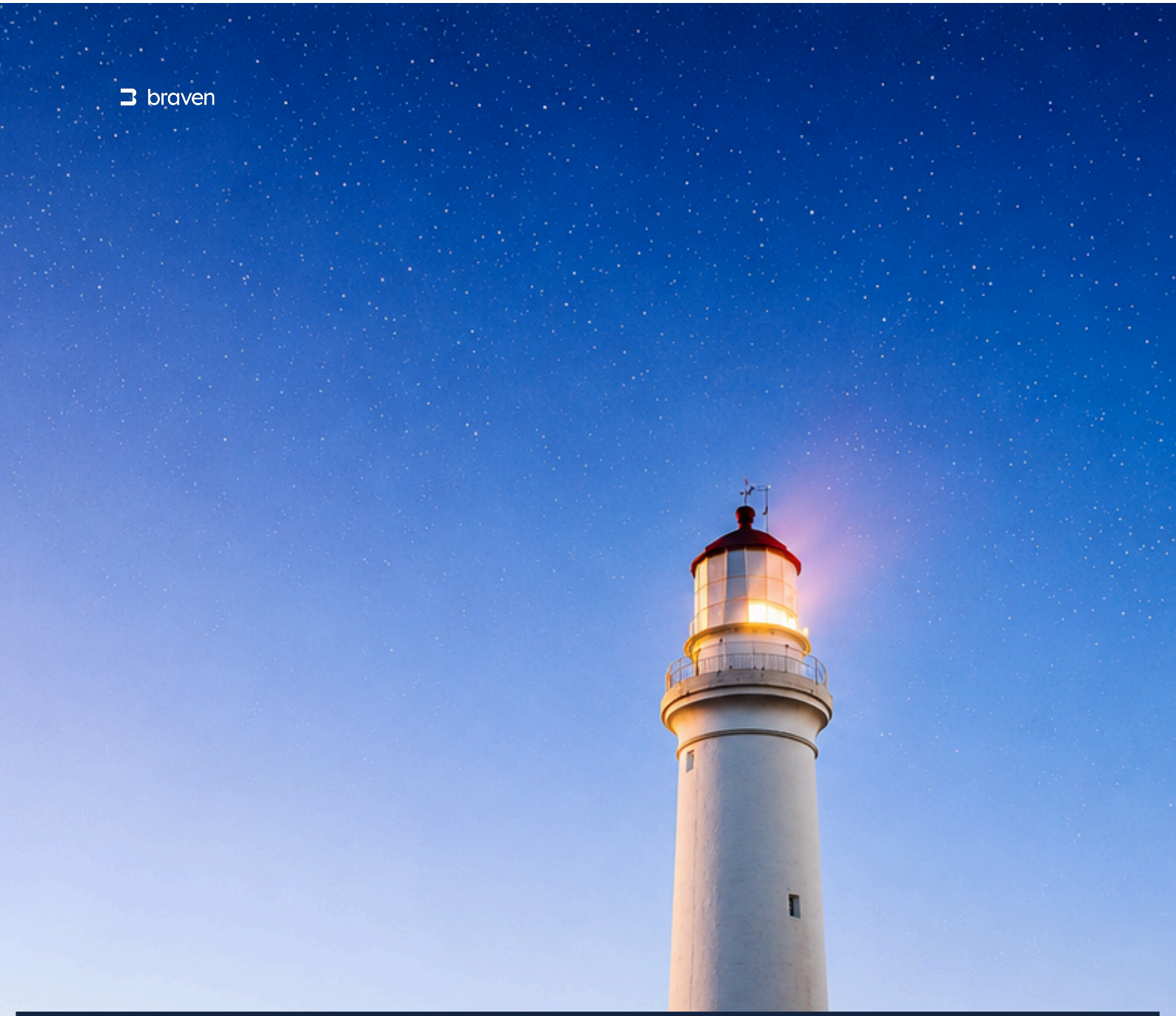
braven



BUYER'S GUIDE · UPDATED AUGUST 2026 · 30 MINUTE READ

The MGA's buyer's guide to AI operations

What to buy, what to skip, and the questions to ask any vendor, including us.

braven

The MGA's buyer's guide to AI operations

Written by Carlos Chávez, Co-founder and CEO of Braven.

Braven is headquartered in San Francisco with an office in London, and works with delegated authority teams across seven countries.

Published by Braven, the AI operations platform for MGAs and reinsurers that do their own underwriting, at gobraven.com. Reviewed quarterly.

Written for the people who run delegated authority businesses, and for the operations lead who will actually have to live with whatever gets bought.

They want to see how you underwrite, not just what you wrote

Your capacity providers have stopped being impressed by growth. They want to see how you underwrite, not just what you wrote. At the same time, most of the AI being sold to you right now will not survive its own pilot, and the reasons have almost nothing to do with the technology.

This is the guide we would want if we were sitting where you are: a scorecard, forty questions to put in writing, and a ninety day test that ends in a decision to sign or walk away.

Overview

01 Carriers are getting pickier. Growth on its own will not hold your binder the way it did three years ago.

02 Roughly 95% of enterprise AI pilots deliver nothing measurable. The causes are commercial and organizational rather than technical, which means you can test for all of them before you sign.

03 Most bad purchases are not bad products. They are the right product bought at the wrong level.

04 Be skeptical. Whoever hands you the evaluation criteria has already won half the argument. Write your own first.

05 Measure your operation in week one, before anyone installs anything. Otherwise you will be arguing about feelings in month three.

06 If a vendor is pressing you about the EU AI Act's August 2026 deadline, they are working from information that is out of date. It moved to December 2027. Section 8 has the dates that are actually real.

Discipline that cannot be evidenced is indistinguishable from luck.

Contents

Sources, page 67.

01

SECTION ONE

Where the delegated authority market actually is

Capital still wants what you do. Proving that you are the one worth backing has simply become harder.

It is 8:40am on a Monday. 61 unread emails. 9 of them are submissions, 2 of those are risks you already declined in March under a different broker's name, and one is a photograph of a schedule of values taken at an angle on somebody's phone — slightly out of focus, thumb in the corner.

Your best underwriter will spend until about 11:00am working out which is which, and none of that time is underwriting.

Nothing in this guide is really about that morning. Everything in it is.

Growth has stopped being the thing that impresses people

$128bn	$102.6bn	45%
Total US MGA premium, 2025[5]	Direct premium written, up 12%[5]	Of Lloyd's premium income is delegated[6]

Conning puts total United States MGA premium at roughly $128 billion in 2025, with statutory filings showing $102.6 billion in direct premium written, up 12% on the year before.[5] That growth ran at more than double the broader property and casualty market, which grew about 5%.[5] At Lloyd's, delegated underwriting accounts for roughly 45% of the market's premium income (a proportion that has barely moved in a decade).[6]

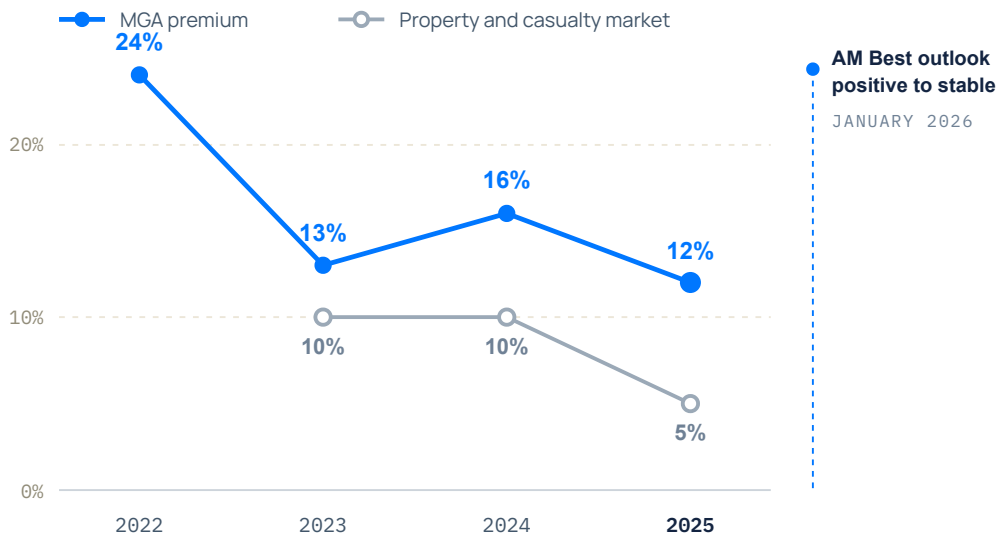
Then, quietly, the mood shifted. AM Best moved its outlook on the segment from positive to stable, and the reasons it gave are worth reading twice: moderate growth, tighter renewal economics, and increased scrutiny of delegated partners.[1] Carriers are asking about underwriting consistency and loss ratio stability rather than volume. Deloitte's read on 2026 dealmaking says much the same thing, that specialty property and casualty, excess and surplus, and MGAs still attract private capital, but buyers are choosier about underwriting quality.[7]

Capital still wants what you do. Proving that you are the one worth backing has simply become harder. And discipline that cannot be evidenced is indistinguishable from luck. Your book might be beautifully underwritten. But when the only proof lives in your head and in a shared inbox, nobody outside the building can tell.

FIGURE 1

MGA growth still outpaces the market, even if it is slowing

Annual growth in total United States MGA premium against total property and casualty premium, on Conning's estimates.



Source: Conning; AM Best. MGA premium grew 12% in 2025, more than double the broader property and casualty market's approximately 5%.[5] Conning published no comparable market growth rate for 2022. AM Best moved its outlook on the segment from positive to stable in January 2026.[1]

Where your underwriters' week actually goes

Capgemini measured it. Its World Property and Casualty Insurance Report 2024 found underwriters spending 41% of their time on administration and operations rather than on risk.[8] Across commercial and personal lines the range runs 41% to 43%, with only about a third of the day going to the actual job of assessing risk, pricing it and managing the book.[9] By the 2026 edition, the consequences were showing up in the outcomes:

41% of an underwriter's time goes to administration and operations rather than to risk[8]

61%	of underwriters struggle to improve quote to bind conversion
57%	find it hard to hold underwriting accuracy and risk quality where they want it
49%	are struggling to keep brokers and clients happy[10]

A carrier can absorb that. With a large expense base, the drag just disappears into it. You have no such cushion. Every hour spent retyping a schedule of values is an hour not spent on the broker who sent it, and it comes out of a commission that does not stretch.

What actually goes wrong when MGAs buy AI

Before you take a single vendor call, consider the following.

FINDING	WHAT IT MEANS FOR YOU
About 95% of enterprise generative AI pilots deliver no measurable profit and loss impact. Only around 5% reach production with value. (MIT NANDA, 2025)[2]	A successful pilot is weak evidence. Ask for production references at firms your size.
Externally built tools succeeded roughly twice as often as internal builds. (MIT NANDA, 2025)[2]	Building your own is the riskier path, not the safer one.
More than 40% of agentic AI projects will be canceled by end 2027, due to escalating cost, unclear business value or inadequate risk controls. (Gartner, 2025)[3]	Not one of those three causes is technical. All three can be tested before you sign.
Widespread "agent washing". Only about 130 of the thousands of vendors claiming agentic AI are the real thing. (Gartner, 2025)[3]	The label tells you nothing. Test the behavior instead.

FINDING	WHAT IT MEANS FOR YOU
90% of workers report daily use of personal AI tools while only 40% of companies hold official subscriptions. (MIT NANDA, 2025)[2]	Your submissions are probably already going into consumer tools. That is a risk you have today.
Generic assistants stall in enterprise use because they do not learn from or adapt to workflows. (MIT NANDA, 2025)[2]	The models are not the bottleneck. Context and memory are.

ONE CAVEAT

Before anyone quotes the 95% at you, including us. The MIT paper is a preliminary study built on 52 interviews and 153 survey responses.[2] While the precise number is arguable, the direction is not, and it matches what Gartner found by a completely different route.

What the 5% did differently

The more useful question is what the 5% did differently. On MIT's evidence, they did two things: first, they bought rather than built, and they pointed the tool at one workflow somebody owned rather than spreading it thinly across a department.[2]

Frontier models are genuinely extraordinary, and your underwriters are right to be impressed by them. But in their consumer form they arrive at your business as a stranger every single morning. They have no idea what is in your book, which markets you work with, what your appetite is, or what you decided about a nearly identical risk in April.

Many of them improve their models using whatever you type unless you go and change a setting. Very few come with the isolation, the processing terms or the retention controls that a capacity provider would sign off on, let alone a regulator.

Intelligence was never the gap.
Context, permissions and evidence are.

The large carrier prescription

Big consultancies have converged on a sensible answer. For example, McKinsey argues that the money is in rewiring a whole domain end to end. However, to do that it recommends a scalable operating model of 20 to 50 pods, a modular stack of reusable components, and a digital talent bench that is ideally 70% to 80% in house.[4] It also makes a point that almost nobody budgets for: expect to spend at least another dollar on adoption for every dollar you spend building, because change management is roughly half the total effort.[4]

If you are one of the many MGAs we meet with on a daily basis, none of that is available to you. You do not have 20 to 50 pods, and you do not have a hundred data scientists on call.

What you do have, and what people at large carriers quietly envy, is a very short decision chain, one book you can hold in your head, and the ability to change how work gets done on a Tuesday afternoon without asking anyone's permission.

20–50	70–80%	1:1
pods in McKinsey's recommended operating model[4]	of digital talent held in house[4]	adoption spend for every dollar of build[4]

So here's the translation.

TABLE 1

The large carrier prescription, translated

LARGE CARRIER PRESCRIPTION	MGA EQUIVALENT
Rewire one to three domains end to end	Rewire one flow end to end: submission in, decision out, evidence attached
Build a reusable component library	Buy a platform whose components are already reusable across your lines
Stand up an AI control tower for governance	Require the audit trail to be a product feature, not a project
Match build spend dollar for dollar with change management	Choose a vendor whose own team does the change management, then hold them to a measured baseline
Recruit 70% to 80% in house digital talent	Recruit none. What makes you money is underwriting judgment, not platform engineering

The honest downside of buying

McKinsey is also honest about the downside of buying, and you should take it seriously rather than assume we would hide it. Bought software deploys faster and arrives proven, but you accept limits on customization, a dependence on somebody else's roadmap, and what the firm calls a reversion to market median performance, because you are using the tools everybody else is using.[4]

The way an MGA gets around that is to be careful about what it buys. Buy the operational layer, where being merely as good as the market would still be a substantial improvement on a shared inbox.

Keep your appetite, your wordings and your pricing judgment to yourself. If a system asks you to adopt somebody else's view of what you should write, that is not a configuration setting, that is the end of your edge.

BUY THIS	KEEP THIS
The operational layer. Reading the mess, checking the rules, finishing the work, leaving the evidence, keeping the book live.	Your appetite, your wordings, your pricing judgment. The view of what you should write.

If a system asks you to adopt somebody else's view of what you should write, that is not a configuration setting, that is the end of your edge.

02

SECTION TWO

The buying ladder, and how to acquire at each rung

Nearly every disappointing purchase in this market comes down to the same mismatch: somebody buys at Rung 1 and expects Rung 3.

Nearly every disappointing purchase in this market comes down to the same mismatch: somebody buys at Rung 1 and expects Rung 3. It usually surfaces around the eight month mark, when everyone agrees the AI did not work. In fact, most of the time it worked exactly as it was sold, and was never going to do what was hoped.

Consider bringing this ladder with you to every initial vendor meeting and ask them which rung they are on.

FIGURE 2 · WHO FINISHES THE WORK

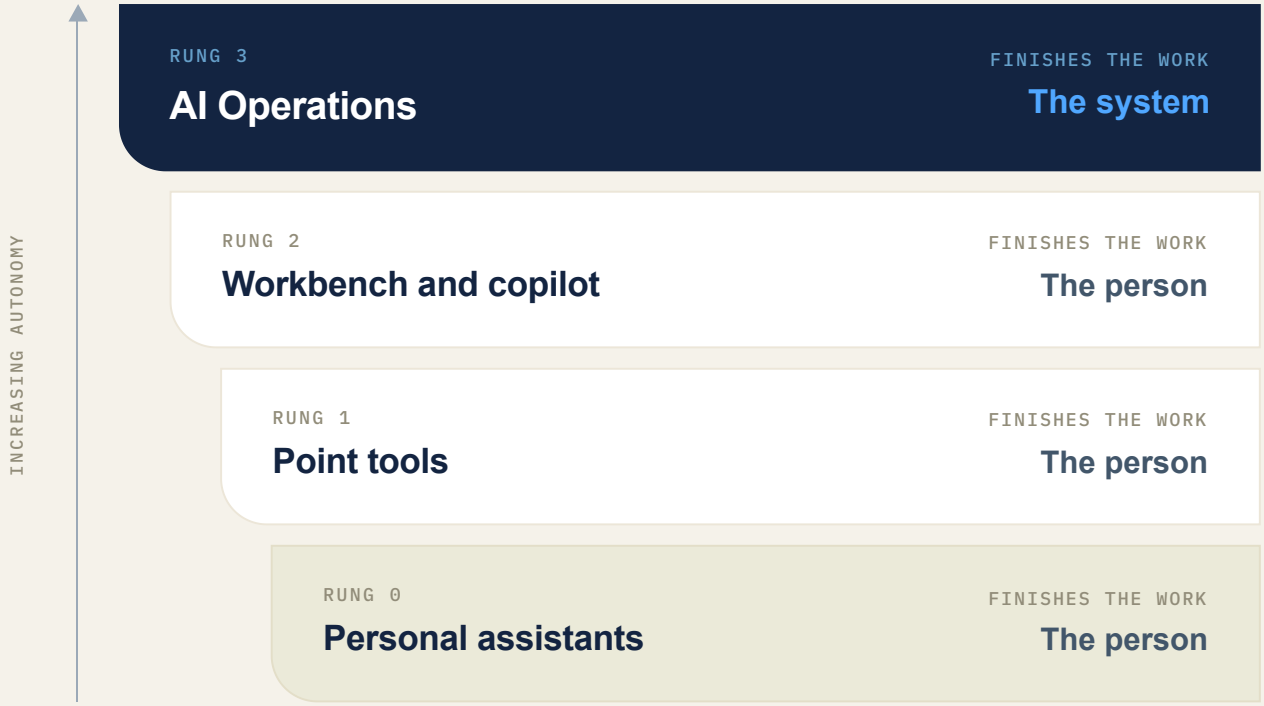


TABLE 2

The four rungs, and where each one stops

RUNG	WHAT IT IS	WHAT IT DOES WELL	WHERE IT STOPS
0. Personal assistants	Consumer chat tools your team already uses on their own	Drafting, summarizing, thinking out loud. Genuinely useful, and free	No memory of your book, no permissions, no audit trail, no contract protecting you. Usually invisible to management
1. Point tools	Document extraction, classification, one task automated	Turning one kind of document into fields, cheaply and reliably	Someone still has to decide what to do with those fields. The work is not finished
2. Workbench and copilot	An interface that gathers the data and suggests what to do next	Fewer clicks, and the context in one place instead of six tabs	All the value depends on people opening it. The day they stop, the benefit stops. Usually asks you to move where work happens
3. AI Operations	A system that runs the operational work end to end, inside boundaries you set, leaving evidence behind at every step	Finishing work rather than helping with it. Quotes drafted, follow ups chased, bordereaux built, the book kept current	Needs real integration, real governance, and a vendor prepared to be judged on your numbers rather than its own

KPMG's way of slicing this is useful in the same conversation. Their taxonomy separates taskers, which handle one repetitive job, from automators, which run a whole workflow, from collaborators, which work alongside people, from orchestrators, which coordinate other agents like a control tower.[11]

Taskers
Handle one repetitive job

Automators
Run a whole workflow

Collaborators
Work alongside people

Orchestrators
Coordinate other agents like a control tower

Ask which of the four you are being sold. Paying orchestrator prices for a tasker is the most common way to overpay in this market.

ONE WORD ON RUNG 0

Do not ban it. As many of our customers learned the hard way, bans merely push it underground, and MIT's numbers suggest it is already everywhere in your firm whether you have noticed or not.[2] Write a short acceptable use rule instead. Nothing that identifies a client, no submission documents, no loss runs, no personal data, into any tool your data processing agreements do not cover.

Then deal with the actual cause, which is that somebody reached for ChatGPT because the sanctioned system could not do the job.

To build, buy, or partner

Once you know which rung you need, the next question is how you get there. KPMG frames it as build, buy or borrow, and it places building with organizations that have deep technical expertise, scalable infrastructure, mature agent operations, well defined governance, and subject matter experts they can pull in whenever the model gets something wrong.[11] Consider this list carefully against your own firm's capabilities before you answer.

WHAT BUILDING REQUIRES, ON KPMG'S ACCOUNT[11]

- Deep technical expertise
- Scalable infrastructure
- Mature agent operations
- Well defined governance
- Subject matter experts you can pull in whenever the model gets something wrong

The desire to build runs strong in this market, and after all, your process really is distinctive. Where it goes wrong is in concluding that the software wrapped around that process must be distinctive too.

TABLE 3

Build, buy, partner

FACTOR	BUILD	BUY	PARTNER
Time to first production value	12-24 months	Weeks	3-6 months
Who encodes the domain knowledge	You do, from scratch	Already built for delegated authority	Shared, and you carry the specification burden
Ongoing model maintenance	Yours, permanently	Vendor managed	Contractual, and usually renegotiated
Integration library	Built one connector at a time	Pre-built for the systems you run	Mixed
Realistic first year cost	2-4 engineers plus infrastructure, plus a matching spend on adoption[4]	Subscription plus implementation	Project fees plus subscription
Odds of reaching production	About half that of buying[2]	The higher probability path[2]	Somewhere between
Right choice when	Your operational workflow genuinely is the moat, and you already employ engineers	**Almost every MGA**	You want to de risk the decision before committing to build

The optimal approach? Keep the judgment proprietary but buy the plumbing.

03

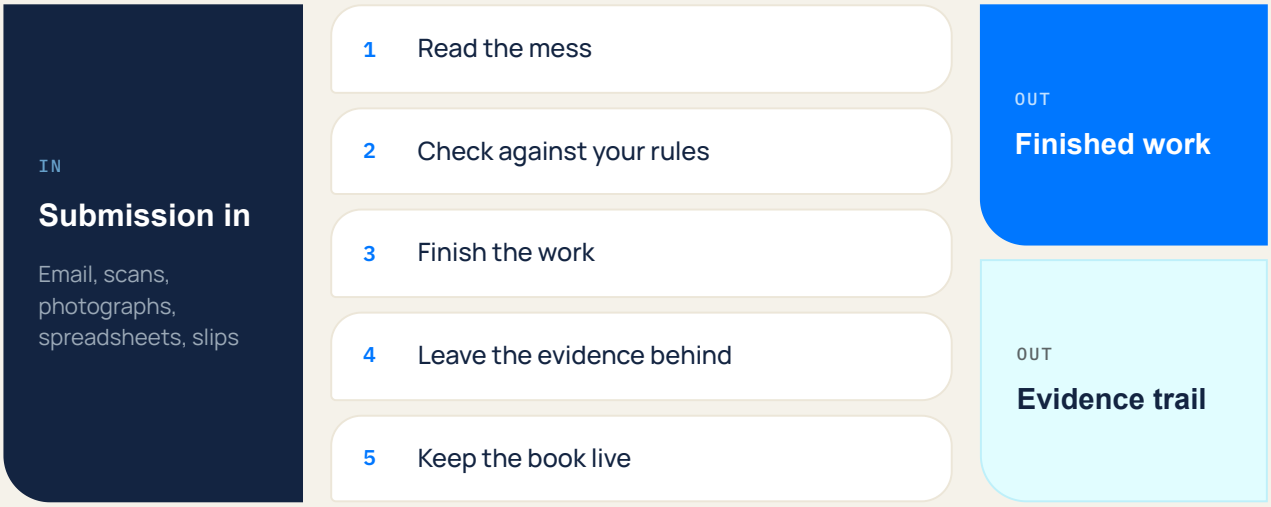
SECTION THREE

What to buy

Five things earn their place in a delegated authority operation. Anything doing fewer than three of them is a Rung 1 tool, and should cost you Rung 1 money.

Five things earn their place in a delegated authority operation. Anything doing fewer than three of them is a Rung 1 tool, and should cost you Rung 1 money.

FIGURE 3 · ONE PASS, TWO OUTPUTS



Both outputs come out of the same pass. The evidence is a by-product of doing the work, not a second system bolted on afterwards.

1 Read the mess, in whatever form it turns up

Submissions do not arrive as clean forms. Instead they arrive as email bodies, scanned PDFs, phone photographs, spreadsheets with merged cells and a hidden tab, and slips in a format that quietly changed at last renewal.

So the test is not whether a vendor can read a tidy ACORD form. All of them can. The test is your worst file.

WHAT GOOD LOOKS LIKE

It reads the whole file, puts it into the fields you expect, and links every field back to the page it came from. And you see how much of the file it actually captured, per submission, so you are not quietly assuming it got everything.

HOW TO TEST IT

Pull the five ugliest submissions from the last quarter (every firm knows exactly which ones they are). Not the tidy ones you would show a visitor; the bad ones. Hand them over and ask to see the output and the source links in the same meeting. Being offered a prepared sample instead is itself the answer.

2 Check against your rules

Your appetite, your binder terms, your wordings, your sanctions checks and your referral thresholds are yours. Anything that classifies and routes risk needs to hold those rules, apply them the same way every time, and let your own team change them without raising a support ticket.

WHAT GOOD LOOKS LIKE

The risks that were never worth your time do not reach an underwriter at all, and the ones that do arrive with the binder check already done and visible.

HOW TO TEST IT

Ask them to load one real appetite statement and one real binder during the evaluation. Next, feed it a risk that breaches them subtly rather than obviously. Just outside territory. A little over limit. An excluded occupancy described in language nobody would recognize unless they knew the class.

Watch whether it gets caught, and more importantly, how the breach gets explained back to you.

3 Finish the work, do not just tee it up

We see a clear dividing line between Rung 2 and Rung 3, and most of the money sits on it. Drafting a quote, chasing a follow up, assembling a bordereau, answering a market invitation: these are pieces of work with an output.

Either the system produces that output or your team still does, and only one of those changes your capacity.

WHAT GOOD LOOKS LIKE

Finished work, produced inside boundaries you set, with a human approving anything that binds, prices, or commits the firm.

HOW TO TEST IT

Count the human touches before and after. Where the number of times a person has to open a document does not fall, you have bought a nicer interface at a higher price.

4 Leave the evidence behind without being asked

Of the five, this is the one MGAs undervalue most and the one capacity providers care about most. Every extracted field linked to its source page. Every action recorded, human and machine alike. Decision trails sitting on every risk, without anybody having to assemble them afterwards.

WHAT GOOD LOOKS LIKE

Your capacity provider asks why a risk was written, and the answer takes four minutes and arrives with sources attached, instead of a week of archaeology across three inboxes and somebody's laptop.

HOW TO TEST IT

Take a bound risk from the evaluation and ask the system to show every input, every rule check and every action that produced it. Time it on your phone.

5 Keep the book live

Bordereaux have quietly become an after action report rather than a control. Data reaches the carrier weeks after the events it describes, having been re-keyed and hand corrected by two or three people along the way, which is exactly why delegated oversight feels reactive to everybody involved.[12]

Lloyd's has been chipping away at the same problem across the market through the Delegated Data Manager and the coverholder reporting standards, both of which exist to validate delegated data at the point it is collected rather than months later.[13]

WHAT GOOD LOOKS LIKE

Composition, appetite adherence and binder alignment as they stand today, per program and per underwriter, with the bordereau falling out of the same data rather than being rebuilt at month end.

HOW TO TEST IT

Ask what the book looked like on a specific day three weeks ago. Now time how long the answer takes.

The test is not whether a vendor can read a tidy ACORD form. All of them can. The test is your worst file.

04

SECTION FOUR

When to pass

There are nine ways to waste money here. And, in fairness, several are things we used to be tempted to sell you too.

There are nine ways to waste money here. And, in fairness, several are things we used to be tempted to sell you too:

01 The rip and replace

Any proposal that opens by replacing your policy administration system, your CRM or your broker portal has quietly moved all the risk to you and all the revenue to them. Your systems are not what is broken. The unstructured work between them is. Insist on coexistence and watch how quickly the conversation changes.

02 The strategy engagement that produces a roadmap

A roadmap is not an outcome and if you have fewer than a hundred staff you do not need a 200 page AI strategy. You just need one flow working by the end of the quarter.

03 The pilot with no end date

A pilot that was never designed to conclude is a way of deferring a decision. Fix the baseline, the metrics and the decision date before it starts.

04 The accuracy claim with no denominator

"99% accurate" is decoration until you know exactly on which fields, on which document types, measured against what, and refreshed how often. A vendor who measured accuracy once (at implementation) is telling you they do not measure accuracy.

05 The chatbot in a trench coat

A conversational wrapper on a model, with no memory of your book, no permissions and no audit trail, is still Rung 0 even if it comes with a Rung 3 price tag. Keep in mind that Gartner estimates only roughly 130 of thousands of self described agentic vendors are the real thing.[3]

06 The data for discount trade

A lower price in return for rights to use your submissions to improve a shared model is not a discount. It is you selling information about your book into a market that contains your competitors. Read the training and improvement clauses with at least as much care as you read the pricing page.

07 Anything that can outrank your underwriter

If an automated action can exceed the permissions of the person who set it off, you have built an accountability gap, and both your capacity provider and your regulator will eventually find it. PwC's standard is the right one: give every agent a verified identity, a defined role, task specific permissions, an auditable record, and hard limits on what it may do alone, with oversight tightening as autonomy and consequence rise.[14]

08 Per seat pricing on an operational platform

If the whole point is that each underwriter can handle more, then per seat pricing charges you for the headcount you were trying not to hire. Push for pricing tied to volume processed, or to outcomes.

09 Vanity integrations

A wall of 150 logos is not the same as the four connectors you need working on day one. Ask which of your systems are live in production today, at a customer you can name.

05

SECTION FIVE

Three criteria that sound neutral and are not

Every published evaluation framework is shaped, consciously or not, around what that vendor happens to be good at.

Many vendors in this market publish an evaluation framework, and every one of them is shaped, consciously or not, around what that vendor happens to be good at. However, distrusting every framework is not the answer. Here we help you understand the three criteria that get engineered most often, and to translate each one back into the question it should have been.

"Live in 48 hours"

Implementation speed is a real signal, and it is also the easiest number in the industry to manufacture, because almost nobody defines what "live" means. Connecting an inbox is not the same as running your book. Somebody can be technically live before lunch tomorrow and still have delivered nothing you can measure six months later.

ASK INSTEAD

How long until a number on our baseline moves, and will you put that date in the contract?

Time to connection is their number; time to measured outcome should be yours.

Agent count

Long catalogs get presented as breadth and can be misinterpreted as maturity. When a vendor advertises 25 production ready agents covering capital modeling, retrocession, insurance linked securities and commutation analysis, pause and consider what validating any one of those in production actually requires.

This is the so-called "agent washing" Gartner was pointing at when it estimated that only about 130 of thousands of self described agentic vendors were genuine.[3]

130 of the thousands of vendors claiming agentic AI are, on Gartner's estimate, the real thing[3]

ASK INSTEAD

Which three of these run every day at a customer whose name you will give me and whose number I can dial?

Breadth is cheap to claim and expensive to verify. Depth is the other way round.

On premise deployment

On premise turns up on requirement lists constantly, sometimes as a hard disqualifier. But hardly anyone actually wants on premise. What they want is isolation, data staying in the right region, control, and auditability.

Those four are perfectly achievable in a cloud deployment with per customer isolation, a choice of hosting region and data you can export on demand, without the servers, the patching, the security burden or the eighteen month project. Making the deployment model the criterion, rather than the four properties underneath it, inadvertently eliminates every vendor that delivers those properties a different way.

Isolation	**Residency**	**Control**	**Auditability**

ASK INSTEAD

Name the four properties and ask how each one is delivered.

If a vendor can evidence isolation, residency, control and auditability, the deployment model is an implementation detail.

Finally, be wary of any framework whose disqualifiers happen to exclude every competitor while admitting its author.

06

SECTION SIX

The Delegated Authority AI Scorecard

Forty questions in eight weighted categories, with a shortlisting threshold and two elimination gates. We actively suggest you copy it and take it to vendors who are not us.

How to use it

What follows is The Delegated Authority AI Scorecard: forty questions in eight weighted categories, with a shortlisting threshold and two elimination gates. We actively suggest you copy it and take it to vendors who are not us.

1. **Set your weights before you contact anybody.** The ones below are just a suggested starting point. Adapt them to your context.

2. **Send the forty questions, not your weights.** A vendor who can see your rubric will always write to it.

3. **Ask for written answers**, before the demonstration rather than after it.

4. **Score the written answers yourself**, one to five per category, using the anchors under each section.

5. **Apply the two gates first, then the 3.5 bar.** A vendor can fail on a gate while scoring well overall, and that is still a fail.

To be clear, none of these questions are ours originally. Most come from control expectations that independent parties already publish. KPMG, for example, lists what any agentic deployment needs before it goes anywhere near production:[11]

- Observability, and audit trails at the level of individual actions
- Error handling and a way to roll back
- Authentication, authorization, and least privilege access
- Secure tool calling with valid schemas
- Escalation to a human, whether in the loop or on it
- Red teaming, and an AI system card you can actually read
- A choice of model, rather than one vendor's bet
- Measurable performance indicators, and somebody maintaining the thing afterwards

PwC adds verified identity, defined role and bounded autonomy for each agent.[14] And the NAIC's Model Bulletin expects a written program with senior accountability, plus oversight of third party AI, where the insurer stays responsible for how that AI behaves.[15]

How to score

Score each category one to five on the written answers, apply the weight, add it up. Around 3.5 is a sensible bar for a shortlist.

Two categories are elimination gates rather than scores. Anything below 3 on autonomy and control or on data ownership and privacy should end the conversation, whatever the total says. Those are the two places where getting it wrong cannot be fixed later with money.

3.5

Weighted total that makes a sensible shortlist bar

< 3

On category C or D ends the conversation, whatever the total says

ABOUT THESE WEIGHTS

As mentioned earlier, every vendor's framework is shaped around its own strengths. Here is ours, so you can see the shape of it.

We have weighted fit to delegated authority heavily, and we would, because that is what we built. We have weighted deployment speed lightly, and we would do that too, because we take five to eight weeks and somebody else will promise you Thursday. Change these numbers. We put that column there to be written over. Our weights being right is not the point; arriving at the meeting with weights of your own is, rather than accepting a scorecard from whoever got to you first.

FIGURE 4 · TEAR OUT AND TAKE IT WITH YOU

braven

The Delegated Authority AI Scorecard

VENDOR _____ REVIEWER _____ DATE _____

CATEGORY	WEIGHT	SCORE 1 TO 5	WEIGHTED
A. Evidence and accuracy	20%		
B. Fit to delegated authority	20%		
C. Autonomy and control GATE	10%		
D. Data ownership and privacy GATE	15%		
E. Security and resilience	10%		
F. Governance and regulatory	10%		
G. Deployment and integration	10%		
H. Commercials and exit	5%		
Total	**100%**		

SHORTLIST BAR

Around 3.5 weighted is a sensible bar for a shortlist.

ELIMINATION GATES

Below 3 on C or D ends the conversation, whatever the total says.

The forty questions

Send these. Keep your weights to yourself.

A. Evidence and accuracy

WEIGHT 20%

1. What is your extraction accuracy, on which document types, and against what ground truth?
2. Is accuracy measured continuously in production, or only during implementation?
3. Does every extracted field link back to its source page? Show me.
4. Do you report extraction completeness per submission, so we know how much of a file was captured?
5. What happens when the system is not confident? Does it guess, flag, or stop?
6. Will you publish your accuracy benchmark, or is it only available under non disclosure?

Score 5: accuracy is measured on every correction a user makes, you can see it, and the number moves over time. **Score 1:** one percentage, no denominator, no date, no method.

B. Fit to delegated authority

WEIGHT 20%

7. How many delegated authority businesses run you in production today, and where?
8. Does the system understand binders, slips, bordereaux, layers, attachment points and cedents natively, or are those just generic documents to it?
9. Can we load our own appetite, wordings and binder terms, and edit them ourselves without raising a request?
10. Can you produce bordereaux in our capacity providers' formats, including Lloyd's coverholder reporting standards where they apply?
11. Which languages do you operate in, and is that the whole system or just the interface?

Score 5: named production references, at firms your size, whom you may contact directly. **Score 1:** "we work with several insurance clients" and no names.

C. Autonomy and control GATE WEIGHT 10%

12. Can an automated action ever exceed the permissions of the user who triggered it?
13. Which actions need human approval before they take effect, and who decides what goes on that list?
14. How do we set and change the boundaries of autonomous action?
15. What is the rollback path when it does something wrong?
16. Does each agent carry its own identity in the audit record, or does everything appear as the system?

Score 5: permissions inherited from the human, boundaries you configure yourself, sensitive actions gated.
Score 1: "the model is very reliable" offered as though it were a control.

D. Data ownership and privacy GATE WEIGHT 15%

17. Is our data ever used to train or improve any model, yours or anyone else's? Point me to the clause.
18. Where is our data hosted, and can we choose the region?
19. Is every customer's data isolated? Could anything we process be seen, inferred or reconstructed by another customer?
20. If we leave, what do we get back, in what format, and how fast?
21. Which sub processors touch our data, and how are we told when that list changes?

Score 5: they point you at a clause. **Score 1:** "we would never do that", with nothing in the contract that says so.

E. Security and resilience WEIGHT 10%

22. Encryption at rest and in transit, backup regime, and recovery objectives?
23. How do users sign in, and does it work with the identity provider we already run?
24. How dependent are you on a single model provider, and what happens if that provider has an outage or changes its terms?
25. Have you been red teamed, by whom, and what did you change afterwards?

Score 5: more than one model provider, infrastructure managed as code, a specific answer on recovery times.
Score 1: a security page covered in badges and short on detail.

F. Governance and regulatory

WEIGHT 10%

26. What does the audit trail cover, is it tamper evident, and how long do you keep it?
27. Can you produce, on demand, the full decision trail for one bound risk?
28. Do you provide documentation we can hand to a capacity provider or a regulator, such as a system card?
29. How do you support our obligations under the NAIC Model Bulletin in the states that have adopted it?
30. What is your position on the EU AI Act, and which obligations do you think apply to this system?

Score 5: a precise account of which obligations apply, which do not, and by when. **Score 1:** urgency about a deadline they cannot cite correctly. Section 8 will tell you which is which.

G. Deployment and integration

WEIGHT 10%

31. How long until a number on our baseline moves, and will you put that date in writing?
32. What do we have to replace? What stays exactly as it is?
33. Which of our systems do you connect to, and are those live in production today or on a roadmap?
34. Who does the configuration work, and what does our team's time commitment look like week by week?
35. What does change management look like, and who owns adoption when people go quiet in week six?

Score 5: their team does the shaping, your systems stay put, and the time you owe them is quoted in hours. **Score 1:** a discovery phase, billed separately, before anything runs.

H. Commercials and exit

WEIGHT 5%

36. What is the pricing basis, and what happens to the bill if our submission volume doubles?
37. What is the total first year cost, including implementation, integration and support?
38. What happens if it does not work? Is there a guarantee, and has anyone ever claimed it?
39. What is the notice period, and what are the exit assistance obligations?
40. Which operational metrics will you be measured against, and will you baseline them before we start?

Score 5: they propose the measurement before you think to ask for it. **Score 1:** a refusal to baseline, on the grounds that value is hard to quantify.

07

SECTION SEVEN

Arriving at a fact-driven decision in 90 days

Easily the most valuable thing you will do in this process costs nothing: measure your operation as it is now.

Easily the most valuable thing you will do in this process costs nothing: measure your operation as it is now.

McKinsey's point that change management is roughly half the total effort[4] has a sharper edge at your size. When nobody can see the thing improving, people stop using it, and the investment dies without anyone ever holding a meeting to kill it.

STEP 1

Baseline in week one, before anything is installed

Most MGAs can put these together in a few days from email and the policy administration system. Duller than a demo, and worth considerably more.

Be sure to measure these 8 numbers before anyone installs anything. Table 4, overleaf.

TABLE 4

The week one baseline

METRIC	DEFINITION	WHY YOUR CAPACITY PROVIDER CARES
Time to quote	Median hours from submission received to quote issued	Runs straight through to hit rate and to whether brokers think of you first
Submission capture rate	Share of inbound submissions logged and triaged, against everything that arrived	Shows leakage that neither you nor your carrier can currently see
Response rate	Share of market invitations answered	An unanswered invitation costs you a relationship and never appears in any report
Hit rate and quote to bind	Bound over quoted, by class and by broker	The clearest read on risk selection and speed at the same time
Appetite adherence	Share of bound risks inside agreed appetite, per underwriter	The number your binder review is really about
Binder alignment	Share of bound risks inside binder terms, scored at bind	Turns oversight from retrospective sampling into something continuous
Bordereaux production time	Working days from period close to accepted bordereau	Late, rekeyed data is the root of why oversight feels reactive[12]
Gross written premium per underwriter	Premium bound per underwriting head	The capacity question, stated plainly

We will detail each of these in depth in our upcoming companion guide, *The bordereau vs the live book*, which covers how to produce each one without hiring a reporting team.

STEP 2

Run it on real work

Live submissions, a real class of business, actual underwriters. Sandboxes prove the software runs. Whether your people will use it on a bad day is a different question entirely.

Put your most skeptical underwriter on it. If the system cannot win that person over inside ninety days, it will not survive contact with your operation, and you would much rather find that out now.

STEP 3

Write down the decision before you start

Before day one: what movement in which metrics counts as success, who decides, and on what date. Put it in the order form. Any vendor who will not agree to a written definition of success has just told you something useful.

STEP 4

At day ninety, look past the headline

Three things predict year two better than any efficiency number.

01 Is usage climbing or drifting down week on week?

02 Are corrections falling as the system learns your book?

03 Did anybody quietly go back to the old way of doing it?

People vote with their behavior long before they say anything in a meeting.

08

SECTION EIGHT

The governance questions you will be asked

There is a lot of misplaced urgency being sold around these dates at the moment, so it pays to know the actual ones.

Now we get to the critical conversations in front of you: one with your capacity providers and one with regulators. There is also a lot of misplaced urgency being sold around them at the moment, so it pays to know the actual dates.

United States: the NAIC Model Bulletin

On 4 December 2023 the NAIC adopted its Model Bulletin on the Use of Artificial Intelligence Systems by Insurers.[15] By the second quarter of 2026, 25 jurisdictions including the District of Columbia had adopted it. California, Colorado, New York and Texas run their own insurance specific frameworks instead, which puts 29 jurisdictions in scope one way or another.[16]

The bulletin does not create new law. What it does is explain how state departments will read the unfair trade practice, market conduct and corporate governance statutes that already exist when AI is involved. It asks insurers for four things:[15]

- A written AI systems program
- Accountability at senior management and board level
- Model validation and testing
- Oversight of third party AI, where the insurer stays responsible for how that AI behaves

There is also an AI Systems Evaluation Tool, a structured framework for examiners, being piloted by a group of states through September 2026.[17]

WHY THIS LANDS ON YOUR DESK

Your carrier is accountable for the AI used to write business on its paper. When their compliance team gets asked how that AI behaves, they will come to you, because you are the one underwriting. So expect the question at your next binder review, and expect it even though you are not the licensed insurer and nothing in your own license requires you to answer.

Europe: the EU AI Act and the latest dates

This is where you will hear the most confident wrong advice.

- Annex III classifies AI used for risk assessment and pricing of natural persons in life and health insurance as high risk.[18] Property and casualty is not named there. That makes most delegated authority underwriting a different question, one that still needs answering rather than assuming.
- Stand alone Annex III high risk obligations were originally due to apply from 2 August 2026. That date has moved. The Digital Omnibus on AI was adopted by the European Parliament on 16 June 2026 and by the Council on 29 June 2026, pushing those obligations to 2 December 2027, and Annex I embedded systems to 2 August 2028.[19]
- Not everything moved, though. Article 50 transparency obligations stand, and legacy systems have to meet the marking and detection duties by 2 December 2026.[19] AI literacy duties on deployers already apply, which in plain terms means your underwriters and compliance people should be able to explain what your systems do and where they stop.[20]

HOW TO USE THIS IN A MEETING

Any vendor leaning on an August 2026 high risk deadline is working from information a month out of date. Take that as a signal about how current the rest of their regulatory reading is. And use the extra runway on documentation, oversight design and evidence rather than treating it as a reprieve.

FIGURE 5

The dates that are actually real

United States and European Union obligations, 2023 to 2028.

Dec 2023 — **NAIC Model Bulletin adopted**

Adopted 4 December 2023. By Q2 2026, 25 jurisdictions including the District of Columbia had taken it up.[15,16]

Jun 2026 — **Digital Omnibus on AI adopted**

European Parliament 16 June, Council 29 June 2026.[19]

~~Aug 2026~~ — ~~**Annex III high risk obligations apply**~~

Superseded. Still widely quoted.

Sep 2026 — **NAIC evaluation tool pilot runs through**

A structured framework for examiners, piloted by a group of states.[16]

Dec 2026 — **Article 50 legacy compliance**

Legacy systems must meet the marking and detection duties by 2 December 2026.[19]

Dec 2027 — **Annex III high risk obligations apply**

2 December 2027. The date that replaced August 2026.[19]

Aug 2028 — **Annex I embedded systems**

2 August 2028.[19]

The London market and Lloyd's

If you hold a binder from a Lloyd's syndicate, your managing agent carries oversight obligations for the authority delegated to you, supported by the coverholder reporting standards, the annual compliance attestation, and the market's delegated data infrastructure.[13]

None of that changes because you started using AI. What changes is how much it costs you to satisfy it.

An operation that produces a decision trail and a clean bordereau as a by-product of doing the work turns an audit from an event into a query.

What your capacity provider will actually ask

Rather than a direct "do you use AI?" it may be something more like:

1. Where does AI touch a decision that affects what you bind on our paper?
2. Can a machine action exceed a human's authority in your shop?
3. Show me the decision trail on this risk.
4. How do you know your extraction is accurate, and how often do you check?
5. Who is accountable, by name, for the AI systems program?

An MGA that can answer those five in the room is negotiating from somewhere quite different than one that needs a fortnight and a call with its vendor.

09

SECTION NINE

Our own answers to our own questions

Publishing forty questions and then dodging them would be a bit rich. So here we are, including the parts where you should buy from somebody else.

Where Braven is strong

Evidence

Every extracted field links back to its source page. We measure accuracy on every correction a user makes, and we are publishing the benchmark rather than asking you to take a number on trust.

Built for this market, not adapted to it

Braven knows what a binder is. It operates in English, Spanish and Portuguese, and 17 or more MGAs, brokers and reinsurers across seven countries run it in production.

Named references

Ask us for customers by name and we will introduce you to them. A case study credited to "a Zurich reinsurer" is not a reference.

Autonomy with a ceiling

AI actions cannot exceed the permissions of the person they run for. Sensitive actions can require approval first. Every business change and every AI action goes into a tamper evident audit trail, with protected verification records kept for seven years.

Your data stays yours

Your data never trains a model. It sits where you choose. It leaves with you if you go. Every customer is isolated from every other.

We fit around what you already run

Braven connects to email, policy administration systems, CRMs and carrier portals. Your broker portal stays broker facing. Your CRM stays where it is.

No single point of failure

Multi model across Azure OpenAI and AWS Bedrock, on cloud portable infrastructure managed as code.

We start by measuring you

Every deployment baselines the metrics in section 7 first, then reports movement against them every cycle. Including the cycles where the movement is disappointing.

A guarantee with teeth

Ninety days, 100% of your money back. Nobody has claimed it yet.

17+	7	90
MGAs, brokers and reinsurers running Braven in production	countries	days, 100% of your money back

Where we are not the right answer

You need to be live this week

We take five to eight weeks, because we shape the system around your appetite, your binders and your systems first, and because we insist on baselining your numbers before we start. Somebody will offer you Thursday — and if that is genuinely the requirement, take it. We would rather lose the deal than win it on a definition of live that means an inbox is connected and nothing else has changed.

You want it on your own hardware

We are cloud deployed. We give you what on premise buyers are usually after, which is isolation, regional hosting, control and auditability, without the servers or the project. But if the deployment model itself is the requirement rather than those four properties, we are not your vendor.

You want a pricing engine or a capital model

Braven runs the operational work around underwriting. It does not price your risk and it does not model your capital. Those should remain yours.

You want to build it, and you actually can

If you employ real engineers and your operational workflow genuinely is your moat, KPMG's build path is a legitimate choice and you should take it seriously.[11] Most MGAs are not in that position, and MIT's finding that external builds succeed about twice as often as internal ones is worth sitting with before you commit.[2]

You want one document type extracted and nothing more

Buy a Rung 1 point tool. It will be cheaper and it will do the job. Come back when the constraint is the work rather than the document.

So try us

Send us your messiest submission. Not a clean one, the one with the thumb in the photograph. Watch Braven run it in thirty minutes, on your documents, at no cost.

Then take the Delegated Authority AI Scorecard in section 6 to every other vendor on your list, and ask them to answer in writing.

Book a walkthrough with Braven,
the AI operations platform for MGAs and reinsurers.

gobraven.com

Definitions

Agent washing Marketing an assistant, chatbot or scripted automation as an autonomous agent. Gartner estimated only around 130 of the thousands of vendors claiming agentic AI were the real thing.[3]

Agentic AI AI built to act toward goals across several steps rather than respond to single prompts. The label alone tells you nothing about capability.

AI operations platform A system that runs the operational work of a business, reading submissions, checking appetite, drafting responses and building bordereaux, and delivers finished work with an evidence trail, rather than helping a person with one task.

Appetite adherence The share of bound risks that sit inside the underwriting appetite agreed with your capacity provider, measured per underwriter and per program.

Binder alignment How closely bound business conforms to the binding authority agreement, scored at the point of bind rather than sampled afterwards.

Bordereau A periodic report from a coverholder or MGA to its carrier, listing risks, premiums or claims written under a binding authority. Traditionally monthly or quarterly. Traditionally late.[12]

Coverholder A company granted delegated underwriting authority by a Lloyd's managing agent under a binding authority agreement, able to bind risks on behalf of the syndicate within agreed limits.

Decision trail The full record of inputs, rule checks, human actions and machine actions behind a single underwriting outcome, retrievable for one risk on request.

Delegated authority An arrangement where an insurer or managing agent delegates underwriting, and sometimes claims handling, to a third party such as an MGA or coverholder.

Extraction completeness How much of a submission file was captured, structured and made searchable. Different from accuracy, which is about whether what was captured is correct. Ask about both.

Human in the loop A person approves before an action takes effect. Appropriate for anything that binds, prices or commits the firm.

Human on the loop A person supervises and can intervene, but the action does not wait for approval. Appropriate for chasing a missing loss run. KPMG treats escalation between the two as non negotiable.[11]

Least privilege The principle that any actor, human or automated, holds only the permissions its task requires. In practice it means an automated action never exceeds the authority of the person it runs for.

Live book A continuously current view of portfolio composition, appetite adherence and binder alignment, as opposed to a bordereau describing a position that has already changed.

Submission capture rate The share of inbound submissions that get logged and triaged. The gap between that and 100% is business you never knew you turned away.

Time to measured outcome How long between deployment starting and a baselined metric actually moving. The honest version of the go live date most vendors quote.

> The gap between your submission capture rate and 100% is business you never knew you turned away.

Frequently asked questions

What is an AI operations platform for MGAs?

An AI operations platform for MGAs runs the operational work of the business, reading submissions, checking appetite, drafting responses and building bordereaux, rather than assisting a person with one task. It hands back finished work with a full evidence trail attached.

How should an MGA evaluate an AI vendor?

Decide your own criteria and weights before any vendor gives you theirs, then score written answers rather than demonstrations. The Delegated Authority AI Scorecard in section 6 sets out eight weighted categories, forty questions, and a shortlisting bar of 3.5 out of 5, with autonomy and control and data ownership as elimination gates. Send the questions but keep the weights to yourself, because a vendor who can see the rubric will write to it. Be particularly careful with three criteria that sound objective and rarely are: implementation speed measured as time to connection, the size of an agent catalog, and on premise deployment treated as a proxy for control.

Does an MGA need to replace its policy administration system to use AI?

No. Work that eats your underwriters' time sits between systems, in email and in documents, not inside the policy administration system. Any proposal that starts with replacement has shifted the risk onto you. Insist on coexistence with what you already run.

How accurate does AI document extraction need to be for MGA submissions?

Measured accuracy with a source link matters more than the headline percentage. Trace a field back to a page and you can verify it in seconds. One you cannot trace needs the same manual check whether the vendor claims 90% or 99%. Ask how accuracy is measured, on what, and how often it is refreshed.

Can an MGA just use ChatGPT or Claude instead of an AI operations platform?

For individual tasks, yes, and plenty of good underwriters do. Those limits are structural, not a question of how clever the model is. Consumer assistants do not watch your inbox, hold your appetite rules, chase your follow ups or leave an audit trail, and they start every session knowing nothing about your book. MIT found that generic tools stall in enterprise use precisely because they do not learn from or adapt to workflows.[2] Separately, read the data processing terms before any client document goes near one.

How long should an AI deployment take at an MGA?

Weeks rather than quarters, but ask for time to measured outcome rather than time to go live. Anyone can connect an inbox in a day and still deliver nothing measurable six months later. Braven takes five to eight weeks, with your metrics baselined in week one.

What does a capacity provider want to see from an MGA using AI?

Evidence of discipline. How completely submissions are captured, how consistently appetite is applied, how each decision was actually reached, and bordereaux that arrive on time and reconcile. Delegated authority is a bet on discipline that a carrier usually cannot see until the account year tells them. Make it visible and the conversation changes.

Is AI use by MGAs and insurers regulated?

Increasingly. In the United States, 25 jurisdictions including the District of Columbia have adopted the NAIC Model Bulletin, with California, Colorado, New York and Texas running their own frameworks.[16] In the European Union, the AI Act treats life and health risk assessment and pricing as high risk, and those obligations now apply from 2 December 2027 following the Digital Omnibus.[18,19] Property and casualty falls outside that particular provision and needs its own analysis.

Does an MGA need its own AI governance policy?

If you underwrite on somebody else's paper, assume yes. Responsibility for third party AI runs down the delegation chain, and the request will reach you from your capacity provider well before any regulator asks. A short written program, naming an accountable person, describing where AI touches decisions and pointing at your audit trail, covers most of what gets asked.

Should an MGA build or buy AI software?

Buy, in almost every case. KPMG puts building with organizations that have deep technical expertise, scalable infrastructure, mature agent operations and subject matter experts on hand.[11] MIT found external builds succeed roughly twice as often as internal ones.[2] What makes you money is underwriting judgment. Keep engineering headcount out of it. Section 2 has the full comparison.

What is agent washing, and how do I spot it?

Agent washing is marketing an assistant or a scripted automation as an autonomous agent. Gartner estimated only about 130 of thousands of self described agentic vendors were genuine.[3] The quickest test is to hold catalog size up against production evidence: ask which three of the advertised agents run every day at a customer you can telephone.

Sources

1. AM Best delegated underwriting authority segment data and outlook revision, reported in *2026 MGA outlook: scaling smarter in a demanding market*, January 2026.
2. Aditya Challapally, Chris Pease, Ramesh Raskar and Pradyumna Chari, *The GenAI Divide: State of AI in Business 2025*, MIT Project NANDA, July 2025. Preliminary findings based on 300+ initiative reviews, 52 interviews and 153 survey responses; the authors describe the figures as directionally accurate rather than audited.
3. Gartner, *Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027*, 25 June 2025.
4. McKinsey & Company, *The future of AI in the insurance industry*, 15 July 2025.
5. Conning, *Managing General Agents: Reconfiguring the Insurance Value Chain?*, reported 28 July 2026.
6. Lloyd's, *Delegated Authority at Lloyd's*, accessed July 2026.
7. Deloitte, *2026 insurance M&A outlook*.
8. Capgemini, *World Property and Casualty Insurance Report 2024*.
9. Capgemini Research Institute for Financial Services, *Unleashing growth: the evolving role of underwriters*.
10. Capgemini, *World Property and Casualty Insurance Report 2026*, May 2026.
11. KPMG, *Agentic AI untangled: navigating the build, buy, or borrow decision*, January 2026.
12. InsTech and distriBind on rekeying and manual editing in the bordereaux chain, *Making delegated authority data exchange simple*.
13. Lloyd's, *Delegated Data Manager*, and LIMOSS, *Data and Reporting*, accessed July 2026.
14. PwC, *AI agent governance for workforce use*, Trust and Safety Outlook 2026.
15. National Association of Insurance Commissioners, *Model Bulletin: Use of Artificial Intelligence Systems by Insurers*, adopted 4 December 2023.
16. National Association of Insurance Commissioners, Big Data and Artificial Intelligence (H) Working Group, *Implementation of NAIC Model Bulletin: state adoption map*, accessed August 2026. content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-map-ai-model-bulletin.pdf
17. NAIC AI Systems Evaluation Tool, pilot status as at second quarter 2026. Pilot runs through September 2026.
18. Regulation (EU) 2024/1689 (the AI Act), Annex III, point 5(c).
19. Digital Omnibus on AI, adopted by the European Parliament 16 June 2026 and the Council 29 June 2026. Analysis: Freshfields and Gibson Dunn, June 2026.
20. EU AI Act, Article 4, AI literacy obligations on deployers.

braven

Architects of the Possible.
Move. Evolve. Win.

Braven is the AI operations platform for MGAs and reinsurers that do their own underwriting. Headquartered in San Francisco with an office in London.

gobraven.com

BUYER'S GUIDE · UPDATED AUGUST 2026 · REVIEWED QUARTERLY