CUTTING THROUGH THE AI NOISE

THE VFR FRAMEWORK

Framework to evaluate, question, and act on any AI initiative. No AI background required.

Why this framework exists

95% of enterprise AI pilots fail to deliver measurable returns.

95%

of pilots fail to generate ROI

3×

more expensive to fix late vs. early

$M+

wasted per failed enterprise AI project

The core issue is not technical. It's an evaluation gap and you're in the best position to close it.

Identify a real initiative

Identify one AI tool or initiative in your organization right now.
Something being considered, piloted, or already live.

Three patterns behind almost every AI failure

01

The Vague Win

Approved because it "improves productivity." No baseline. No metric. No way to know if it worked — or failed.

02

The Demo Didn't Match Reality

Clean demo. Messy workflow. Integration took 3× longer. The team that would use it was never consulted.

03

Risk Was Outsourced

"Legal will handle it." No named owner. When something went wrong, no escalation path and no rollback plan.

None of these are technical failures. They are evaluation failures and are preventable if someone knew what to ask.

The Framework

VALUE

"Is this worth doing, and how would we know?"

Prevents: The Vague Win

FIT

"Will it actually work here, in our reality?"

Prevents: Demo Didn't Match Reality

RISK

"What can go wrong, and who owns it?"

Prevents: Risk Was Outsourced

Value

"Is this worth doing, and how would we know?"

WHAT GOOD LOOKS LIKE

- Baseline metric exists today
- Target is defined and specific
- ROI logic is explicit
- Named beneficiary (a role, not "everyone")
- Tradeoffs are acknowledged

SCORING GUIDE

5	Baseline + target + ROI + named beneficiary + tradeoffs
3	Clear problem. Target vague or baseline not yet measured.
1	"Innovation." No baseline. No owner. "Everyone will benefit."

RED FLAGS

- No baseline
- "Everyone benefits"
- Value = innovation
- ROI unstated
- No named beneficiary

Fit

"Will it actually work here, in our reality?"

WHAT GOOD LOOKS LIKE

- Specific workflow step is named
- Integration path is confirmed (SSO, permissions, logs)
- Data is available, clean, and permitted
- Adoption plan exists with a named owner
- Human checkpoint is defined

SCORING GUIDE

5	Workflow named. Integration confirmed. Data verified. Adoption owned.
3	Workflow plausible but integration is TBD. Adoption assumed, not planned.
1	"We'll figure out integration later." Users never consulted.

RED FLAGS

- Integration deferred
- Data not verified
- Users not consulted
- No adoption owner
- No human review

Risk

"What can go wrong, and who owns it?"

WHAT GOOD LOOKS LIKE

- Top 3 failure modes are documented
- A single named person owns risk acceptance
- Privacy/security review path is defined
- Monitoring plan exists (drift, incidents, quality)
- Rollback plan is written and tested

SCORING GUIDE

5	Failure modes documented. Named owner. Monitoring live. Rollback written.
3	Risks acknowledged but undocumented. Team owns risk. Monitoring TBD.
1	"Legal will handle it." No monitoring. No rollback. No named owner.

RED FLAGS

- "Legal will handle it"
- Committee owns risk
- No rollback plan
- No monitoring
- "Just an assistant"

The Question Toolkit

12 questions. Each one prevents a specific failure.

The best questions aren't aggressive. They're specific.
A vague answer to a specific question is data.

4 Value Questions

Prevent the Vague Win

4 Fit Questions

Prevent Demo vs. Reality gap

4 Risk Questions

Prevent Outsourced Risk

Value Questions

What specific metric changes if this works? What is the baseline today?

Prevents the Vague Win. "Productivity" is not an answer.

Where in the workflow does the value appear? Which step becomes faster, safer, or higher quality?

Prevents abstract claims that don't map to anything a human actually does.

What number tells us to continue — and what number tells us to stop?

Prevents pilot drift. Pre-commits to success and failure thresholds.

What do we stop doing if we do this? What gets deprioritized?

Prevents initiative stacking. Every new thing displaces something.

Fit Questions

Exactly who uses this, how often, and at what moment in their day?

Prevents hypothetical-user solutions. "The team" is not a user.

What systems does it touch, and what is the integration path — SSO, permissions, logging, audit trails?

Prevents demo-to-reality gaps. "We integrate with everything" is a claim.

What data does it require, and is that data available, clean, and permitted for this use?

Prevents pilots that collapse on data access. Real data is messy and legally constrained.

What human check exists, and what happens when the output is wrong?

Prevents no-oversight deployments. Every AI output has a failure rate.

Risk & Accountability Questions

What are the top 3 ways this breaks? Walk me through each failure mode.

Prevents risk minimization. The quality of the answer tells you everything.

Who owns risk acceptance? Not "we" — an actual role or named person.

Prevents diffuse ownership. A committee is not an owner.

How will we monitor quality over time — drift, false positives, user feedback, incidents?

Prevents set-and-forget deployments. AI models degrade. Monitoring is not optional.

What is the rollback plan? If this harms outcomes, how do we reverse it — and how fast?

Prevents irreversible deployments. No rollback plan = betting on success.

The one question that works in any conversation:

"What evidence would change your mind about this either way?"

Use it on a vendor you're skeptical of.

Use it on an initiative you're championing.

It surfaces assumptions. It keeps the conversation honest in both directions.

Your 7-Day Action Map

Go back to the initiative you typed at the start. Pick one move in each category.

A

One Conversation

Ask for the baseline + success threshold. Or map the workflow. Or ask who owns rollback. Schedule it before you close your laptop today.

B

One Metric

Define the baseline and target that doesn't exist yet. Or push for adoption or quality metrics. Name where it's tracked and who owns it.

C

One Risk Flag

Surface the privacy/security review status. Ask about the monitoring plan. Ask for the rollback procedure. Name who signs off.

Three rules that keep it high-trust

01

Score claims, not people.

VFR is a tool for evidence, not judgment. The goal is clarity, not conflict. Apply it to yourself as readily as you apply it to others.

02

Demand evidence, not certainty.

Pause is not failure. Missing evidence is information. Knowing what's absent is as valuable as knowing what's present.

03

Name owners, not committees.

A decision owned by everyone is owned by no one. Putting a name next to a risk is one of the most powerful things you can do.

You leave with four tools.

- **VFR Scorecard**

  Evaluate any initiative. Output a defensible Proceed / Pause / Stop.

- **12 Questions**

  Use in demos, planning meetings, and leadership conversations.

- **Decision Memo**

  Half-page format. Turns your evaluation into something that travels.

- **7-Day Action Map**

  One conversation, one metric, one risk flag — with names and dates.