EXECUTIVE FIELD GUIDE

The AI Vendor BS Detector

How to Cut Through the Hype and Make Smarter AI Investments

JJ Shay | Global Gauntlet AI
bit.ly/jjshay

The $4.4 Trillion Problem

87%
of AI projects never make it to production

$4.4T
projected AI spend by 2030

Most executives can't distinguish between transformative AI and expensive vaporware. This guide changes that.

> *"We spent $2.3M on an AI solution that our team refuses to use. The demo was incredible. Production was a disaster."*
>
> — Fortune 500 CTO (anonymous)

Why You Need a BS Detector

The Demo Illusion

Vendors perfect demos on cherry-picked data. Production reality is messier.

Benchmark Gaming

Accuracy numbers are meaningless without context on test conditions.

Hidden Costs

The sticker price is 20-40% of total implementation cost.

Timeline Fantasy

Average AI project takes 3x longer than vendor estimates.

The 7 Phrases That Should Kill Any AI Pitch

■ What Vendor Says	→	What It Actually Means
"Our AI is 99% accurate"	→	On our hand-picked test set, not your data
"It works out of the box"	→	Basic version works; your use case needs months of tuning
"Trusted by Fortune 500"	→	Someone bought a pilot license we never renewed
"Our proprietary algorithm"	→	We wrapped open-source in a nice UI

■ What Vendor Says	→	What It Actually Means
"No technical expertise needed"	→	Until something breaks (it will)
"AI-powered" / "Uses ML"	→	Could be anything from basic rules to actual AI
"ROI in 6 months"	→	Based on theoretical best-case nobody achieves

The Demo vs. Production Gap

Why Demos Look Amazing

✗ Pre-selected "golden" data samples

✗ Controlled environment, no edge cases

✗ Sales engineer manually fixes issues live

✗ Hardcoded responses for common questions

✗ No integration complexity shown

Production Reality

✗ Messy, inconsistent real-world data

✗ Edge cases that break the model

✗ Integration nightmares with legacy systems

✗ Performance degrades over time (model drift)

✗ Users find creative ways to break it

The Benchmark Manipulation Playbook

How vendors make numbers say whatever they want:

Cherry-Picked Test Sets

Testing on data that looks like training data. Real world is messier.

Metric Shopping

Reporting accuracy when precision matters. Or vice versa.

Lab Conditions

"99% accuracy" achieved with unlimited compute and perfect preprocessing.

Hiding Failure Modes

Averages hide that the model fails catastrophically on 5% of cases.

Questions That Make Vendors Squirm

■ Can we test on OUR data before signing?

If they hesitate, the demo was a magic trick. Real confidence means real testing.

■ What's your churn rate and why do customers leave?

Evasive answers = red flag. Good vendors know their weaknesses.

■ Can I speak to a customer who FAILED implementation?

Success stories are curated. Failures teach you what breaks.

■ What percentage of customers achieve the ROI you quote?

Watch them pivot to "it depends." Because it does. A lot.

More Killer Questions

■ What happens when the model is wrong? How do users correct it?

No feedback loop = no improvement. The model stays dumb forever.

■ Show me the model's confidence scores on edge cases.

Good models know when they don't know. Bad ones guess confidently.

■ What's the total cost of ownership over 3 years?

License is the tip of the iceberg. Implementation, training, and maintenance sink ships.

■ Who owns the model trained on our data if we leave?

Your data makes their product better. Make sure you can take your work with you.

The Hidden Cost Iceberg

Cost Category	Amount
License	$150,000
Implementation Services	$180,000
Data Preparation	$75,000
Integration Development	$120,000
Training & Change Mgmt	$45,000
Ongoing Maintenance	$60,000/yr
YEAR 1 TOTAL	**$630,000**

The sticker price ($150K) is only 20-40% of the real cost!

Are You Ready for AI?

Most failed AI projects fail because of organizational readiness, not technology.

- ■ **Data Quality:** Is your data clean, labeled, and accessible?
- ■ **Clear Problem:** Can you define success metrics precisely?
- ■ **Executive Sponsor:** Who owns this when it gets hard?
- ■ **Change Management:** Will users actually adopt this?
- ■ **Technical Talent:** Who maintains this after go-live?
- ■ **Process Clarity:** Is the current workflow documented?

Score: Less than 4 checks? Fix foundations before buying AI.

Green Flags: Signs of a Legitimate Vendor

Offers Paid Pilot First

Confident vendors let you test before committing.

Admits Limitations Upfront

"Here's where we struggle" beats "We do everything perfectly."

Connects You to Recent Failures

Mature vendors learn from failed implementations.

Provides Explainability

Can they show WHY the model made a decision?

Clear Data Ownership Terms

You own your data. Period. No gray areas.

Documented Integration APIs

Public, well-documented APIs = they expect scrutiny.

SLAs with Teeth

Performance guarantees with financial penalties.

Customer Success Beyond Sales

Dedicated post-sale support team.

The POC Trap

Why Most POCs Succeed...

✓ Vendor's best engineers assigned

✓ Your best data selected for testing

✓ Scope narrowed to easy wins

✓ Timeline pressure creates shortcuts

✓ Success defined by vendor, not you

...And Implementations Fail

✗ Junior consultants take over

✗ Real data is 10x messier

✗ Scope expands to actual requirements

✗ Integration complexity explodes

✗ Users resist adoption

> *"A successful POC is evidence of vendor capability. It's not evidence of production viability."*

How to Run a Real POC

Define Success First

Write acceptance criteria BEFORE the POC starts.

Use Your Worst Data

Test edge cases, messy records, and exceptions.

Include Skeptics

Put your most critical users in the pilot.

Test Response Time

Demo speed ≠ production speed. Measure latency.

Break It Intentionally

Feed bad data. See how it fails gracefully (or doesn't).

Measure Time Savings

Track actual hours saved, not projected.

Contract Red Flags

"Results may vary"

No performance guarantees = no accountability.

Multi-Year Lock-In

3-year minimum with limited exit? They know you might want to leave.

Usage-Based Pricing Opacity

"Per API call" without clear definitions = surprise bills.

Data Rights Ambiguity

Any clause letting them train on your data.

Contract Protections to Demand

Protection	Why It Matters
Performance SLAs	95%+ uptime, max latency, accuracy floors with financial penalties
Data Portability	Export ALL your data in standard formats within 30 days
Model Ownership	Custom models trained on your data belong to YOU
Price Lock	Cap annual increases at CPI + 3%, not "market rate"
Exit for Convenience	90-day notice termination after year 1
Security Audit Rights	Annual SOC 2 reports + right to conduct pen tests

The Vendor Scorecard

Rate each vendor 1-5 on these dimensions before making a decision:

Dimension	What to Evaluate	Score
Technical Fit	Solves your actual problem, not adjacent problem	__ /5
Proof Quality	Performance on YOUR data, not demos	__ /5
Customer Evidence	References in your industry with similar scale	__ /5
Integration Readiness	APIs, documentation, compatibility with your stack	__ /5
Transparency	Willingness to discuss limitations and failures	__ /5
TCO Clarity	All-in pricing without hidden costs	__ /5

Scoring: 25+ = Proceed | 20-24 = Negotiate harder | Below 20 = Walk away

Key Takeaways

1■■
Demos Lie

Never buy based on demos alone. Test on YOUR data with YOUR edge cases.

2■■
Total Cost = 3-4x License

Implementation, integration, training, and maintenance are the real costs.

3■■
Ask Uncomfortable Questions

Good vendors welcome scrutiny. Evasion is a disqualifier.

4■■
Fix Foundations First

Bad data + great AI = expensive garbage. Readiness matters more than vendor.

LET'S CONNECT

Want the Full AI Due Diligence Framework?

I help executives cut through AI hype with battle-tested evaluation frameworks from $4B+ in M&A; transactions.

Connect with JJ →

bit.ly/jjshay

JJ Shay | Global Gauntlet AI
M&A; Executive → AI Strategy Consultant