BOMBADIL LLC · FIELD GUIDE

Putting a Dollar Figure on Cyber Risk

Field notes from a security program that runs on quantified risk

Brandon Karpf

Bombadil LLC · brandon@bombadil.pro · July 2026

Feynman Technology Solutions is an invented name for a real company: a venture-backed deep-tech infrastructure firm, roughly fifty people, selling into government and large-enterprise buyers. Its figures are disguised (rounded and rescaled), but every ratio, every ordering, and every conclusion survives the disguise. Shared with the client's permission.

1. The map on the wall

Nearly every security briefing I have sat through ends at the same slide. A five-by-five grid. Red in the top-right corner, green in the bottom-left, a scatter of numbered dots in between. The board studies it. Someone asks the only question that matters: how bad is it?

And the honest answer, reading off that slide, is: it's red.

Nobody prices anything in red. Your CFO cannot reserve against red. An underwriter cannot quote red. A board cannot weigh red against the four other risks on its agenda, each of which arrives with a dollar figure attached, because finance, legal, and supply chain all learned to quantify decades ago. Security walks into that room holding a coloring book.

Compare how we handle water. When you buy a house near a river, nobody hands you a color. You get a flood map: a one percent annual chance the water tops your first floor, and a table for what a foot of water does to a wood-frame house. Probability and dollars. You can buy insurance against that. You can move the electrical panel upstairs against that. You can do neither against orange.

Cyber risk can work like the flood map. The method is called cyber risk quantification (CRQ), the output is a loss exceedance curve, and I run an entire security program on one. The seat I watch this from: security executive at Feynman Technology Solutions, hired to build the program from close to zero, with a board that thinks in runway and a sales team blocked on customer security questionnaires. That seat does not tolerate theater.

Feynman's curve puts expected annual loss at about $2.6M, with a four percent annual probability of an eight-figure incident. And it surfaced the number that reshaped every conversation after it: a bit more than half of that expected loss cannot be removed at any level of spend.

Three numbers. They now run the security budget, the insurance negotiation, and the business case for the program itself. These notes cover how we built them, what they changed, and the system diagram we'll assemble piece by piece as we go.

2. The method, in one sitting

Strip CRQ to its frame and it fits in five questions. I have walked non-technical founders through this in under an hour.

How often does this happen to companies like us? The outside-in rate. You do not guess it; you look it up. Verizon's Data Breach Investigations Report, Cyentia's IRIS dataset, and IBM's Cost of a Data Breach report publish incident frequencies by sector and size. For most mid-sized companies the annual rate of a material event lands between 5 and 15 percent. Sector and size both matter. Cyentia's IRIS 2025 study puts a typical firm's annual chance of a significant incident at 9.3 percent,

but the eight sectors it charts run from about 4 percent (entertainment) to about 11 percent (manufacturing), and per firm, the largest corporations log far more incidents than the smallest. Using the all-industry average for your company is like using the national crime rate to price a lock in downtown Baltimore.

How much better are we than the base rate? The inside-out adjustment, and the question every security team flunks by flattery. Answer with evidence. Enforced MFA everywhere, around-the-clock detection coverage, 72-hour critical patching, a tested response plan: claim a few points. And credit culture where it shows up in the numbers, because it does show up: a team that believes in the work patches faster, clicks less, and reports the weird email instead of deleting it. Culture moves the human-oriented hazards hardest, and it earns its adjustment the same way everything else does, as measured behavior (patch latency, phish-simulation failure rates, offboarding speed) rather than as a slide that says "strong security culture."

If it happens, what's the floor? The best-case bad day. Forensics alone bills six figures. Legal bills five. Notification runs $5 to $15 per affected individual. For a mid-market company the realistic minimum sits between $250K and $1M. Zero is not on the menu.

If it happens, what's the ceiling? The 95th-percentile day: ransomware across production, regulators on the phone, customers reading about you. Sum the worst realistic case across response costs, fines, downtime, and litigation. The number should be uncomfortable. Comfortable numbers belong on heat maps.

What loss actually matters? The material threshold. That number belongs to finance, and your CFO already knows it: the figure that triggers board notification or disclosure, often near one percent of revenue.

Five inputs. Then a Monte Carlo simulation runs the year 500,000 times: in each simulated year, events either happen or don't at your combined frequency, and each one that happens draws a loss from a lognormal distribution fitted between your floor and ceiling. Lognormal, in kitchen terms: most incidents cluster small, a few run catastrophic, and the shape has a long right tail. The output is the loss exceedance curve, which answers one question at every dollar amount: what is the annual probability we lose more than this?

Three numbers fall straight out of the curve. Expected annual loss (EAL), the probability-weighted yearly cost of cyber risk. Combined annual probability, the odds anything material happens. And the probability of exceeding the material threshold, which is the one your CFO will remember.

A spreadsheet does all of this. Half a million simulated years, one chart. The five-variable version is a weekend project, and I mean that as a compliment: the barrier to entry for pricing your cyber risk is one honest afternoon, not a consulting engagement.

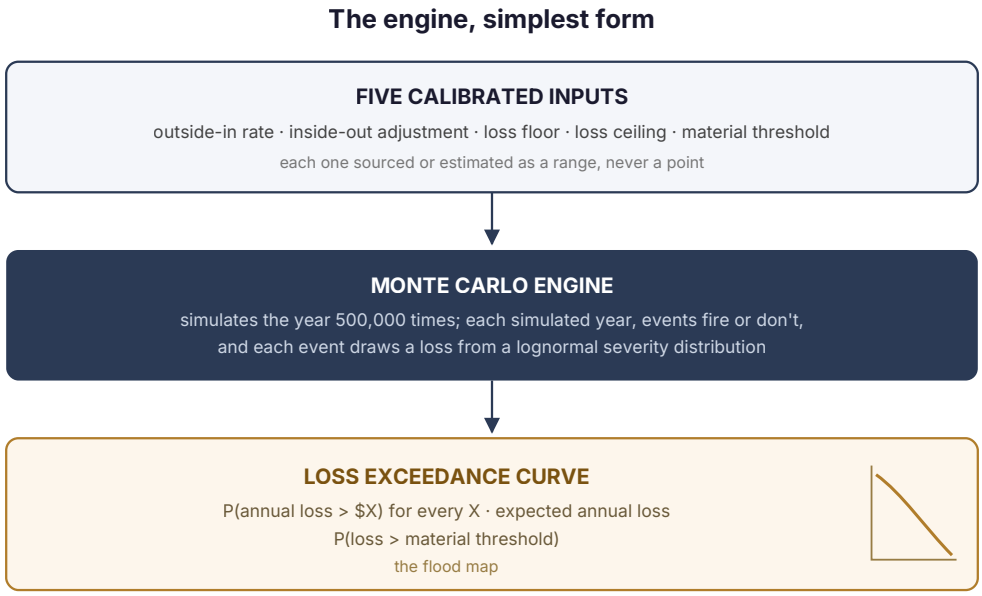


Figure 1. What you're looking at: the whole method at its simplest. Five calibrated inputs at the top, a Monte Carlo engine in the middle simulating the year half a million times, and the loss exceedance curve at the bottom. Every section that follows adds one layer to this diagram until the full production system stands at the end of the piece.

3. When one curve stops being enough

The five-variable model treats all cyber weather as the same storm. One probability, one loss distribution. For a first baseline that's exactly right. But the moment the board asked "so where does the risk actually live?", one curve had no answer. A ransomware event and a stolen laptop draw from very different distributions, and a model that averages them can rank nothing.

The step up is decomposition: split the risk into hazard categories, give each its own annual frequency, its own severity distribution, and (hold this thought) its own irreducible floor. Each fires on its own clock in the simulation; statisticians call the clock a Poisson process, and you never need to say that word again. The aggregate curve sums the small storms across 500,000 simulated years.

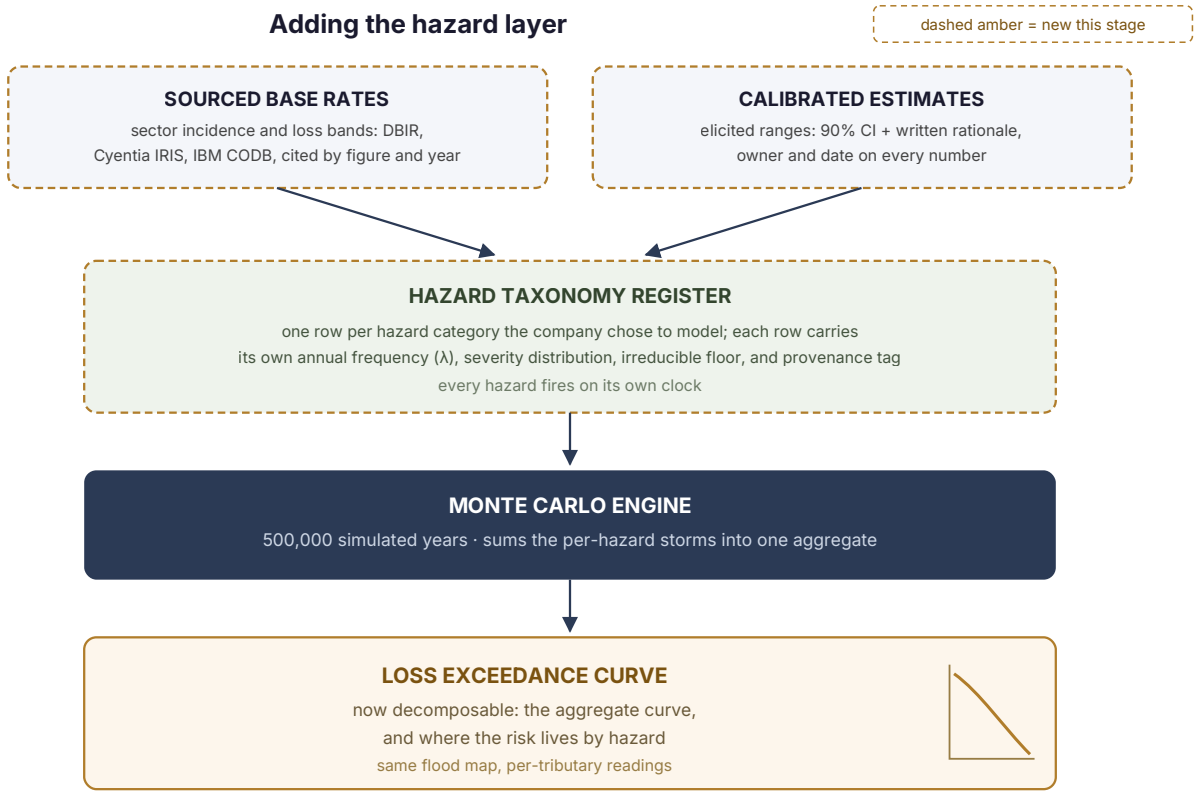
Which categories? That decision belongs to you, and it deserves an hour of real argument, because the categories you model are a statement about the losses you believe you face. The menu below is the full set I consider defensible, cross-walked against the taxonomies the loss data actually uses (DBIR/VERIS action categories, MITRE ATT&CK, ENISA's threat landscape, FAIR's threat communities, NIST SP 800-30):

Candidate hazard category	What lands in this bucket
Credential abuse	Stolen, phished, or misused legitimate access; the classic "valid accounts" breach
Vulnerability exploitation	Unpatched or zero-day software flaws exploited from outside
Supply chain & third-party software	Compromise arriving through vendors, dependencies, or build pipelines
Malware & ransomware	Malicious code as the primary loss driver, extortion included
Social engineering	Phishing, pretexting, business email compromise; the human as the vector
Insider misuse	Authorized people doing unauthorized things, malicious or negligent
Misconfiguration	Exposed buckets, default credentials, permissive policies; self-inflicted openings
Human error	Misdelivery, lost devices, fat-fingered commands; mistakes without malice
Availability & denial of service	Attacks on uptime itself
Physical	Theft, intrusion, tampering, hardware interdiction
Cryptographic failure	Key compromise, broken or misused crypto, harvest-now-decrypt-later exposure
Fraud & payment abuse	Financial-process manipulation, invoice fraud, account-takeover monetization
Third-party service outage	A vendor's downtime becoming your loss, no attacker required
Information manipulation	Disinformation, brand impersonation, data-integrity attacks

How many should one company model? Enough to separate losses that behave differently, few enough that a real incident lands in exactly one bucket. The working test: could an incident commander or an insurance adjuster assign an event to a single category without calling a meeting? Somewhere between eight and fourteen buckets passes that test for most companies; the right subset depends on what you sell, what you hold, and who wants it. Document what you deliberately leave out and why. An omission with a written rationale is a scoping decision; a silent omission is a hole.

The discipline that keeps a taxonomy honest is provenance. Every number in the model carries a tag: SOURCED or CALIBRATED. Sourced numbers come from published data, cited by figure and year (a sector base rate off Cyentia's IRIS incidence tables, for instance). Calibrated numbers are

estimated ranges, and estimating them well is a craft with a literature; Hubbard and Seiersen's How to Measure Anything in Cybersecurity Risk is the manual, and a room full of calibrated human experts is the classic instrument. My practice is a compressed version of the same protocol: I elicit each estimate independently from two frontier AI models from different vendors, set those against my own expert assessment and a third party's, and reconcile the four by hand into a 90 percent confidence interval with a written rationale. Where the sources converge, the interval tightens. Where they fight, it stays wide and says so. Nothing enters the model naked.



> **Figure 2. What changed:** the single input block from Figure 1 has split in two (sourced base rates on the left, calibrated estimates on the right), and a new register sits between the inputs and the engine: the hazard taxonomy, one row per category the company chose to model, each row carrying its own frequency, severity, floor, and provenance tag. The curve at the bottom is now decomposable by hazard.

4. The two ideas that changed how the company behaves

The machinery above is plumbing. Two ideas that fall out of it did more work at Feynman than any control we shipped.

The floor. Sum the per-hazard residual minimums and you get the loss the program cannot remove: the zero-day nobody has patched, the nation-state actor with a budget, the trusted insider, the

genuinely novel attack. At Feynman, that floor computes to about $1.4M of the $2.6M expected annual loss. More than half.

Sandbags do not move the river. You can stack controls to the ceiling and the river still runs past the house; the floor is the river. The moment the board saw the split, the program's objective changed shape: stop promising zero, and compress the roughly $1.2M addressable band down toward the floor. That reframing killed two bad conversations permanently. Nobody asks "why do we still have risk after all this spend" anymore, and nobody proposes buying the risk to zero, because the model shows zero was never for sale. What the floor is for, it turns out, is insurance. More on that in section 9.

Status credit, or: work in progress isn't protection. Every finding in Feynman's risk register carries a status, and status sets the credit the curve grants. A finding merely identified earns nothing. A finding in remediation earns a tenth of its loss reduction. Only remediated earns full credit.

A pallet of sandbags in the driveway holds back no water.

This rule sounds harsh and prices reality. Half-deployed MFA is not half protection; the attacker logs in through the half you haven't covered. Midway through the program's first year, the register held a few dozen calibrated findings, most still at identified, a handful in remediation, none complete. Total credit earned against the baseline: about $60K, on a $1.2M addressable band. The curve sat nearly flat, and the flat line went to the board as-is, because it was the true state of the program: skeleton built, little finished. A dashboard that can embarrass you is a dashboard you can trust, and the embarrassment did more for remediation velocity than any escalation email I have ever written.

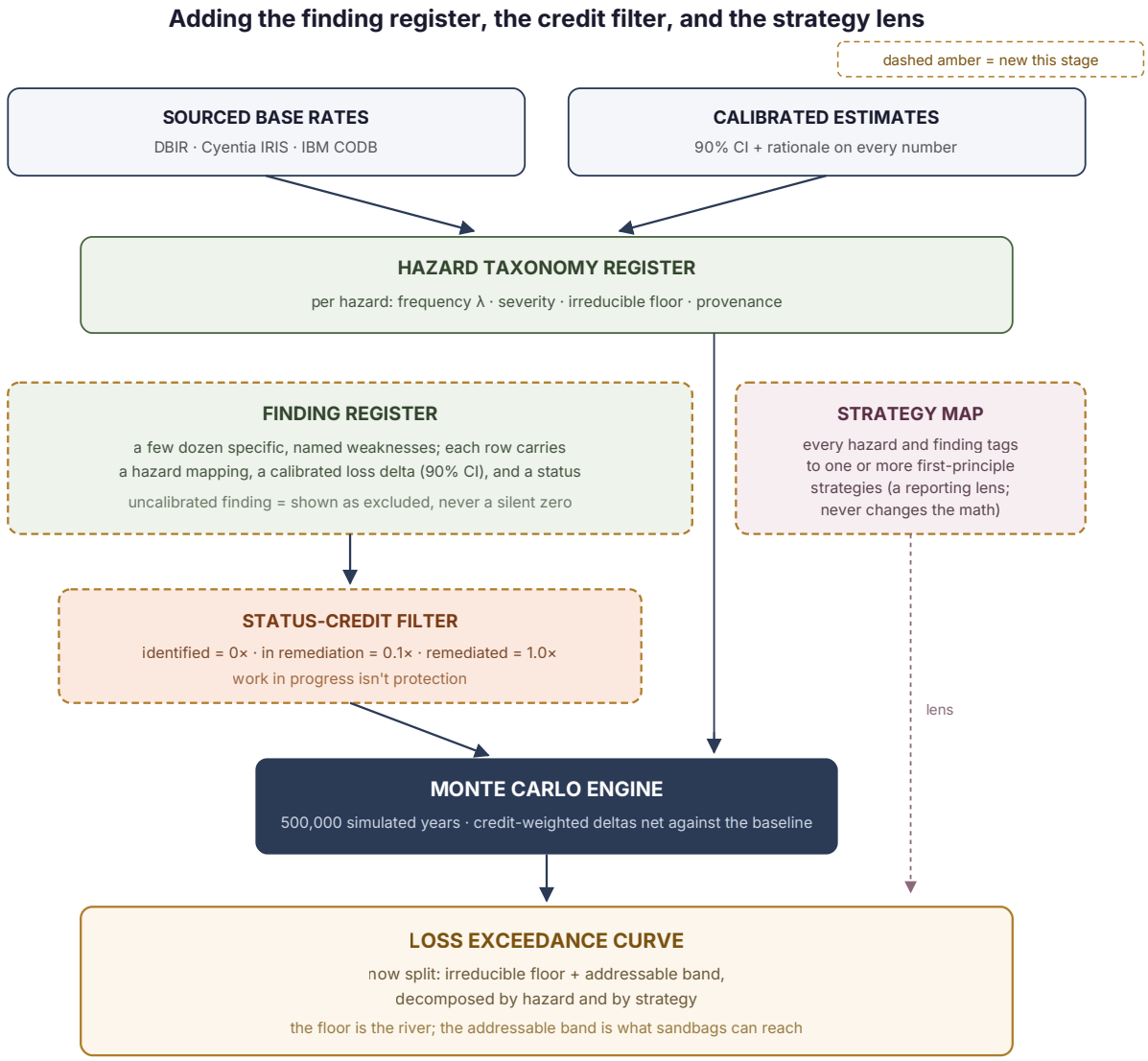


Figure 3. What changed: two new components between the hazard register and the engine. The finding register holds the company's specific, named weaknesses, each mapped to hazards, priced with a calibrated loss delta, and tracked by status; the status-credit filter converts status into earned credit (0x / 0.1x / 1.0x) before anything reaches the math. On the right, the strategy map tags every hazard and finding to one or more first-principle strategies; it's a reporting lens and never changes the curve. The output now shows the floor/addressable split.

5. The strategy lens: designing the program, not just measuring it

A curve that only measures is an audit. The step that turns it into a design tool came from Rick Howard's Cybersecurity First Principles: A Reboot of Strategy and Tactics (Wiley, 2023), a book I edited and have been field-testing ever since. Rick's argument: a security program should reduce the probability of material impact through a small set of first-principle strategies rather than a shopping list of controls. The strategies we run the lens on: zero trust, intrusion kill chain prevention,

resilience, risk forecasting, and automation, straight from the book, with workforce development alongside them as a sixth.

The mechanics are simple and load-bearing. Every hazard, every finding, every remediation, every control tags to one or more strategies. Sometimes one: an MFA rollout is zero trust, full stop. Sometimes several: release signing earns credit under both resilience and kill chain prevention. Roll the tags up and the curve decomposes a second way: not just where the risk lives by hazard, but what strategy would remove it, priced in dollars of addressable expected loss.

That decomposition is the program design. Instead of asking "which tools should we buy" (the vendor's question), you ask "which strategies hold our addressable dollars, and what's the cheapest way to buy them down" (the owner's question). At Feynman, the top three strategies held over three-quarters of the addressable band, two of them at near-zero maturity; that single readout sequenced eighteen months of work. You buy the tools and build the processes under the strategies that carry your exposure, at the best return on invested capital, and you stop funding strategies whose headroom you've already spent. The lens never touches the math (Figure 3 shows it hanging off to the side, on purpose), but it converts a measurement instrument into a capital-allocation instrument.

Risk forecasting itself is one of the six, which produces a pleasant recursion: the curve is the program auditing the program. Rick would call that a first principle doing its job.

6. The worked example: Feynman's baseline

Feynman's current curve, disguised per the note up top:

Metric	Value
Expected annual loss	~$2.6M
Irreducible floor	~$1.4M (over half)
Addressable	~$1.2M
Combined annual incident probability	~10%
P(loss > $10M)	~4%
P(loss > $50M)	~1%

The exceedance ladder reads like the flood map it is: roughly an 8 percent annual chance of losing more than $1M, 5 percent above $5M, 4 percent above the $10M material threshold, about 1 percent above $50M. Small floods come often; the big one is rare, real, and priced.

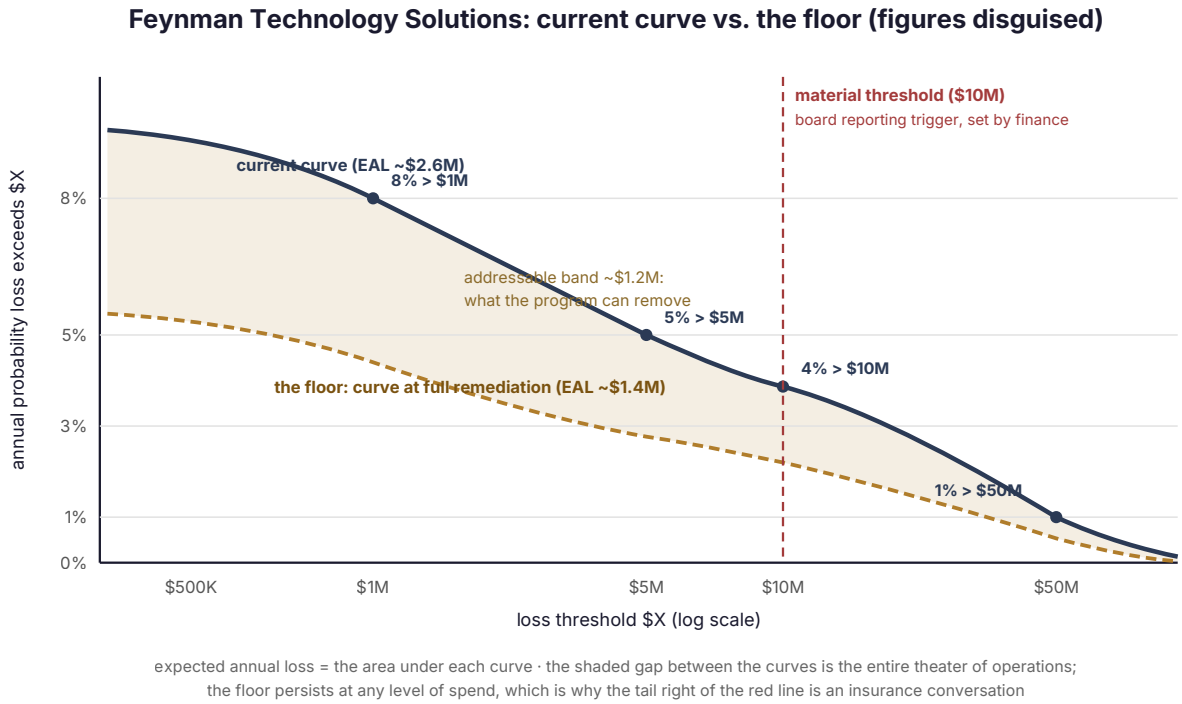


Figure 4. What you're looking at: the actual output artifact, disguised, with both curves drawn. Loss thresholds run left to right on a log scale; the y-axis is the annual probability of exceeding each threshold. The solid line is today's curve (EAL ~$2.6M). The dashed amber line beneath it is the floor: what the curve becomes at full remediation (EAL ~$1.4M). The shaded band between them is the ~$1.2M addressable band, the program's entire theater of operations. The red dashed line is the material threshold finance set ($10M); everything to its right is the tail, which is insurance territory.

Decomposed by hazard, three categories carry roughly 60 percent of Feynman's expected loss: vulnerability exploitation (~$600K), credential abuse (~$550K), and supply chain (~$450K). That concentration is the remediation priority list, ranked in dollars. The findings mapped to those three hazards are where completing work moves the most money, and every point of combined incident probability the program removes is worth about $250K a year at the current curve.

One register entry, to make it concrete. A product-security review turned up an access-control gap between customers in one product's early design. As a finding it reads mild; calibrated as a cross-tenant exposure path in a product headed for multi-tenant deployment, it carries a 90 percent confidence interval on its loss contribution, a hazard mapping, strategy tags, and a status that earns credit only when the fix ships. The register holds a few dozen of these, each calibrated the same way, and the curve is nothing more than their disciplined sum plus the sector base rates.

That is the entire trick. No oracle. A list of specific, priced, tracked findings, and arithmetic.

7. Operating the model: the part everyone skips

A risk model built once is a brochure. The operating discipline is what makes it an instrument, and the discipline turns out to be light if you build it in from the start.

The rule that carries everything: you never edit the curve. You maintain the registers, and the curve recomputes from them. The finding register drives; everything else is a reporting lens over it.

Updates come in two layers, and the split is what keeps the operating cost near zero. Layer one runs free and frequent: a finding changes status, and the curve moves on the next rebuild. No re-estimation, no meeting. This happens as a side effect of normal program work; close a finding in the weekly sync and the model already knows. Layer two fires rarely: a new or re-assessed finding needs a calibrated loss delta before it can contribute, which costs one elicitation pass and a written rationale. That's the only moment a human feeds the model, and it comes with a hard rule: a finding with no calibrated delta contributes nothing and renders as excluded. Never a silent zero, never a fabricated one.

Three habits round out the discipline. The weekly dashboard freezes a dated snapshot, so the trend line accumulates whether or not anyone is watching. The simulation seed is pinned to the inputs, so the headline moves only when posture moves, never because the dice rolled differently on Tuesday. And the full calibration gets redone every six months, because base rates drift and last year's honest estimate is next year's stale one.

Total operating cost between recalibrations: minutes per week. The model rides along on work the program does anyway.

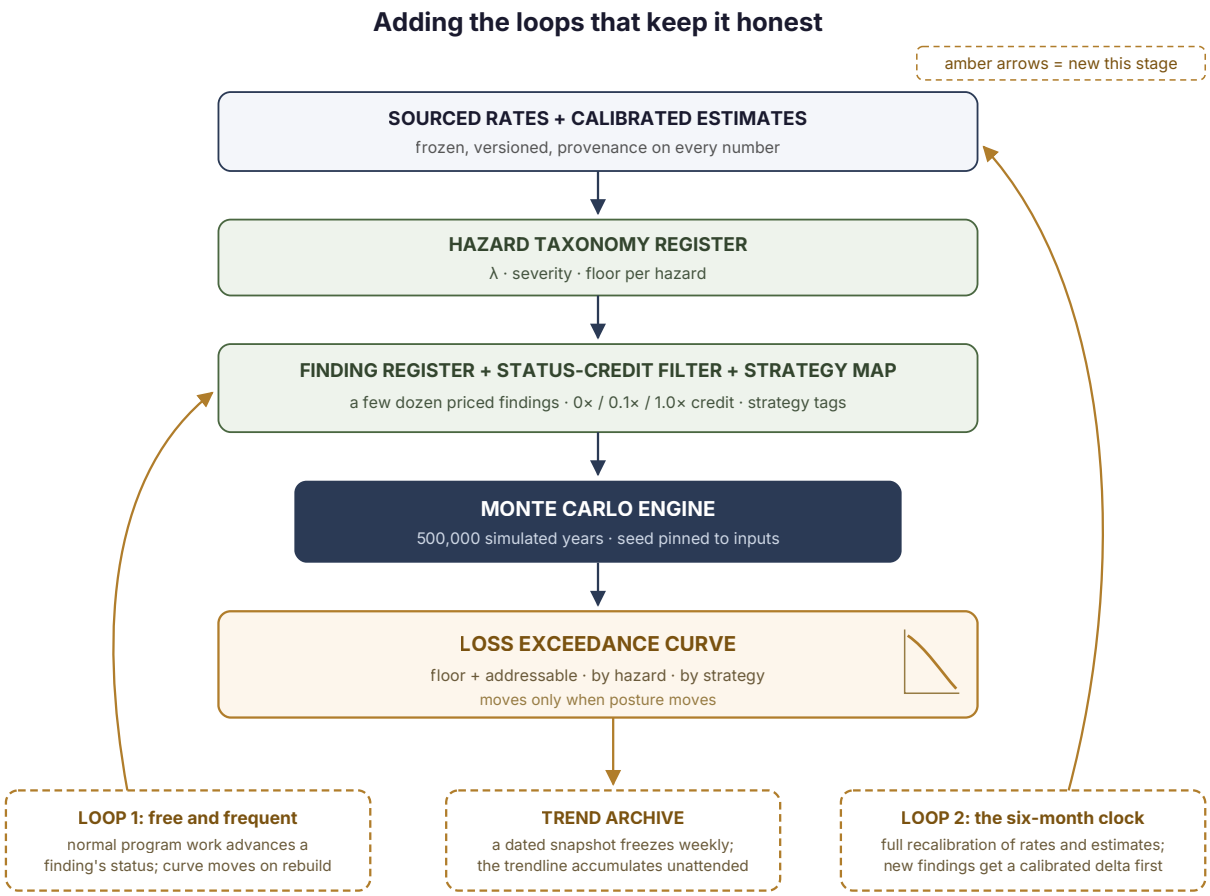


Figure 5. What changed: the system from Figure 3, compressed, plus the three loops that keep it alive. Loop 1 (bottom left) feeds status changes from normal program work back into the finding register, free and frequent. Loop 2 (bottom right) is the six-month recalibration clock feeding the input layer. The trend archive (bottom center) freezes a dated snapshot every week so the curve's history accumulates unattended.

8. Decision one: ranking every security dollar

The first thing the curve got used for was the budget, and the logic compresses to one sentence: rank every proposed dollar against the ~$1.2M of addressable loss, because the floor doesn't care about your tooling.

The strategy lens from section 5 does the first cut. Feynman's addressable band, decomposed by strategy:

Strategy	Addressable headroom	Share
Zero trust	~$350K	30%
Kill chain prevention	~$290K	25%
Resilience	~$260K	22%
Workforce development	~$140K	12%
Automation	~$120K	10%
Risk forecasting	~$12K	1%

Three strategies hold over three-quarters of the recoverable loss, and when we first built this table, two of them sat at near-zero maturity. That is the shape of a company that stood up governance and identity first and had not yet built detection or recovery. Risk forecasting's tiny headroom is by design: the curve measures risk, it doesn't remove it.

Within each strategy, each candidate control gets scored by the addressable loss it removes divided by its annual cost. Walk down the ranking and stop where the next dollar stops beating itself. At Feynman, that walk split the program into two tiers with a bright line between them.

Tier one costs about $350K a year and removes roughly $800K of expected loss, two-thirds of the addressable band. It is aggressively unglamorous: single sign-on, device management, a password manager, backup and restore, software supply-chain hygiene, phishing simulation, and a block of controls the company already owned but hadn't turned on. Every line returns more than a dollar per dollar. There is no vendor booth at RSA for "turn on the things you already bought," which is a shame, because it beat every product pitch we received.

Tier two costs about $1M a year and buys the rest: managed detection and response, endpoint response tooling, centralized logging, vulnerability management, email security, offensive testing. On pure expected-value math at Feynman's size, every one of these lands below the break-even line. And we're buying most of them anyway, for two reasons the ranking forces into the open. The tail: expected annual loss is a mean, and a mean under-weights the rare event that ends the company; the detection-and-response stack barely moves the average while it collapses the worst-case ceiling, so it gets sized against the far end of the curve in the same breath as the insurance decision. And the evidence: the audits the sales pipeline requires demand exactly these controls as proof. The curve confirms they're worth doing and tells you their honest place in the sequence: after the cheap wins are banked, on the audit calendar's clock.

The ranking also says no, which is where it earns its keep. One quote made the decision for us: a heavily marketed monitoring add-on priced at about $120K a year, mapped cleanly to a single hazard slice carrying about $30K of expected annual loss even if we left it completely uncontrolled.

Four dollars a year to chase one. We passed, wrote the rationale into the register, and set the trigger that would reopen the question (material growth in that hazard's exposure). The vendor still emails me quarterly, and the answer is still the curve's.

Workforce investment runs through this same lens differently, and deliberately. Training, security champions, the leadership bench: on a per-control expected-value test, people always lose to software, because a person costs more than most hazard slices. That test misreads what the money buys. Workforce development is one of the six strategies precisely because it's the one that makes the other five execute; a purchased stack with nobody skilled enough to run it removes nothing. So Feynman budgets workforce as a strategy allocation (its ~$140K of direct headroom plus its share as the operating multiplier on everything else), and judges it at the program level, at return on invested capital, never as a line item competing with a password manager.

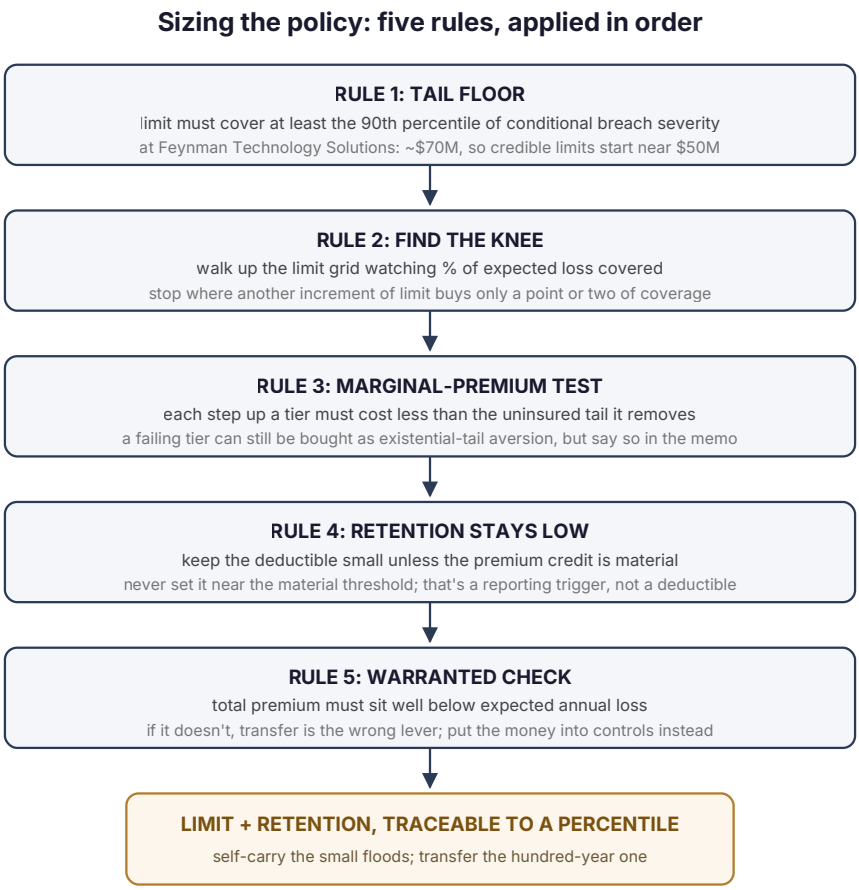
One more number the ranking surfaced: the board had approved a tooling budget roughly half the full program's ~$1.4M annual run rate. The curve turned that gap from a vague ask into a priced decision with a ranked list behind it.

9. Decision two: sizing insurance to the curve

Feynman carried general liability and no cyber policy. The standard way to fix that is a broker's rule of thumb keyed to revenue. We sized it off the curve instead, because an insurance limit is a bet on the loss distribution, and we had one with provenance.

The curve hands the underwriting conversation four quantities per candidate limit L: the annual probability a loss exceeds L; the expected dollars per year that punch through the cap and land back on the company; the share of expected annual loss the policy would absorb; and the conditional severity percentiles, meaning if a breach happens, how bad at the 50th, 90th, 99th.

Five decision rules turn those quantities into a policy, applied in order; Figure 6 walks them as a flow. The tail floor comes first: the limit must cover at least the 90th percentile of conditional breach severity, which at Feynman computes to roughly $70M, so credible limits start near $50M. Then find the knee: walk up the limit grid watching the share of expected loss covered, and stop where another increment buys only a point or two. The marginal-premium test prices each step: a tier clears if its added premium costs less than the uninsured tail it removes, and a tier that fails can still be bought as existential-tail aversion, stated out loud in the memo. Retention stays low, and never near the material threshold, which is a board-reporting trigger and not a deductible. Last, the warranted check: total premium sits well below expected annual loss, or transfer is the wrong lever and the money belongs in controls.



> **Figure 6. What you're looking at:** the five rules as a decision flow, top to bottom, each with its test and its Feynman reading. The output at the bottom is the point: a limit and retention that trace to a percentile on the curve, defensible in front of a board or an underwriter, instead of a multiple of revenue somebody remembered from a webinar.

The reframing that matters most sits underneath the rules: anchor the transfer conversation on the floor, not the headline. The program will compress the addressable band over time, but the $1.4M floor and its tail persist by definition, and the tail is precisely the thing insurance exists to absorb. You self-carry the small floods; the hundred-year one is exactly what an underwriter is for, and nothing else on this list is.

The far tail of any Monte Carlo model carries wide confidence intervals, a fact some actuary will eventually email me about: the 99th-percentile figures are order-of-magnitude estimates, not quotes. The rules above survive that fuzziness because they lean on the curve's shape, not its last decimal.

10. Decision three: revenue at risk, on its own curve

The third use surprised me by becoming the most persuasive one in the boardroom, and the mistake to avoid is bolting it onto the security model. Revenue risk gets its own register and its own curve. Same engine, separate instrument.

The setup: Feynman sells to buyers whose procurement teams gate every deal on demonstrable security: questionnaires, audit reports, a named security leader who can sit across from their CISO. Two contracts on the table sat gated exactly this way, both with enterprise buyers in heavily regulated sectors, against roughly $20M of annual revenue.

The model treats each gated deal as a hazard in its own taxonomy: the value at stake is the deal's annual contract value, the frequency is the probability the deal stalls or dies on the security gate, and the program's effect is the reduction in that probability once the evidence exists. Calibrate those probabilities exactly like everything else (two models, two experts, reconciled, 90 percent CI, rationale) and the same Monte Carlo machinery produces a revenue-at-risk exceedance curve and an expected enabled revenue figure. Illustrative, with disguised numbers in the spirit of the rest of the piece:

Input (illustrative)	Deal A	Deal B
Annual contract value	$4M	$2.5M
P(stalls or dies on the security gate), no evidence	~40%	~35%
P(same), with program evidence in hand	~10%	~10%
Expected enabled revenue	~$1.2M/yr	~$625K/yr

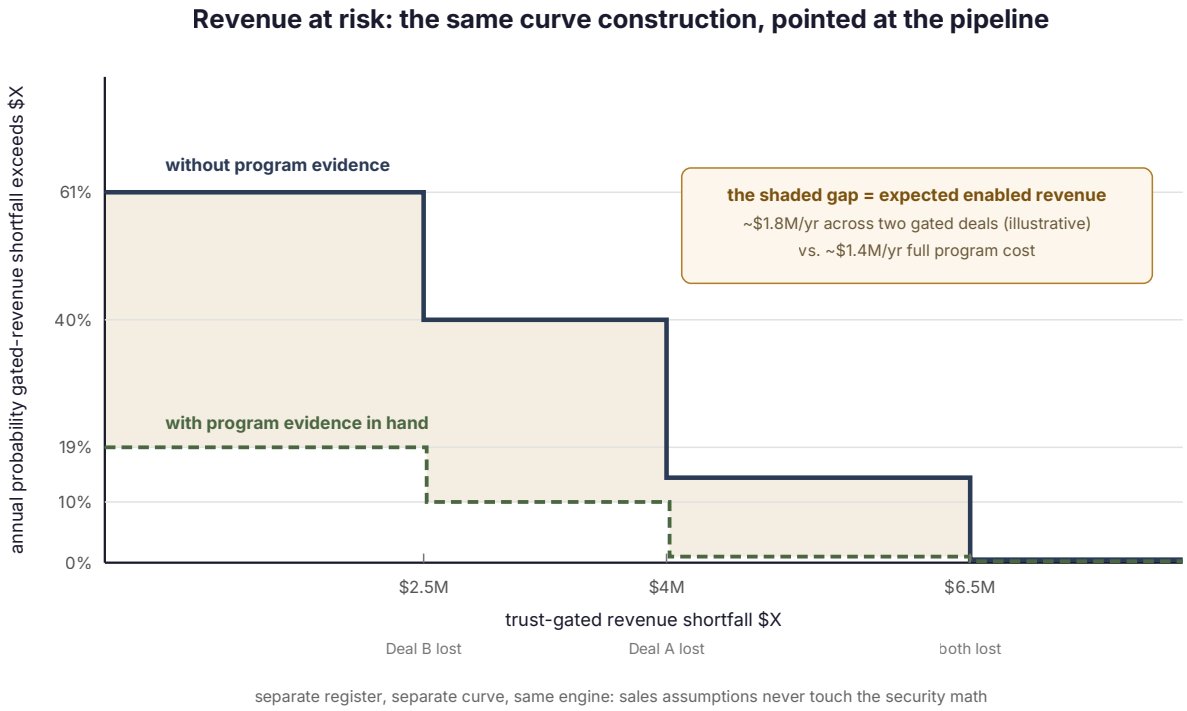


Figure 7. What you're looking at: the revenue-at-risk exceedance curve, built exactly like the loss curve but pointed at the pipeline: for every X, the annual probability that trust-gated revenue shortfall exceeds X. The upper solid line is Feynman without evidence in hand; the lower dashed line is with the program's evidence. The steps mark the individual deals. The shaded gap between the curves is the expected enabled revenue, about $1.8M a year here, and it's the number the program budget gets sized against from the growth direction. Illustrative figures, disguised like everything else.

Two deals, roughly $1.8M a year of expected enabled revenue, against a ~$1.4M annual program cost. The loss-avoidance case for the program (the $1.2M addressable band plus tail protection) is solid and roughly break-even; it wins a polite nod. The revenue curve is what changes the meeting, because it prices the program as growth infrastructure rather than loss prevention, and every future regulated buyer rides behind the first two.

Keep the two curves separate, permanently. Security hazards measure what you might lose; the revenue register measures what you might fail to win. Mixing them lets an optimistic sales assumption quietly subsidize a control decision, and the whole system's credibility rests on nobody being able to do that. Kept apart, the pair right-sizes the program from both directions: the loss curve bounds what protection is worth, the revenue curve bounds what trust is worth, and the budget defends itself against either.

I still label the revenue side carefully: deal values belong to the sales team, gate probabilities are calibrated judgment, and the causal share of security in any won deal is arguable. The direction

survives every quibble. When one gated contract covers the program's budget, the program stops being a cost line.

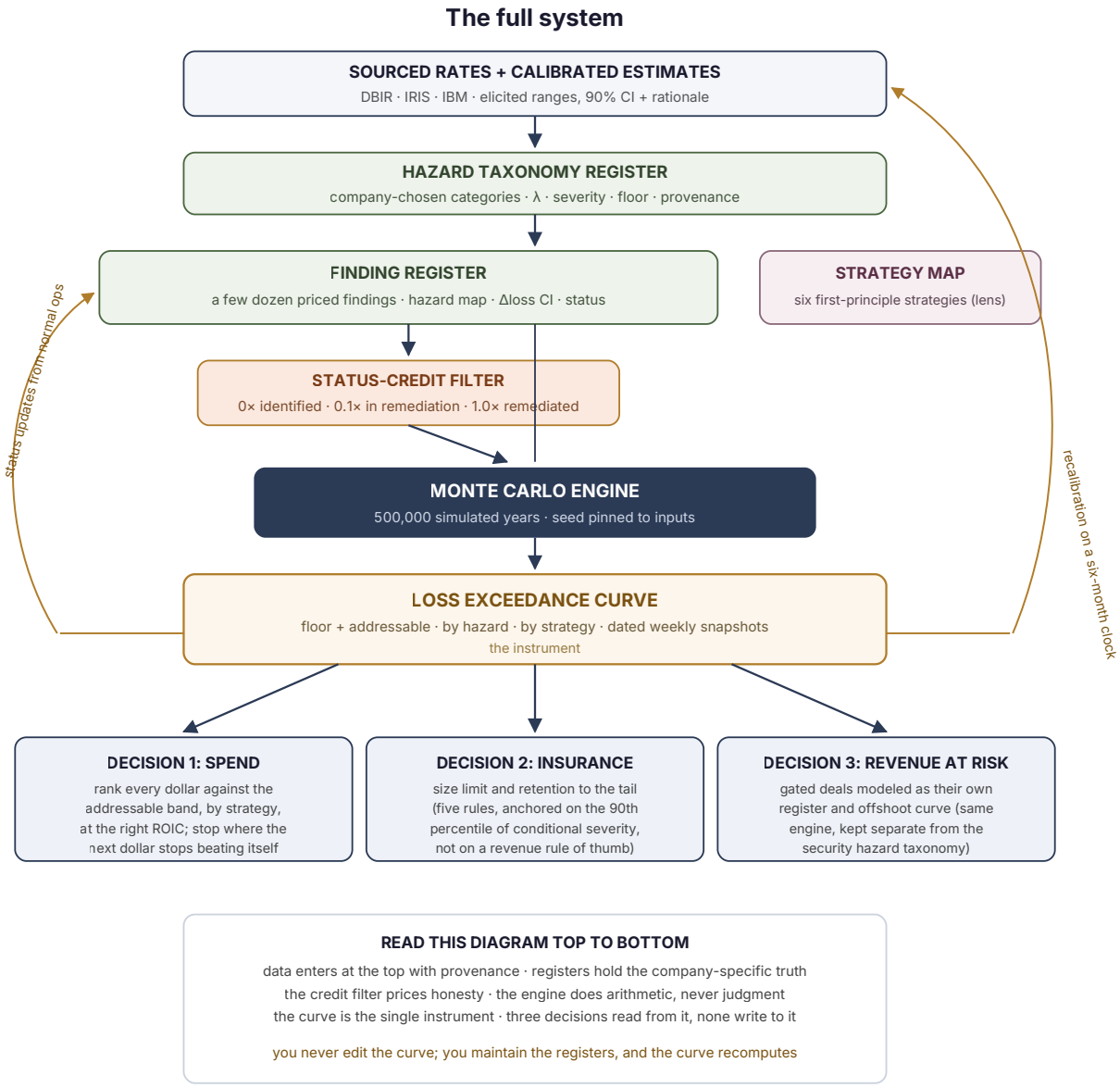


Figure 8. The finished diagram: everything the piece has assembled, top to bottom. Provenance-tagged data enters at the top; the hazard taxonomy and finding registers hold the company-specific truth; the status-credit filter prices honesty before the engine sees anything; the engine does arithmetic, never judgment; and the curve at the center feeds three decisions that read from it and never write to it. The two amber loops (status updates, the six-month recalibration clock) are what keep the instrument alive between builds.

11. What it takes, the case against it, and whose shoulders this stands on

The bill of materials, for anyone pricing the exercise. A five-variable baseline: one afternoon, one spreadsheet, sourced base rates, and a CFO willing to name the material threshold. The multi-hazard production model: a few working days spread across two weeks, most of it in calibration and provenance writing, none of it requiring a statistician. Operations: minutes a week, plus a semiannual recalibration measured in hours. The scarce ingredient at every step is the willingness to write down honest ranges and defend them.

The strongest objection to all of this arrives in five words: you made these numbers up. A 90 percent confidence interval elicited from experts remains a guess wearing a lab coat; the Monte Carlo machinery launders subjective inputs into objective-looking outputs; and a board that sees "$2.6M" will treat it as an actuarial fact when it sits closer to a structured opinion. False precision, the argument runs, beats admitted vagueness at exactly one thing: inviting confident wrong decisions.

Every clause of that is half right, and the half-rightness matters. The inputs are estimates: tagged, sourced where possible, interval-bounded where not, with the author and rationale written on each one. What the critique skips is the comparison class. No purer rigor waits offstage. The working alternative, at every company I have watched face this choice, is the heat map: the same subjective judgment, compressed to a color, stripped of its assumptions, immune to audit, and unable to be wrong because it asserts nothing. The curve can be wrong in public, one finding at a time, and that property (checkability) is the entire upgrade. Decades of calibration research show structured estimation with feedback beats unstructured expert intuition. The flat curve we had to hand the board mid-year was the system working.

The model's real limitations, so the reader holds them rather than discovering them: the hazards simulate independently, and real attacks correlate (the phish that becomes the credential abuse that becomes the exfiltration); the lognormal severity shape is an informed convention, and other heavy-tailed shapes fit breach data too; the far tail is order-of-magnitude; and expected annual loss is a mean, which is exactly why the spend and insurance decisions above read the whole curve instead. None of these breaks the instrument. All of them belong in the memo footnotes, permanently.

Credit where it's due, because none of the foundations here are mine. The measurement stance and the calibration protocol descend from Doug Hubbard and Richard Seiersen's How to Measure Anything in Cybersecurity Risk, the book that demolished "cyber can't be quantified" a decade ago. The decomposition tradition runs through Jack Jones's FAIR. The sector data that makes outside-in rates possible comes from the Verizon DBIR and the Cyentia Institute's IRIS research. And the strategic frame, risk forecasting as a first-principle obligation of the security program rather than a consultant's add-on, comes from Rick Howard's Cybersecurity First Principles (Wiley, 2023), which I edited and have been field-testing at Feynman ever since. What these notes add is the operating

pattern: the registers, the credit rules, and the provenance discipline that keep the curve honest between calibrations.

A year and a half in, here is what actually changed at one company. The budget fight runs on a ranked list instead of adjectives. The insurance limit traces to a percentile instead of a revenue multiple. The sales team hands the curve to customer security reviewers instead of scheduling a call about the questionnaire. The board hears "the flood line sits here, the sandbags cost this, and the river keeps some of it no matter what we do," and allocates accordingly. The heat map still exists somewhere in a slide archive, red corner and all.

Nobody has asked for it back.

Practitioner appendix

The main body earns the argument; this appendix carries the spec. Reference register from here down.

Appendix A: The data model

The production model is a frozen, versioned JSON input set plus registers that change daily. Separating the two is the design's load-bearing decision: slow-moving calibrated structure in one file, fast-moving finding status in another.

Hazard schema. Each modeled hazard category carries:

Field	Meaning
`id`	Category slug (e.g., `credential-abuse`)
`lambda_sector_base`	Annual event frequency, allocated from the sourced sector incidence by prevalence share
`lambda_floor`	Irreducible residual frequency; the model-wide floor equals the sum of per-category floor runs
`severity.median_usd / p95_usd / ceiling_usd`	Lognormal severity anchors per category
`severity.tail_shape`	Qualitative tail class (e.g., `heavy`) recorded with the anchors
`strategy_tags`	First-principle strategy assignment, one or more (reporting lens)
`provenance`	`kind` (SOURCED or CALIBRATED), owner, date, written rationale, and a `ci90` interval

Frequency sourcing. The sector base incidence comes from published data (Cyentia IRIS sector incidence tables). Per-category allocation shares come from prevalence data (e.g., DBIR access-vector rankings), then each category's floor is set as a calibrated fraction of its base.

Severity calibration. Category severity anchors to published loss bands for the company's revenue band (Cyentia IRIS loss tables), then gets shaped per category with a written rationale and a 90% CI.

Uplift factors. Multipliers for company-specific exposure (e.g., IP attractive to well-funded collection) applied with a minimal-overlap rule: at most one factor per hazard category, so multipliers never stack silently.

Nation-state modifier. A per-hazard probability that a given event is state-grade (highest on the categories touching the company's core IP, low on commodity hazards), carrying a severity multiplier (~3x) and a success uplift. This prices the difference between a criminal operator and a funded one without doubling the taxonomy.

The finding register. Each finding carries a calibrated loss delta (90% CI on frequency and loss), a status, strategy tags, and a hazard mapping with weights. Status multiplies credit: `identified` 0x, `in_remediation` 0.1x, `remediated` 1.0x. A finding with a null delta is excluded and rendered as excluded.

Strategy lens. Each finding and hazard tags to one or more of the six first-principle strategies (zero trust, kill chain prevention, resilience, risk forecasting, automation, workforce development). The lens groups reporting and drives the spend decomposition; it never changes the curve. Keeping the lens out of the math means reorganizing the program never silently rewrites the risk.

Revenue-at-risk register (separate instrument). Gated deals modeled as their own hazard set: value at stake, gate-failure probability with and without program evidence, calibrated CI. Same engine, separate curve, never merged with the security taxonomy.

Engine invariants. 500K trials; simulation seed pinned to the input hash (renders are reproducible; the headline moves only when inputs move); every consumer asserts the schema version on read; the frozen baseline file is never recomputed, only superseded by a new frozen version.

Appendix B: Calibration and elicitation protocol

1. **Classify every number.** SOURCED (published, cited by document, figure, and year) or CALIBRATED (estimated). No third class. A number that cannot name its source is CALIBRATED and gets the full treatment below.
2. **Assemble the evidence pack** for the quantity: sector reports, incident precedents, the company's own architecture and exposure facts.
3. **Elicit independently, at least twice.** The classic instrument is a room of calibration-trained human experts (Hubbard & Seiersen's protocol). The compressed instrument: two frontier AI models from different vendors, given the same evidence pack, each returning low/median/high on frequency and loss with reasoning. Add the analyst's own expert assessment and a third party's expert assessment as further independent estimates where available.
4. **Reconcile by hand.** Where the independent estimates converge, set the 90% CI at roughly +/-20% around the consensus. Where they diverge, the CI spans the disagreement. The human reconciler owns the final number and signs the rationale.

5. **Write the provenance record** on the value itself: kind, owner, date, rationale, CI. A future reader must be able to re-derive or challenge any number without archaeology.
6. **Freeze and version.** Calibrated structure freezes as a dated, versioned input file. Changes produce a new version; nothing edits history.
7. **Guardrails.** Sum-of-deltas stays under the baseline headroom (no register can claim more reduction than the curve contains); elicitation past the recalibration due date triggers a warning; a failed cross-check against the prior model version (combined probability and EAL tolerance bands) halts promotion.
8. **Recalibrate on a six-month clock,** or immediately when the reference class shifts (funding stage, headcount band, new regulated market).

Appendix C: Building your own hazard taxonomy

The goal is a set of categories where any incident, claim, or finding lands in exactly one bucket, and every bucket can carry its own frequency and severity. The candidate menu lives in section 3; this is the selection procedure.

1. **Start from DBIR/VERIS action categories.** They are incident-derived, maintained annually, and map to the loss data you will source rates from.
2. **Cross-walk** against MITRE ATT&CK tactics, ENISA's threat landscape, FAIR's threat communities, and NIST SP 800-30's threat events. Where a category exists in fewer than two frameworks, question whether it earns a bucket.
3. **Collapse to claim level.** The test for granularity: could an insurance adjuster or incident commander assign a real event to exactly one category without a meeting? Eight to fourteen buckets passes for most companies; twenty-five will not.
4. **One severity family per bucket.** If a proposed category contains events with wildly different loss shapes (lost laptop and cloud-wide ransomware), split it; if two categories share a shape and a defense, consider merging.
5. **Document deliberate omissions.** An omission with a rationale is a scoping decision; a silent omission is a hole.
6. **Give every category a floor.** The irreducible share is per-hazard, and the model-wide floor should emerge from the sum, never get asserted top-down.
7. **Tag every category to strategies.** The strategy lens (section 5) only works if the taxonomy carries the tags from day one.

Appendix D: Glossary and sources

Loss exceedance curve (LEC). For every loss amount X, the annual probability that total losses exceed X. The output artifact of the whole method.

Expected annual loss (EAL). The probability-weighted mean annual loss; the area under the exceedance curve. Useful for budgets; misleading alone for tail decisions.

Irreducible floor. The expected-loss residual that persists at full control maturity: novel attacks, state actors, insiders. The anchor for the insurance conversation.

Addressable loss. EAL minus floor; the only pool control spend should be ranked against.

Material threshold. The loss amount that triggers board or regulatory attention. A finance number, owned by the CFO.

Conditional severity percentile. Given that an event occurs, the loss at the Nth percentile. The 90th conditional percentile anchors the insurance limit floor.

Status credit. The multiplier a finding's remediation status earns: identified 0x, in remediation 0.1x, remediated 1.0x.

First-principle strategies. The strategy set the program designs against: zero trust, intrusion kill chain prevention, resilience, risk forecasting, automation (Howard, 2023), plus workforce development in our implementation.

Expected enabled revenue. From the revenue-at-risk register: deal value times the reduction in gate-failure probability the program's evidence buys.

Lognormal severity. The heavy-right-tail distribution fitted between the loss floor and ceiling; most events small, few catastrophic.

Primary sources. Rick Howard, Cybersecurity First Principles: A Reboot of Strategy and Tactics (Wiley, 2023); Douglas Hubbard & Richard Seiersen, How to Measure Anything in Cybersecurity Risk (Wiley, 2nd ed. 2023); Jack Jones & Jack Freund, Measuring and Managing Information Risk: A FAIR Approach; Verizon Data Breach Investigations Report (annual); Cyentia Institute, IRIS (sector incidence and loss bands); IBM Cost of a Data Breach (annual); Sophos State of Ransomware (recovery durations); NIST SP 800-30; MITRE ATT&CK; ENISA Threat Landscape.

Method and structure: Bombadil LLC. Feynman Technology Solutions is an invented name; the client's data is disguised and shared with permission. The model described operates in production; the author will happily compare notes with anyone running a similar instrument.