

ANTON by openEXPERT

The ANTON family whitepapers · Autumn 2026 · Paper 6 of 11

Setup Beats Size

Why the right setup matters more than the largest model, and how to choose a model per task

> *Spend on setup every time. Spend on size when the task earns it.*

Version 1.0 · 19 November 2026

Created by Daniel Bardun · FutureChain AB

Open source ANTON is EUPL · github.com/altspace-hub/anton_desktop

See it live ai-portal.futurechain.eu (showcases with fictional data)

Before we start

Where we left off

Last week we opened the seven layers of knowledge: the foundation of house rules, the area's context and its regulatory packs, the module's expert method, a persona, skills, your own sources, and the setting for how hard to think and what to record. Stacked in a fixed order, those layers turn a general model into something that behaves like an experienced professional — and behaves the same way next time.

That paper ended with a question, and this paper is the answer to it. **If the context is right, do you still need the biggest model?** In the language of the first paper, this is the time and cost gap.

It is not an academic question. It is a budget question, a speed question and, as we will see next week, a question about where your data goes and who processes it. And it is a question most teams never ask out loud, because the market has trained us to assume the answer.

The question nobody asks

Every few months a larger model arrives, with a larger price and a launch event to match. The message underneath is always the same: *if your answer was weak, the model was too small.* So teams pay for the largest model for every task, "to be safe" — including the tasks a smaller, well-set-up model would handle perfectly well.

I want to argue the opposite, carefully. Not that model size does not matter. It does, and I will be specific about where. But that size is one lever among six, and usually not the first one to pull. **With the right setup, a mid-sized or smaller model will often do the job — at lower cost, faster, and, if you choose, hosted in the EU.**

I also want to be unusually honest about the evidence. What we have is a small internal test, a second small synthetic comparison, and a great deal of daily practice across the family. That is enough to change how I choose models. It is not enough to make sweeping claims, and I will not make any. You will find every caveat in this paper, in plain sight.

In this paper

- **Part 1, the problem:** the market sells size, three scenes, and why size is the wrong first question.
- **Part 2, our thesis:** model size is one lever among six — what we measured, how small it was, and what we are not claiming.
- **Part 3, our take:** a six-step framework for any AI task, the cost picture, and what we have built.
- **Part 4, a worked example:** three tasks on one Monday, three model choices.
- **Part 5:** what this means for you on Monday, and next week.

Who this is for

Heads of compliance, MLROs, FCP leads, risk managers and internal auditors who are asked to approve an AI budget — and anyone who has wondered whether the most expensive model is really buying them anything on a Tuesday afternoon. No technical background is needed.

Part 1 · The problem

The market sells size

Think about how AI models are presented to a buyer. There is a league table. There is a larger number than last time. There is a demonstration of something remarkably hard — a mathematical proof, a long chain of reasoning, a program written from a single sentence. And there is a price list in which the newest, largest model sits at the top.

None of that is dishonest. Larger models really are better at many hard things. But the way the market talks about them quietly teaches a lesson that does not hold in professional work: that quality is something you buy by buying size.

In a compliance function, most of what goes wrong with AI output is not a shortage of raw intelligence. It is a shortage of everything around it — the gaps we named in the first paper. The model did not know the firm was a payment institution and not a bank. It did not know the reader was the board. It did not have the final text of the regulation, only what it half-remembered from the proposal. It did not know what a good result looked like, because nobody had said.

A bigger model does not fix any of those things. It writes the same misunderstanding more fluently.

Three budgets, one assumption

These are composites. No names, no clients.

Scene 1: the expensive summary

A team decides, sensibly enough, that it does not want to worry about which AI model to use. It picks the largest one available and runs every request through it — "to be safe". That includes the two-paragraph summaries of meeting notes, the quick reformatting of a table, the first pass at mapping a document to a checklist.

A few months later, the monthly bill has tripled. Someone in finance asks what the team got for it. Nobody can show that the summaries got better. They were fine before. They are fine now. The team has been paying a partner's rate to take the minutes.

Scene 2: the big model with no context

A firm pays for access to a frontier model — one of the largest models from the major vendors — and gives it a one-line prompt: *"Assess our CDD procedure against the AMLR."* The output is long, fluent and generic. It reads like a summary of the regulation with the firm's name added.

At the next desk, a colleague runs the same task through a smaller model inside a well-built module. The module carries a method. Her firm's procedure is attached. The output format is pinned down: a matrix by article, a two-page summary for the board, an action plan with owners and dates. Her draft is the more useful one, by some distance.

Nobody in the room concludes that the smaller model is cleverer. It is not. **It was simply asked a better question, with better material, and told what a good answer looks like.**

Scene 3: the data-residency question

Legal asks a simple question: may client documents be sent to a model hosted in the United States? The honest answer takes weeks. There are contracts to read, data protection impact assessments to consider, and nobody is quite sure what the provider does with the data after the call.

Meanwhile, an EU-hosted model with zero data retention exists — one where the provider has committed to keep nothing after the answer is returned. But nobody has checked whether it is good enough for the task, because the working assumption is that only the largest model will do. So the project waits, for a question that might never have needed asking.

> **If the answer was weak, ask about the setup before you ask about the size.**

Why size is the wrong first question

There is a pattern in those scenes worth naming. In each of them, the team started at the end — with the choice of model — and worked backwards. In Scene 1 the choice was made once for everything. In Scene 2 it was made instead of a setup. In Scene 3 it closed off an option before anyone had tested it.

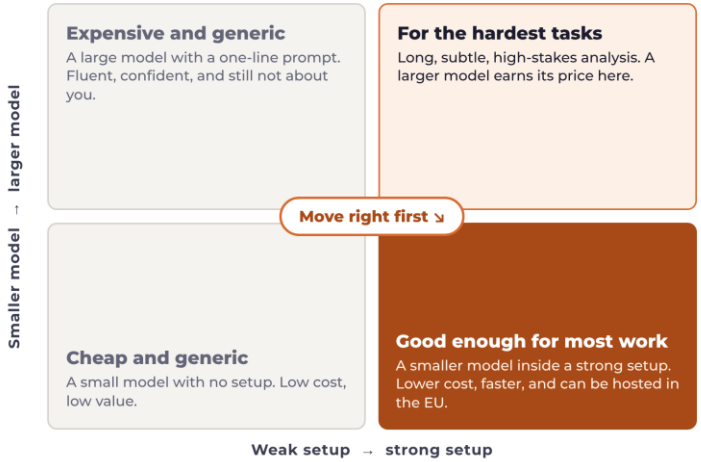
Starting with the model feels natural, because the model is the part of AI that is advertised. But it is a little like hiring for a role by asking only how many degrees the candidate has. Degrees matter. They are not what makes someone good at *this* job, in *this* team, for *this* reader. The briefing does that. The method does that. Knowing what good looks like does that.

There are three costs to getting the order wrong, and they compound.

You pay for capacity you do not use. A summary of meeting notes does not need a model that can reason through a hundred-page contract. Paying for that capacity on every call is like sending every letter by courier.

You blame the wrong thing. When a large model with a thin setup produces a generic answer, the natural conclusion is that "AI is not ready for this" — or that the next, even larger model will fix it. Neither is true. The lesson that should have been learned, *the setup was thin*, is never learned at all.

You narrow your options. If only the largest model will do, then only the handful of providers who offer it will do — on their terms, in their regions, at their prices. That is a concentration of dependence that a regulated firm would question in any other part of its supply chain. We come back to that next week.



Move right before you move up: improve the setup first, and buy size only for the tasks that need it.

Three words, glossed once

- **Frontier model:** one of the largest, most capable models from the major AI vendors, usually the most expensive.
- **Open-weights model:** a model whose weights — the trained model itself — are published, so that others can host and run it.
- **Zero data retention:** the hosting provider's commitment to keep nothing from a request after the answer is returned, and not to use it for training.

Part 2 · Our thesis

> **Model size is one lever among six. It is usually not the first one to pull.**

The thesis

Fix the problem statement, the method, the knowledge, the format and the thinking level first. Then choose the smallest model that does the job well — within your data rules — and keep a larger one for the tasks that genuinely need it.

That is all. It sounds modest, and in a sense it is. But it inverts the way most organisations buy AI. It says that the bulk of the quality comes from things you own and control — your method, your sources, your definition of a good result — and that the model is a component you choose per task, like choosing which colleague to give a piece of work to.

Why we believe it

We hold this view for four reasons. One of them is a measurement, and I will give it the care a measurement deserves.

1. The layers carry the expertise. This is the argument of the last paper, and it is the foundation of this one. The method, the regulation's text, the house format and the reader all sit in the setup, not in the model. A smaller model that is *given* the AMLR's article text does not need to *remember* it. A smaller model that is *told* the reader is the board does not need to guess. Much of what we pay a large model for is its ability to fill gaps from general knowledge. **Close the gaps in the setup, and there is far less left to fill.**

2. A small, early test pointed the same way. On 25 September 2026 we ran ten of ANTON's expert modules, with exactly the same setup, through four model set-ups. The ten tasks were deliberately varied: an FCP gap analysis, a BWRA, an HR policy, a contract review, a board summary, an operational risk assessment, a legal memo, a variance analysis, a press release and a data-subject request in Swedish. The inputs were fictional.

A separate model acted as a blind judge — Claude Opus 5.5 — and scored each answer from 1 to 5 without knowing which model had written it. The two results that matter here:

- **GLM 5.3 Flash**, a small open-weights model, run for us by an EU provider with zero data retention, averaged **2.89**.
- **Claude Sonnet 5**, a well-known model from a major vendor, averaged **2.80**.

The ten GLM answers cost about **$0.025** in total. The ten Sonnet answers cost about **$0.90** — roughly 36 times more.

Read those numbers slowly, and then read the box below, because the caveats matter as much as the result.

What the 25 September test does and does not show

- **Small.** Ten tasks, one run each, one judge — and the judge is itself a model, from the same maker as one of the models being judged.
- **Low effort for everyone.** All models ran at a low reasoning setting. Sonnet 5 at a higher setting may well do better.
- **A strict judge.** No model averaged above 3 out of 5.
- **One failed call.** One GLM call did not run, so its average is over nine tasks, not ten.
- **Fictional inputs, ANTON's own modules.** It measured module outputs on fictional material, not client work.
- **What it shows:** that with the right setup, on these tasks, a small model's answers were comparable with a larger one's, at a small fraction of the cost — and no more than that.

The averages also hide something useful. Task by task, the picture was mixed — which is exactly what you would expect if the setup, not the size, is doing most of the work. The small model scored higher on the FCP gap analysis (4 against 3) and the BWRA (3 against 2). The larger model scored higher on the contract review (3 against 2) and the variance analysis (3 against 2). Several tasks were ties. **Neither model was simply better. Each was better at some things.** That is an argument for choosing per task, not for choosing once.

A fourth model in the same test, a much weaker one, averaged 1.90. On the Swedish data-subject request it did something no compliance officer would forgive: it told the requester, in effect, that a suspicious-activity report existed — a tip-off — and got a date wrong. The judge rejected the answer. I mention it because it is the most important result in the test. **There is a floor below which no setup will save you, and a person must still read what goes out.**

3. A second small comparison pointed the same way. On 11 October 2026 we ran a different kind of test inside Valentin, our review expert. Six synthetic review cases — invented documents with known problems deliberately planted in them — went through Valentin's review method on three models: GLM 5.3, GLM 5.3 Flash and Mistral Large 4.

Model	What it did on the six cases	Cost of the run
GLM 5.3	Completed all six; found every planted issue; every quote verified; the most false findings	About 3 US cents
GLM 5.3 Flash	Completed all six; found every planted issue; every quote verified; fewer false findings	Under 1 US cent
Mistral Large 4	Completed four of six (two ran out of room while reasoning); found every planted issue; every quote verified; by far the fewest false findings	About 26 US cents

What strikes me about that table is not which model "won". It is that **the process carried the quality floor.** All three found every planted issue. Every quote any of them gave was checked by code against the source document, and every one was found. The knowledge layers, the review lens, the pinned-down format and the quote check did that — not the size of the model. Even the cheapest, at under a cent for all six cases, met the bar.

And the models still differed, in ways that matter. The larger model raised far fewer false findings, which saves a reviewer's time. The smaller models raised more, which costs it. And the larger model failed twice when it spent its whole budget thinking — and the process caught both failures and showed them *as failures*, not as empty answers that looked like a clean result.

What the 11 October comparison does and does not show

- Six synthetic cases, one run, all inside one product's review method.
- No independent judge: "found the planted issue" and "quote verified" were checked by the harness, not by a person.
- It illustrates the argument. **It is not a benchmark.** A larger, fairer test — more cases, more than one judge, higher reasoning settings — would be needed before anyone should rely on these numbers.
- Mistral Large 4 is built into Valentin but not yet released there. This was a test, not a live service.

4. Choosing per task is normal professional practice. Nobody sends a partner to take minutes. Nobody asks a trainee to write the legal opinion on the hardest question of the year. We match the person to the task, every day, without thinking about it — and we give each of them a proper briefing first. Matching the model to the task is the same discipline, applied to a new kind of colleague.

Back to the process

In the fourth paper we argued that trust in AI does not come from the model. It comes from the process around it: roles, evidence, checks, sign-off and a record. The second comparison shows why that matters for model choice too, and the lesson goes beyond cost.

If you approach AI by asking *"which model can I trust?"*, you will always be drawn to the largest one, because it fails least often. But it still fails. Every model in both tests invented facts somewhere, or raised false findings, or ran out of room. **No model is trustworthy on its own. A process can be.**

A process that proposes, checks the quotes against the source, asks a second model from a different maker for its view, shows its confidence, says what it did not check — and then lets a person decide, with four eyes where it matters. Inside a process like that, the question changes from *"is this model good enough to trust?"* to *"is this model good enough to propose, given everything that will check it?"* And for a great deal of compliance work, the honest answer to the second question is: a smaller model is.

> **The process carried the quality. The model carried the cost.**

An analogy: staffing the work

Think about how a well-run advisory engagement is staffed. Its quality does not come from putting the most senior partner on every task. It comes from everything that happens before anyone starts writing: an engagement letter that says exactly what must be answered and for whom, a method the firm has refined over years, the client's documents gathered and read, a template for the deliverable, and a manager who reviews every page before it leaves the building. With that in place, a capable associate produces excellent work, engagement after engagement. Without it, even the partner produces something expensive and generic — because the partner, too, is guessing what the client needed.

The partner still matters — for the judgement calls and the hardest questions, the ones that genuinely need a partner. **But nobody serious staffs the partner first and writes the brief later.**

What we are not claiming

A thesis built on thin evidence needs firm limits. Here are ours.

We are not saying a smaller model is as good as a large one in general. We are saying that, on the tasks we tested, with our setup, the results were comparable — and that the setup moved the result more than the size did.

We are not saying size never matters. It does, and here is where, in our experience:

- **Hard, multi-step reasoning.** A legal analysis that has to hold several interacting provisions in mind at once, and follow them through.
- **Very long documents.** Long contracts, or subtle contradictions spread across hundreds of pages.
- **Polished language other than English.** In the September test, the judge noted weaker Swedish from the smaller models on the Swedish task. For client-facing Swedish text, test before you trust.

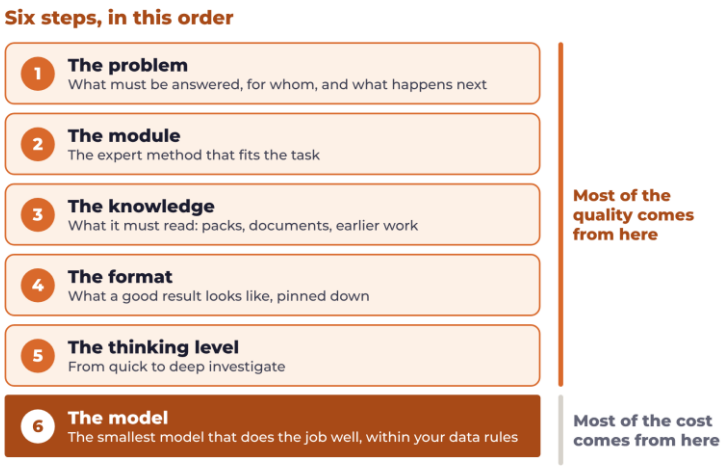
We are not saying setup makes any model perfect. Every model in the September test averaged around 3 out of 5 from a strict judge. Setup lifts every model. It does not remove the need for a person to read, correct and decide.

And we are not saying invented facts are a size problem. Every model, large and small, invented details in the press-release task — because the module gave it room to. That is a setup problem, and the fix is in the setup: a tighter method, a narrower brief, and a rule that facts must come from the material provided.

Part 3 · Our take: decide the model last

A framework for any AI task: six steps, in this order

This is the most practical idea in the paper, and you can use it tomorrow with any tool. Before you choose a model, answer five questions. Then choose the model.



Decide the model last. The first five steps bring most of the quality; the sixth brings most of the cost.

1. The problem. What exactly must be answered, for whom, and what will be done with the answer? *"Is our CDD procedure ready for the AMLR? For the board; it will decide the remediation budget."* That sentence alone rules out half the generic answers a model might give.

2. The module. Which expert method fits? That might be one of ANTON's modules, a focused expert from the family, or simply a method your team has written down. For an AMLR gap analysis, ANTON's gap-analysis module, or Gabriel for a full framework run.

3. The knowledge. What must it read? Which knowledge pack, which of your documents, which earlier work? The AMLR pack and the article text; the firm's own CDD procedure; last year's assessment.

4. The format. What does a good result look like — exactly? Pin it down: a gap matrix by article, a two-page board summary, an action plan with owners and dates. **This step does more for quality than most people expect**, because a model that knows what a good answer looks like spends far less effort guessing.

5. The thinking level. How hard should it think? ANTON offers a range — quick, think, think hard, investigate, plan first, deep investigate. A board deliverable deserves "think hard", and "plan first" so that the plan can be seen before the draft. A meeting summary does not.

6. The model. Only now: the smallest model that does this task well, within your data rules. A mid-sized model hosted in the EU for the draft; a model from a different maker for the second opinion; a larger model for the hardest legal points.

The order matters. Steps one to five are where most of the quality comes from. Step six is where most of the cost comes from. Teams that start at step six pay for size and still get generic answers. Teams that start at step one often find they need less size than they thought.

Reasoning settings are part of the setup

Some models now reason by default, and that changes how they must be set up. One small model we use reasons at maximum effort unless told otherwise, which slows it down and costs more than the task needs. Mistral Large 4 reasons by default too, and needs both an effort level and room for its thinking: given too small an output limit, it spent everything on reasoning and returned nothing. Choosing a model includes choosing how it should think — and checking what it does if nobody says.

The cost and speed picture

Prices differ by an order of magnitude or more. Between the smallest models we use and the frontier models from the large vendors, the price per unit of text can differ ten-fold or more, and even between our own small and mid-sized choices there is a clear step. I will not print a price list: prices change too often for that, and a paper that says *"about 36 times less, on 25 September 2026"* ages better than one that quotes prices per million tokens. But the ratio is the point. If a task is done well by a model that costs a thirtieth as much, every routine call to the largest model is money that bought nothing.

Speed follows size and settings. In our September test the smaller model was often faster to its first words. Reasoning-heavy settings slow every model down. For a reviewer waiting on a draft, that matters more than it sounds.

A cap turns cost into a control. Cost should never be a surprise at the end of the month. Across the family, spending is capped before the call, not discovered after it:

- **ANTON** caps priced calls per installation and per user per day, and reserves the worst-case cost of each call before it is sent. If the reservation would break the cap, the call does not go.
- **ALEXANDER** has four caps: per run, per engagement per month, per person per day and per server per day.
- **Each public showcase** has its own daily cap.

Model choice per task is built in. In ANTON, the small "utility" calls — summarising, extracting — run on a mid-sized model by default, so they are not quietly promoted to the largest one just because the user chose it for the main work. The user can pick a model per session. And an endpoint can be limited to an allow-list of models, so a team rule such as "small models for summaries" can be enforced, not just hoped for.

What we have built

Multi-model by design

ANTON works with Anthropic's Claude (through the API, or through the subscription engine), OpenAI, Azure OpenAI, Google Gemini, Mistral, local models through Ollama, and any endpoint that speaks the common OpenAI-compatible protocol — routing services such as OpenRouter included. **The seven layers work the same way on all of them.** Your method, your knowledge packs, your format and your thinking level do not change when the model does. That is what makes it possible to choose per task at all.

Endpoints with guardrails

An endpoint in ANTON is not just an address. It can carry a **provider pin** — for example, "only EU providers with zero data retention, and no data collection" — an **allow-list** of the models that may run on it, an **output limit**, and the **prices**. Every priced call is written to a spending ledger and checked against the caps before it is sent.

Each showcase runs only the models its operator allowed

The public showcases run on GLM 5.3 by default, through an EU provider with zero data retention and no data collection. Gabriel, Valentin, ALEXANDER and ARIS all work this way on their showcases, with fictional data. Valentin goes further: it pins each model to its own provider in code, and refuses any request whose pin is not exactly right.

GLM is an open-weights model. In the family it is run by an EU provider; that is a statement about *where* it runs, not *who made it*. That distinction — hosted in the EU versus made in the EU — is important enough to have its own place next week.

Independence as a model choice

Where a second opinion matters, the family deliberately uses a model from a *different* maker. Valentin refuses to review a document with the same model that wrote it. ALEXANDER's showcase uses GLM 5.3 as its main model, Kimi K2.6 for the second opinion, and DeepSeek V4 Flash as a reviewing model that checks the work before a person sees it. Different makers, trained differently, make different mistakes — which is exactly why one can catch the other. And in every case, a person decides.

Evaluation tools, so you can check rather than assume

ANTON has an evaluation tool that first produces a plan and a cost estimate without spending anything, and then runs under a hard spending cap. It is what we used for the September test. Valentin has its own evaluation on a fictional worked example, and since 11 October a comparison tool for the three models in the table above. The point is not our particular numbers. **The point is that "which model for which task" should be a question you test, cheaply and on fictional data, rather than a question you answer by reading a league table.**

Picture: ANTON's model settings: provider pin, allow-list and daily cap (showcase, fictional data)

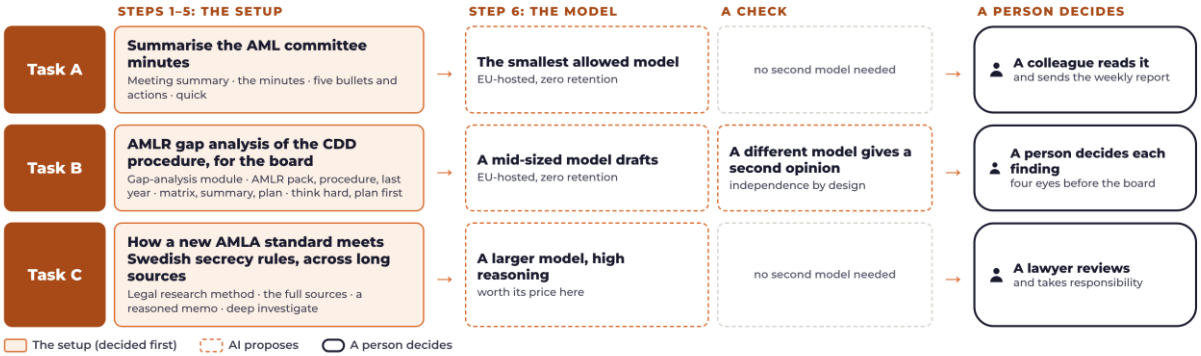
Status at a glance

- **Live** (public showcases, fictional data): Gabriel, Valentin, ALEXANDER and ARIS on GLM 5.3 through one EU provider with zero retention; ALEXANDER's second-opinion and reviewing models; a daily cap per showcase.
- **Built, not released:** Mistral Large 4 in Valentin (tested on 11 October, not live).

Part 4 · A worked example

Three tasks, three model choices

Exempelbolaget AB is a fictional Swedish payment institution. On one Monday morning, its compliance team has three AI tasks in front of it. The team uses the same six steps for each — and ends up with three different models. **The setup is never skimped. The size varies with the task.**



Same six steps, three different models. Spend on setup every time; spend on size when the task earns it.

Task	The setup: steps 1 to 5	Size	The model, the check, and who decides
A. The committee minutes	**Problem:** a half-page summary for the weekly report. **Module:** meeting summary. **Knowledge:** the minutes. **Format:** five bullets and the actions. **Thinking:** quick.	Small	The smallest model allowed, hosted in the EU with zero retention. A colleague reads it before it goes out. Cost: a fraction of a cent.
B. The AMLR gap analysis	**Problem:** is the CDD procedure ready for the AMLR? For the board. **Module:** ANTON's AMLR gap analysis, or Gabriel. **Knowledge:** the AMLR pack and text, the procedure, last year's assessment. **Format:** gap matrix, two-page board summary, action plan. **Thinking:** think hard; plan first.	Mid-sized	A mid-sized model hosted in the EU drafts; a model from a different maker gives the second opinion. A person decides each finding, with four eyes before the board. Cost: cents.
C. A hard legal question	**Problem:** how a new AMLA standard meets Swedish secrecy rules, across long sources. **Module:** a legal research method. **Knowledge:** the standard and the Swedish law and guidance, in full. **Format:** a reasoned memo with sources, assumptions and what was not checked. **Thinking:** deep investigate.	Larger	A larger model at a high reasoning level. A lawyer reviews it and takes responsibility. Cost: a few dollars — and worth it.

What the example shows

Look first at what is the same in all three columns. Every task gets a clear problem statement, a method, the right material, a pinned-down format and a deliberate thinking level. None of them starts with a blank box and a hopeful sentence. **The setup is the constant.**

Then look at what changes. Task A does not need a large model; it needs to be fast and cheap, and the setup makes a small one entirely adequate. Task B is board work, so it gets more thinking, a second opinion from a model by a different maker, and a person who decides every finding — but the draft itself does not need the largest model, because the method, the regulation's text and the format have done most of the work. Task C is different in kind. It is long, subtle and high-stakes, with provisions from two legal systems that have to be held in mind together. **That is exactly the kind of task where a larger model earns its price**, and we would not hesitate to pay it. The team spends where the difficulty is, not everywhere at once.

And notice where the person is. In every column, the last step is a human one: a colleague, a decision-maker with four eyes, a lawyer. Model choice never changes that. **A smaller model is good enough to propose inside a process that checks it. It is never good enough to decide** — and neither is a larger one.

Part 5 · From here

What this means for you on Monday

You can start with any tool you already have. Five things to do this week:

1. **Use the six steps for your next AI task, in order.** Write one line for each step — the problem, the method, the knowledge, the format, the thinking level — before you choose a model.
2. **Pin down the expected result.** Write what a good output looks like: the sections, the length, the reader, the evidence it must show. It will do more for quality than a bigger model.
3. **Run a small comparison yourself.** Take three of your recurring tasks, on fictional or public data, and run them on two models with the same setup. Ask a colleague to score them without knowing which is which. You will learn more in an afternoon than from any league table.
4. **Set a model per task type, and write it down.** Small for summaries and mapping; mid-sized for drafts, with a second opinion from another maker where it matters; large for the hardest analysis. Make it a team rule, next to your data rules, and let your tools enforce it with allow-lists — so good practice does not depend on good memory.
5. **Cap the spend.** Set a daily cap per person and per team. Review the spending monthly and ask two questions: where did the money go, and did it buy quality?

Next week: your knowledge, your side

If models are this interchangeable, what must stay yours? Next week: models are utilities you switch; your business information, your knowledge and your way of working must stay on your side, in a format you control. We look at why Europe should build and use its own AI, why *hosted in the EU* and *made in the EU* are two different questions, and how a client's signed permission can name the exact model allowed to read its data.

A closing thought

There is something freeing in this argument. If quality came mainly from size, most organisations could only wait, and pay, for whatever the largest vendors release next. If it comes mainly from setup, most of it is in your own hands.

Our small tests point the same way as everything else we see in daily work: **the setup moved the result more than the size did.** That is reason enough to change the order in which we make decisions — the problem first, the model last.

Spend on setup every time. Spend on size when the task earns it.

— Daniel Bardun

See it, try it, talk to us

- **See it:** the ANTON family's public showcases, with fictional data, at ai-portal.futurechain.eu.
- **Try it:** ANTON is open source — github.com/altspace-hub/anton_desktop.
- **Read the series:** all papers in the ANTON family whitepaper series, and the earlier whitepapers, are linked from the portal. [Daniel: confirm where the PDFs will live.]
- **Talk to us:** Daniel Bardun, on LinkedIn — questions, disagreements and "this is where AI stopped working for us" are all welcome.

Important notes

Product facts were checked on 11 October 2026; statuses are summarised in "Status at a glance", and both tests in Part 2 are small, internal and run on fictional or synthetic data — not benchmarks. The public showcases use fictional or public data only, and no client uses them. ANTON is open source (EUPL); the other family products are not. The products support your work and help you evidence it; they do not make anyone compliant, and nothing in this paper is legal advice.