether it can answer the question at hand, even when it shares most of its vocabulary with a dozen other documents that aren't relevant for the situation.



General-purpose AI retrieval models, systems designed to retrieve information in response to a question, can't discern this kind of nuance. They were built for breadth, perform well across an enormous range of tasks, and are good at domain classification, like identifying that a document is about coatings formulation versus personal care formulation. What they cannot do, however, is discriminate within a domain, telling apart two passages that share the same vocabulary and the same material categories, but that answer different questions.

For chemistry knowledge retrieval, that distinction is everything.

Consider this example, drawn from a real query run against a real corpus of documents. A formulation chemist asked: "What hydrophobic compounds are entrapped in wax seeded latex polymers to improve scuff resistance in architectural coatings?" Two passages were candidates, and both mentioned paraffins, silicones, coatings, and hydrophobic chemistry, and sat in the same patent classification. Even the ontology, the functional roles each ingredient plays and the application context each sits in, matched up to two levels.

“What hydrophobic compounds are entrapped in wax seeded latex polymers to improve scuff resistance in architectural coatings?”



The answer

US11230645B2 • Scuff resistant and chip resistant architectural compositions (2022, CPC C09D)

“...wax... in combination with a hydrophobic aqueous latex resin produces paint films that resist scuffing and chipping... a hybrid latex polymer, wherein the latex comprises a hydrophobic or wax center entrapped by a polymeric shell, hereinafter ‘wax seeded latex’... The hydrophobic compounds can be **silicones**, silicone modified alkyds, low molecular weight polyethylene, low molecular weight polypropylene and **paraffins**”

ONTOLOGY Coatings & materials → hydrophobic compound system → wax core entrapped in polymer shell → scratch resistant architectural film



What the general model retrieved

US11230654B2 • Calcium carbonate coated materials (2022, CPC A47L | C09K | C11D)

“...the base fluid includes an oil-based fluid, such as for example an oleaginous fluid... Examples of oleaginous fluids suitable for use in a WSF include, but are not limited to, petroleum oils, natural oils, synthetically-derived oils... diesel oil, kerosene oil, mineral oil... polyolefins, **paraffins**, **silicone fluids**, polydiorganosiloxanes, oxygenated solvents, esters...”

ONTOLOGY Coatings & materials → hydrophobic compound system → **oleaginous carrier fluid** → **coated calcium carbonate for servicing fluids**

Every surface signal matches:

paraffins

silicones

coatings

hydrophobic

CPC C09

The correct passage described wax seeded latex polymers specifically, with a hydrophobic compound entrapped in a polymeric shell designed to resist scuffing. The other described oil-based carrier fluids for calcium carbonate coatings, a completely different application that shared surface-level vocabulary. **A leading general-purpose retrieval model ranked the correct answer beyond position 300 out of roughly one million passages, while placing an incorrect passage at the top.** In practice, this ranking is a significant miss, because retrieval systems only hand over their top 100 or so results to the next step in the pipeline, whether that is a re-ranker or the AI model writing the answer—anything ranked in the hundreds never gets seen.

More telling, the model assigned the correct and incorrect passages nearly identical similarity scores. This means it not only failed to surface the correct answer to the formulation question, it also could not distinguish between two passages that a coatings chemist would tell apart immediately. Relative to the query, its top recommendation and a passage buried hundreds of positions down looked essentially the same to it.

When organizations bolt a general-purpose AI tool to their R&D knowledge, the gap between classification and nuanced discrimination is what causes pilots to fall short.

Why the chemical industry needs chemistry-native AI

Albert's AI team exposed this failure mode through extensive benchmarking and developed a custom retrieval model trained on chemistry to power Ask Albert, chemistry-native AI built for enterprise R&D. **In a recent study, the team found that Albert's chemistry-native retrieval model was 45x more accurate than general-purpose retrieval models at finding the correct answers to scientific queries.**

Benchmarking and scoring

The study was built on a foundation of 590,000+ US formulation patents spanning the chemical industry, including pharmaceuticals, adhesives, coatings, agrochemicals, food, and personal care. From this corpus, the team constructed two distinct test subsets.

The first used easy, cross-domain distractors: given a query about automotive adhesives with a known passage where the correct answer can be found, the competing passages came from completely different chemistry domains. This tests whether a model can distinguish between different fields of chemistry, which is a relatively straightforward task.

The second used hard, within-domain distractors. Here, competing passages came from the same chemical domain as the correct answer, sharing vocabulary, material categories, and patent classifications, but differing in functionally important ways, such as in the architectural coatings example discussed earlier. This is the test that probes a retrieval model's ability to identify nuanced differences between similar documents with the expertise of a senior chemist.

Study

590k+

US formulation patents

55k+

expert-validated queries

1M

passage corpus

The team generated **55,000+ queries written in the style of a formulation chemist**, each with a known correct passage where the answer lives, validated by domain experts. **Models were then scored on whether they returned the correct passage in their top 100 results out of a corpus of roughly one million passages.** This is an important metric because passages surfaced within the top 100 results reach the next step in the pipeline. A score of 100% would mean the model surfaces the correct passage in its top 100 results every single time.

How chemistry-native AI performs against general-purpose tools

2-4.5%

of the time, general-purpose models found the correct passage in the top 100 results

On the easy subset, leading general-purpose retrieval models, including models from OpenAI and Voyage, performed reasonably well. On the hard subset, however, **general-purpose retrieval models only surfaced the correct passage in the top 100 results between 2% and 4.5% of the time.** These models recognize that a passage is about formulation chemistry, but cannot tell what kind of formulation it describes.

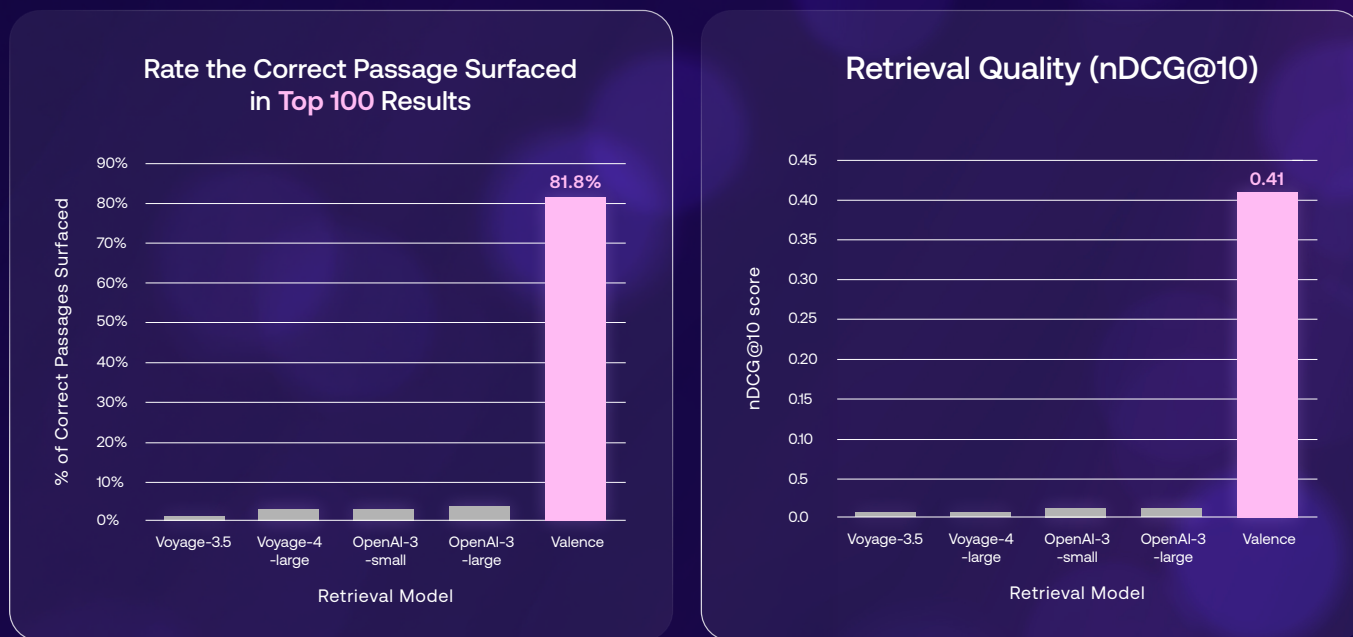
81.8%

of the time, Valence found the correct passage in the top 100 results

On the same hard subset, **Albert's chemistry-native retrieval model, Valence, surfaced the correct passage in its top 100 results 81.8% of the time.** Because Valence was trained on tens of thousands of near-miss examples, it learned what a formulation chemist would use to tell two similar passages apart, and that signal is embedded in how it represents chemical text.

On a stricter test that scores a model only on whether it puts the correct passage near the top of its first ten results, Valence scored 0.41 against 0.009 for the leading general-purpose retrieval model, roughly a **45x improvement.** Analysis of the results revealed that the general-purpose models compress all within-domain passages into a narrow band of similarity scores, leaving almost no difference between the passage ranked first and the one ranked 100th. Meanwhile...

Valence achieves discrimination to the level of chemical composition and function, the way an experienced formulation chemist would.



Higher is better. Scores a model on whether the correct passage appears in its first ten results and how near the top it ranks.

The performance of general-purpose retrieval models also worsened as corpus size grew, mimicking the scale of retrieval required for enterprise R&D. At small scale, the general-purpose retrieval models performed adequately, but as the corpus approached one million passages, scores dropped sharply, in some cases by more than 80%. This is the pattern that explains why AI tools often seem to work in early pilots and degrade in practice. **On the other hand, Valence's performance held as corpus size scaled toward one million passages, validating that its accuracy is unaffected at enterprise scale with large knowledge bases.**

Developing Valence required building a purpose-designed corpus from hundreds of thousands of patents, generating and validating tens of thousands of chemist-style queries, constructing a training regime capable of teaching a model to see what only a domain expert would catch, and running everything at scale. This critical yet resource-intensive work is an advantage of working alongside a digital transformation partner, allowing your internal AI team to direct their energy on other unique challenges.

AI trained on your chemistry

Valence is trained on chemistry, giving it a strong foundation that is useful across the chemical industry from day one. The next level is tuning it to a specific company's proprietary knowledge base: the same training paradigm, applied to your documents, your chemistry, and your hard negatives to create a retrieval model that is tailored to the nuances of your product development. In other words, **you will have a custom retrieval model trained on your chemistry.**

Just like a senior formulation chemist develops intuition specific to their domain and their organization over their career, Valence can develop that same kind of specificity from your data, all while maintaining enterprise-grade security and zero data retention.

Valence is one layer of a system built for chemistry at every level. **Ask Albert's reasoning engine also understands how data objects captured in a lab relate to each other and the way a chemist thinks about them.** The underlying data ontology from Albert OS structurally represents R&D, not just documents. And every response from Ask Albert comes with a reasoning audit trail, so your chemists know exactly where an answer is sourced from.

This is what chemistry-native AI looks like end to end: not a general-purpose tool pointed at chemistry data, but a system built foundationally around chemistry.

If you want to see how Ask Albert performs on your knowledge base, and what customer-specific tuning could look like for your chemistry, we'd like to show you.

Talk to our team

Albert

albertinvent.com