AUSPICA · AI GOVERNANCE ADVISORY

The Modality Method

Identifying and prioritizing AI use cases across machine learning, generative AI, and agentic AI. Choose the work by value. Choose the modality by fit. Choose the order by governance load.

Third in a set of three. The Declared Behavior Framework establishes whether governance is real. Continuously Enforceable AI decides how much autonomy a use case has earned. This method decides which use cases to pursue, in which modality, and in what order.

Version 1.0 | August 2026

PART 0

The premise

Most AI portfolios are scored on a single sheet: value on one axis, feasibility on the other, everything ranked together. The sheet is not wrong so much as insufficient, because it treats three genuinely different technologies as one, and the differences between them determine almost everything that happens after the scoring.

Machine learning predicts. Generative AI produces. An agent *acts*. Those verbs carry different failure modes, different evidence requirements, and governance costs that differ by roughly an order of magnitude at each step. A ranked list that mixes them will consistently over-rank agentic work, because agentic work has the most visible upside and its costs arrive latest.

> THE THREE DECISIONS
>
> **Which work is worth doing.** Answered by value, and modality-independent. **Which modality fits.** A decision, not an assumption. Most organizations decide the modality first and then hunt for problems, which is backwards and expensive. **In what order.** Answered by governance load, not by value. The highest-value item is often the one an organization is least equipped to run.

Why modality is a decision

The same business problem can usually be addressed in more than one modality, and the choice is consequential. Consider invoice exception handling. As machine learning it is a classifier that routes exceptions by predicted type. As generative AI it is an assistant that reads the invoice and drafts an explanation for a human. As an agent it investigates, corrects the record, and notifies the vendor.

All three are legitimate. They differ in value, in build cost, in time to first result, and by a wide margin in what has to be true before they can safely run. Choosing among them deliberately is the single highest-leverage decision in the whole exercise, and it is the one most often made by default.

The governance-load asymmetry

This is the fact the method is built around, and it is why ordering cannot follow value.

MODALITY	WHAT IT DOES	GOVERNANCE REQUIRED
Machine learning	Predicts, scores, classifies over structured data.	Model risk management: validation, bias testing, drift monitoring. A mature discipline with known cost and existing owners.
Generative AI	Produces content, language, synthesis for a human to consume.	Output review, data leakage control, provenance and intellectual property, hallucination exposure. Newer, but the human in the path is the control.
Agentic AI	Plans and takes action in systems, often without a human in the path.	Everything above, plus a declaration of permitted behavior, conformance measurement, containment of side effects, ownership of downstream consequence, and evidence a third party can verify.

The practical consequence is blunt. **A use case that is marginal on value should not be done agentically,** because the governance cost is the same whether the payoff is large or small. Agentic autonomy has to be paid for, and only high-value work justifies the payment.

PART I

Discovery: finding the candidates

Discovery fails in a predictable way: a workshop is convened, people suggest what they have read about, and the list reflects the market's imagination rather than the organization's actual costs. Three lenses produce a better list, and each surfaces a different modality naturally.

Lens 1 · Friction: where attention is being spent

Work where a capable person spends attention on something that does not require their judgment. Measure manual touch time, handoff count, wait states, rework rate and exception volume. This lens surfaces generative and agentic candidates, because friction is usually reading, assembling, drafting and chasing.

Lens 2 · Decision latency: where judgment arrives too late

Places where a decision is made well but slowly, or made late because the analysis is expensive. Measure elapsed time from trigger to decision, and the cost of the delay itself. This lens surfaces machine learning candidates, because the answer is usually a prediction that could have been available earlier.

Lens 3 · Knowledge scarcity: where expertise is the bottleneck

Work that only three people can do, that queues behind them, and that degrades when they are unavailable. This lens surfaces generative candidates, and occasionally reveals that the real problem is documentation rather than AI.

Where to look, in order of signal quality

- **Shadow AI telemetry.** The strongest source available and the most consistently ignored. Employees already using unsanctioned tools are telling you, with revealed preference rather than stated preference, exactly which work is painful enough to risk a policy violation over. Data loss prevention logs, network egress to model providers, and endpoint discovery are a ranked candidate list that somebody else compiled for free.
- **Process mining and workflow telemetry.** Objective measurement of touch time, wait states and rework. Expensive to stand up, unarguable once you have it.
- **Service desk and exception queues.** Every recurring ticket category is a candidate, and the volume is already counted.
- **Frontline interviews.** Ask what people do that they consider beneath them, and what they postpone. Do not ask where AI could help; that question returns the market's answers, not theirs.
- **Executive strategy.** Necessary for legitimacy and funding, and the weakest signal for actual friction. Use it to weight, not to generate.

A DISCOVERY RULE

Record the problem, never the solution. A candidate entered as “chatbot for HR policy questions” has already had its modality chosen, its value assumed and its alternatives foreclosed. The same candidate entered as “employees wait two days for policy answers that exist in a document nobody can search” stays open to a search index, a generative assistant, or a rewrite of the document, and one of those is much cheaper than the others.

PART II

Modality selection

Four questions, asked in order, applied to each candidate. Order matters: the questions run from the most consequential commitment to the least, so a candidate settles at the lightest modality that will actually do the job.

Q1 · Must something happen in a system without a human initiating each instance?

If no, the candidate is not agentic, whatever anyone hopes. Proceed to Q2. If yes, it is an agentic *candidate*, and must clear three conditions before it is agentic in practice:

- **Declarable.** Can you write down in advance, specifically enough to check, what the agent is permitted to do?
- **Observable.** Can you tell afterward what it actually did?
- **Reversible.** Can an incorrect action be undone or compensated?

Any of the three failing means the candidate is not agentic yet. It falls back to a generative assistant inside a human workflow, which usually captures most of the value at a fraction of the governance cost. This is a demotion, not a rejection, and it is the most common correct outcome of the whole method.

Q2 · Is the output content, language or synthesis that a human consumes?

If yes, it is a generative candidate, subject to one condition: **can a human evaluate the output reliably, at the rate the work arrives?** If quality cannot be judged, or can only be judged more slowly than the output is produced, the review becomes theater and the use case needs an evaluation harness before it needs a model.

Q3 · Is the output a prediction, score or classification over structured data with historical ground truth?

If yes, it is a machine learning candidate. Ground truth is the binding constraint: without labeled history and a feedback loop that keeps producing labels, the model cannot be validated at build time or monitored afterward.

Q4 · None of the above

Then it is probably not an AI problem. It is deterministic automation, a systems integration, a data quality problem, or a process that should be redesigned or stopped. **Recording this outcome honestly is what makes the rest of the portfolio credible,** and in most inventories it accounts for a meaningful share of the candidates. An organization that never returns this answer is not applying the method.

THE MOST EXPENSIVE ERROR IN THE FIELD

Choosing agentic for a problem a workflow would solve. It arrives with the best demonstration, the highest enthusiasm and the largest governance bill, and it is usually chosen because the modality is fashionable rather than because the problem requires action without a human. The test is Q1, asked strictly: does something have to *happen*, without a person initiating it, every time?

PART III

Qualification gates

Binary conditions, checked before anything is scored. A candidate failing a gate is not low-priority; it is not yet a candidate, and the gate itself is the first piece of work.

GATE	APPLIES TO	FAILS WHEN
Named owner	All	No person will accept accountability for the outcome. A committee is not an owner.
Observable outcome	All	Nobody can tell afterward whether the result was right.
Ground truth	ML	No labeled history, or no feedback loop that will keep producing labels.
Signal stability	ML	The underlying relationship changes faster than the model can be retrained.
Evaluability	GenAI	Output quality cannot be judged reliably, or not at the rate output arrives.
Source quality	GenAI	The knowledge the output depends on is absent, stale or contradictory.
Declarability	Agentic	Permitted behavior cannot be written down specifically enough to check.
Reversibility	Agentic	An incorrect action cannot be undone or compensated. Permanent recommendation rung at best.
Tool surface	Agentic	The systems the agent must act in have no usable interface.

The declarability gate deserves particular attention, because failing it is usually a symptom rather than a technical limitation. If nobody can state what an agent should be permitted to do, that is rarely because the task is ineffable. It is because the task has never been specified, and it was previously being held together by a human applying judgment nobody wrote down. Discovering that is valuable in itself and often worth more than the automation.

PART IV

Scoring value and feasibility

Two scores per candidate, each out of twenty. Value dimensions are common across modalities; feasibility dimensions are modality-specific, because what makes a machine learning use case tractable has almost nothing in common with what makes an agentic one tractable.

Value, scored 1 to 5 on each dimension

DIMENSION	SCORE 5 WHEN	SCORE 1 WHEN
Friction removed	Hours of skilled attention per instance, with handoffs and rework.	Minutes of unskilled attention, single touch.
Decision quality	Better or faster decisions change a material outcome.	The decision was already good and timely.
Volume	Thousands of instances per period, growing.	Occasional, and unlikely to grow.
Leverage	Creates reusable capability: data, declarations, integrations others will use.	Solves one problem once, reusable by nobody.

Feasibility, scored 1 to 5, by modality

MACHINE LEARNING	GENERATIVE AI	AGENTIC AI
Labeled data depth	Knowledge source quality	Tool and API surface
Signal stability over time	Output evaluability	Observability of actions
Ground-truth feedback loop	Human review capacity	Reversibility of actions
Integration point exists	Context stability	Declarability of behavior

Rank on Value × Feasibility, which ranges from 1 to 400 and separates candidates far more usefully than a sum. A candidate scoring 18 on value and 4 on feasibility is not comparable to one scoring 11 and 11, and addition would rank them almost identically.

Score with the people who do the work, not about them, and score in one session per domain rather than asynchronously. The conversation that produces the numbers is worth more than the numbers, and disagreement about a score is nearly always disagreement about scope that would otherwise have surfaced during delivery.

PART V

Governance load and wave assignment

This is where the method departs from a conventional priority matrix. Governance load does not reduce a candidate's score. **It determines which wave the candidate can be delivered in,** and a wave cannot start until the governance capability it requires exists.

Ranking by value alone produces a portfolio that stalls: the top items turn out to need capabilities nobody has built, the program discovers this during delivery, and the schedule absorbs it as delay rather than as sequencing.

Computing the load

COMPONENT	SCORE	BASIS
Modality floor	1 / 2 / 4	Machine learning 1, generative 2, agentic 4. Not arbitrary: it reflects how many of the five governance layers must be operating before the use case can run safely.
Blast radius	0 to 2	How far a wrong result propagates. 0 if contained to one record and one reviewer; 2 if it reaches customers, money or systems of record.
Data sensitivity	0 to 2	0 for public or internal-general; 2 for regulated, personal or privileged data.
Irreversibility	0 to 2	0 if trivially undone; 2 if the effect is external, permanent or reputational.

Waves

WAVE	LOAD	TYPICAL CONTENT	PREREQUISITE
Wave 1	1 to 3	Machine learning on existing data. Generative assistance with a human reviewing every output.	An inventory, a named owner per use case, and a basic acceptable-use policy.
Wave 2	4 to 6	Generative AI at scale with sampled review. Agents at the recommendation rung, proposing but never acting.	Declarations in force, observation running, evidence retained. Wave 1 delivered.
Wave 3	7 to 10	Agents acting within a declared envelope, gated or reviewed by exception.	Conformance measured continuously, containment tested, reversal rehearsed, evidence verifiable. Wave 2 delivered.

THE RULE THAT MAKES WAVES WORK

A high-value candidate in wave 3 does not get promoted into wave 1 because it is high-value. It gets scheduled for wave 3, and the governance capability it needs becomes an explicit, funded workstream with its own dates. Pulling it forward is precisely how organizations end up with an ungoverned agent in production and a two-year freeze afterward.

PART VI

The portfolio view

The output of this method is not a ranked list. It is a portfolio with a deliberate shape, and shape matters more than rank because a portfolio concentrated in one modality fails in one way, all at once.

Shape for a first year

- **Roughly half in wave 1.** These fund the program's credibility. They deliver inside a quarter, they are defensible, and they buy the patience that later waves require.
- **Roughly a third in wave 2.** These build the governance capability that wave 3 depends on. Their real deliverable is not the use case; it is declarations, observation and evidence in production.
- **At most a fifth in wave 3, and none of it early.** One or two agentic use cases, deliberately chosen for high value and tested reversibility, entering at the recommendation rung.

A portfolio that is entirely wave 1 after a year is not cautious; it is a program that never built anything. A portfolio that is mostly wave 3 in the first six months is not ambitious; it is a program that has not yet discovered what it cannot evidence.

Two cross-cutting checks

Capability dependencies

Each modality rests on foundations that are themselves projects. Machine learning needs a feature store and a retraining loop. Generative AI needs retrieval infrastructure and an evaluation harness. Agentic AI needs a declaration practice, observation, and containment. **If three candidates all depend on the same absent foundation, that foundation is the higher-priority item** and should appear in the portfolio as work in its own right rather than as an assumption inside three business cases.

Concentration risk

Count how many candidates depend on a single model provider, a single data source, or a single team. Concentration is invisible in a ranked list and obvious in a portfolio view, and it is the failure that turns one vendor change into a program-wide event.

PART VII

Operating the pipeline

Prioritization is not an exercise that concludes. Candidates arrive continuously, scores go stale, and the modality decision for a candidate rejected in March may be correct in September because a gate has since been cleared.

CADENCE	ACTIVITY	OUTPUT
Continuous	Intake. New candidates recorded as problems, not solutions. Gates checked at entry.	A candidate record, or a documented rejection with the failing gate named.
Monthly	Scoring session by domain. Value and feasibility scored with the people who do the work.	Updated Value × Feasibility, wave assignment, capability dependencies flagged.
Quarterly	Portfolio review. Shape, concentration, foundations, and re-scoring of anything whose gates have changed.	A revised portfolio and an explicit list of candidates that moved wave, in either direction.
Annually	Method review. Are the scores predicting outcomes? Which delivered use cases were mis-scored, and in which direction?	Recalibrated dimensions and, where warranted, revised modality-floor weights.

The annual review is the one most often skipped and the one that makes the method improve. A scoring model that is never checked against delivered outcomes is a ritual. Track, at minimum, whether high-scoring candidates delivered and whether any wave-3 item was pulled forward and what happened.

PART VIII

Anti-patterns

Seven failures. Each looks like progress from inside the program.

- **Modality fashion.** The modality is chosen before the problem is understood, because it is the one being funded this year. Every candidate becomes agentic, and the governance bill arrives all at once.
- **Solution-shaped intake.** Candidates recorded as “chatbot for X” rather than as the underlying cost. Alternatives are foreclosed before scoring begins.
- **One score for three modalities.** Feasibility dimensions that make sense for a classifier applied to an agent. The scores are numerically comparable and mean nothing.
- **Governance discovered during delivery.** Load was never scored, so the capability requirement surfaces as schedule slippage in month four and is recorded as a delay rather than as a sequencing error.
- **The pilot that cannot graduate.** A wave-3 use case delivered as a wave-1 pilot by simply not governing it. It works, it cannot go to production, and it consumes the program's credibility on the way to being shelved.
- **Ignoring shadow AI.** The best candidate list in the organization sits in the data loss prevention logs, and the program runs workshops instead.
- **No rejections.** If the method never returns “this is not an AI problem” or “this fails the declarability gate”, it is not being applied. A method that approves everything ranks rather than decides.

> THE COMMON ROOT
>
> Six of the seven come from the same substitution: **deciding the modality before understanding the problem, and discovering the governance cost after committing to the modality.** The method exists to reverse that order, which is why the modality question sits in Part II and the load calculation in Part V, rather than both at the end.

APPENDIX

Candidate scoring worksheet

One per candidate. Complete in order; stop at the first failed gate.

FIELD	ENTRY
Candidate	Stated as a problem and its cost, not as a solution.
Source	Shadow AI telemetry / process mining / service desk / interview / strategy.
Named owner	A person.
Modality (Q1 to Q4)	ML / GenAI / Agentic / None. Record the question that decided it.
Gates	Each applicable gate: pass or fail. Any fail stops scoring.
Value	Friction __ + Decision quality __ + Volume __ + Leverage __ = __ / 20
Feasibility	Four modality-specific dimensions, 1 to 5 each = __ / 20
Rank score	Value × Feasibility = __ / 400
Governance load	Modality floor __ + Blast radius __ + Sensitivity __ + Irreversibility __ = __ / 10
Wave	1 (load 1 to 3) / 2 (4 to 6) / 3 (7 to 10)
Blocking foundation	Any capability this candidate needs that does not yet exist.
Re-score trigger	The specific condition that would change the modality, the wave or a failed gate.