

The modular AI tech stack

How to build for a
world where models
change every year

In most insurance organizations today, AI didn't arrive through a strategy. It arrived through urgency.

Does this sound familiar?

●
An underwriting team starts using a general-purpose model to summarize submissions.

●
A claims manager experiments with a tool to extract data from loss reports.

●
A data engineer builds a small pipeline to structure PDFs for downstream analysis.

None of these efforts are coordinated, and none were designed to work together. They solve immediate problems in isolation, often without a shared understanding of how they fit into the broader system.

For a while, this works. Reviews move faster, teams feel more productive, and leadership sees early signals that AI is delivering value.

Then, the cracks start to show.

This ebook explores what's actually breaking in today's AI workflows, and how to build a modular AI stack that can [keep up with ongoing model change](#).



What **breaks** when AI is layered onto existing workflows

Consider a mid-sized carrier processing commercial property submissions. A broker sends in a 300-page submission with schedules of values, loss runs, inspection reports, and supplemental documents.

Historically, reviewing a file like this could take days. Now, underwriters rely on a mix of tools to speed things up.

One underwriter pastes sections into a chatbot to summarize risk details. Another uses a separate tool to extract tables from PDFs. A third relies on an internal script built by IT to capture key fields. Each approach works in isolation, but none of them are connected.

The outputs vary, the formats differ, and none of the results tie cleanly into the policy administration system.

At first, the gains feel meaningful. But when the same submission is reviewed by two different underwriters, inconsistencies emerge. One tool misses exclusions buried in the document. Another mislabels a location value. The internal script fails on rotated pages.

When discrepancies appear, there is no reliable way to trace the data's origin or determine which version is correct.



Too many disconnected AI tools

Teams use different tools across underwriting, claims, and operations, with no coordination or shared logic.



No shared data foundation

Without a common structure or system of record, data can't flow cleanly across systems or teams.



Inconsistent document inputs

Submissions arrive in different formats and are handled inconsistently, creating issues before processing even begins.



Manual reconciliation between systems

Teams spend time resolving discrepancies instead of moving workflows forward.



Conflicting outputs from the same document

Different tools extract and interpret the same data differently, leading to mismatched results.



Limited traceability in decision-making

When outputs conflict, there's no clear way to trace how a decision was generated or which source is correct.

The consequences start subtly but compound quickly. Teams delay quotes as they double-check outputs. Underwriters lose confidence in the tools and revert to manual review for critical decisions. In some cases, incorrect data enters pricing models, introducing risk that surfaces later.

The issue isn't that AI fails. It's that most AI isn't designed for the workflows it's being asked to support.



The **accumulation problem** no one plans for

As more teams adopt AI independently, the problem doesn't just repeat. It compounds.

What starts as isolated experimentation in underwriting or claims spreads across the organization. Each team builds its own workflows, selects its own tools, and defines its own way of working. Over time, those decisions stack up, creating a system that no one intentionally designed.

IT is expected to support this environment, but has limited visibility into how teams use these tools or how they

interact. There is no shared standard for data structure, no consistent way to validate outputs, and no clear ownership of decision-making across systems.

This is where "death by a thousand agents" becomes real: [dozens of disconnected AI tools](#), each solving a narrow task but adding complexity to the system as a whole.

At that point, the issue is no longer about individual workflows. It's now a system-level problem.

Scenario	What teams expect	What actually happens
Adding a new AI tool	Faster processing	Another disconnected workflow
Automating a step	Less manual work	More validation and rework
Scaling usage	Consistency across teams	Conflicting outputs across workflows
Underwriting decisions	Consistent risk evaluation across submissions	Different tools interpret the same submission differently, leading to inconsistent quotes
Claims validation	Faster, more accurate claims processing	Data mismatches and gaps require manual review, slowing decisions
Auditing decisions	Clear traceability	No clear data lineage or defensible decision path

Where fragmentation becomes a business risk

As AI becomes embedded in core workflows, the risk shifts from efficiency to accountability.

When an auditor or compliance team asks how a decision was made, your team can't give a clear answer. The data exists, but no one can trace how it moved through the system or how the final outcome was determined.

This breakdown becomes more visible in complex scenarios. In a claim involving multiple documents, jurisdictions, and policy conditions, teams extract data in one system, enrich it in another, and pass it manually between groups.

Each step produces outputs, but no system captures how those outputs connect or how the final decision was reached.

At that point, the cost extends beyond inefficiency. Without clear traceability or consistency, decisions are harder to defend, and each new tool adds complexity rather than reducing it.

What began as innovation starts to resemble fragmentation.



Why the promise of “end-to-end AI” doesn’t hold up

Faced with this complexity, many organizations look for a way to simplify. The most appealing answer is also the most ambitious: a single system that can handle everything from ingestion to decision-making.

The pitch is compelling. Replace fragmented workflows with an end-to-end AI platform that processes documents, extracts data, applies business logic, and updates downstream systems automatically.

In controlled environments, this works. But real insurance workflows are not controlled environments.

Take the example of a large carrier evaluating an end-to-end AI platform for underwriting. In a typical product demo, the workflow appears clean and linear: the system ingests submissions, extracts data, evaluates risk, and generates quotes. Everything looks seamless.

Once deployed, the variability of real operations emerges. Different lines of business evaluate risk differently. Regional teams follow distinct guidelines. Certain decisions depend on portfolio exposure and market conditions, not just document content.

The workflow is not linear, and it is rarely uniform.

What “end-to-end AI” promises vs. what insurance workflows actually require

Promise	Reality
One system handles everything	Workflows vary by line of business, region, and team
Fully automated decisions	Human judgment remains central
Seamless implementation	Heavy customization and rigid workflows
Reduced complexity	New constraints and workarounds

As a result, the end-to-end system begins to strain under real-world conditions. Teams start working around the system rather than through it, and the promise of simplicity gives way to a different kind of rigidity.

In practice, most organizations end up back where they started, with multiple systems and no clear way to unify them.



The real risk isn't choosing the wrong tool; it's getting **locked** into one

Most conversations about AI focus on which model performs best. Accuracy, speed, and cost dominate the discussion.

The problem emerges later, once a system becomes embedded in workflows.

Consider a carrier that adopts a document processing platform that becomes core to underwriting operations. It feeds structured data into downstream systems and shapes how decisions get made.

Then a better model becomes available.

Switching should be straightforward. It isn't.

If the system is tightly coupled to the original model, replacing it means reworking integrations, revalidating outputs, and retraining teams. What should be a simple upgrade becomes a disruptive project.

At that point, the organization faces a tradeoff: absorb the cost of change, or continue with a system it knows is no longer optimal. Many delay, not because the value isn't clear, but because switching is too expensive.

Over time, that decision locks the system in place.

Where a **modular approach** changes insurance operations

A modular approach changes the system by separating concerns instead of forcing everything through a single platform. Rather than asking one system to ingest documents, structure data, apply logic, and drive decisions, each layer has a defined role.

Where to start when building a modular AI tech stack for insurance



Start with document ingestion and parsing as a separate, stable layer

Normalize submissions, loss runs, and policy documents before they enter underwriting or claims workflows. coordination or shared logic.



Standardize how data is structured before applying models

Ensure key fields like locations, coverage limits, and claim details are consistently formatted across documents.



Treat models as interchangeable components, not core infrastructure

Apply models to specific tasks like extracting submission data or summarizing claim reports without tying them to the entire workflow.



Keep business logic outside the model layer

Preserve underwriting guidelines, risk rules, and claims validation logic outside of the model so they remain consistent and auditable.



Ensure downstream systems depend on structured data, not model outputs

Feed policy admin, pricing, and claims systems with validated data rather than raw model responses.



This shift is not about adding more components. It's about redefining where responsibility sits.

In most organizations today, models are asked to do too much. They ingest documents, interpret context, structure data, and support decisions all at once. [A modular approach breaks that apart.](#) Documents are processed consistently, data is structured once, and models operate on clean inputs without being embedded into the workflow itself.

When a better model becomes available, it can be introduced at the appropriate layer without forcing the entire system to be rebuilt.



What a modular AI tech stack looks like in a **real** workflow

A claim file arrives with incident reports, medical records, and policy documents.

In a fragmented system, each document might be processed differently. In a rigid system, everything is forced through the same pipeline, regardless of fit.

A modular system takes a more deliberate approach.

Where AI fits across insurance workflows

Function	How AI is applied
Distribution	Lead qualification, marketing automation
Underwriting	Submission triage, risk summarization
Service	Customer support automation
Claims	FNOL triage, damage assessment

These applications sit at the workflow level. But to make them reliable at scale, they need a consistent foundation underneath.



A **simplified** modular AI tech stack for insurance workflows

- 1 Documents enter (PDFs, emails, submissions)** Broker submissions, loss runs, medical records, and supporting documents enter the workflow in inconsistent formats.
- 2 Parsing layer standardizes format** Documents are normalized so tables, layouts, and file types can be processed consistently.
- 3 Extraction layer structures data** Key fields like locations, coverage limits, claim details, and exposures are extracted into a usable format.
- 4 Models operate on structured inputs** Models handle tasks like risk summarization, claims triage, or document classification without needing raw documents.
- 5 Structured data feeds downstream systems** Policy admin systems, claims platforms, data warehouses, and analytics tools consume consistent data.
- 6 Rating engines and actuarial models evaluate risk and pricing** Structured data feeds underwriting models, pricing engines, and risk scoring systems.
- 7 Outputs support underwriting and claims decisions** Underwriters quote policies, and claims teams validate and adjudicate based on consistent, structured inputs

The key shift is not adding more components, but redefining where responsibility sits. Each layer can evolve independently without breaking the rest of the system.

The system improves incrementally, rather than requiring periodic reinvention. In practice, this allows teams to remove bottlenecks without rebuilding their entire workflow.

- [See how Amwins streamlined underwriting and reclaimed hours per quote with document-centric AI →](#)



What a modular AI approach means for your organization

If you're leading data or technology within an insurance organization, your teams are already using AI. The decision to adopt it has effectively been made.

Most organizations start by adopting AI and then try to reshape their workflows around it. That approach creates more complexity than it solves. AI should adapt to how your

business operates, not the other way around.

A modular AI stack starts with understanding how work actually flows through your organization and where flexibility matters most.

From there, the focus shifts to a more practical question: can your system absorb change without breaking?

How to tell if your AI architecture will scale

As AI adoption accelerates, the real risk isn't what you've implemented, it's how well it will hold up over time. Use these questions to assess whether your current approach is built to adapt or already showing signs of strain.

Can we swap models without rebuilding workflows?

Do all teams operate on the same structured data?

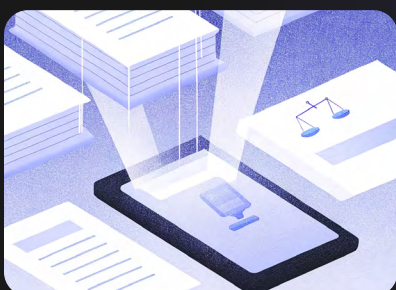
Can we trace how a decision was made?

Are we adding tools faster than we can govern them?

Do our systems reflect how work actually happens?

Because the real test will not come from a single decision. It will come from a series of them. Each new model, each new tool, each new use case will introduce pressure.

The organizations that navigate constant model change successfully won't be those that picked the right model early. They'll be the ones who built systems where adopting the next model is straightforward.



The 90-day path to underwriting reinvention

See how Fortune 500 companies eliminate the bottleneck where **70% of submissions** arrive incomplete.

<https://arc.net/l/quote/mabadeva>