

AI Operations Center — Full Demo Script

Audience: CFO, CIO, VP Finance, VP Supply Chain, IT Architecture board

Duration: 45 minutes (30 min demo + 15 min Q&A)

Setting: Single screen-share. Presenter drives the live app at <https://ai-operations-center-leovc4.abacusai.app>.

Pre-flight: Browser open to [/](#). Sidebar visible. Dark mode optional.

OPENING — 3 minutes

Frame the room (60 sec)

“Today I’m not going to show you a demo of an AI feature. I’m going to show you an **AI Operations Center** — the same way your ERP is an Operations Center for transactions, this is an Operations Center for **decisions**. Five agent swarms running across Trade Deductions, Collections, Invoice Triage, Buyer Assistance, and Demand Sensing. They’re not isolated tools — they share a governance plane, a value ledger, and a common enterprise trust framework. By the end of this 30 minutes you’ll have seen autonomy, control, and measurable value.”

The CFO question (60 sec)

“Every CFO I talk to asks two questions. **One:** ‘Are these agents actually right?’ **Two:** ‘How do I keep control as they get more autonomous?’ We’re going to answer both — with live data, not slides.”

Show the home page (60 sec)

- **CLICK:** [/](#) (Trade Deduction Intake Dashboard)
 - **POINT TO:** Live KPI cards (deductions today, \$ amount, classified, valid claims)
 - **NARRATE:** “Real database, real Prisma schema, real Decimal-precision financial data. This isn’t a mockup — every number you see is computed.”
-

ACT 1 — MVP1: TRADE DEDUCTION DEPTH (8 minutes)

1.1 The problem (45 sec)

“This \$4B food manufacturer takes ~\$120M per year in customer deductions. 41% of recoverable claims slip past the deadline. Their team triages 1,400 deductions a week. Each one is 11 days end-to-end.”

1.2 Multi-channel intake (90 sec)

- **CLICK:** Any deduction in the table → opens [/deduction/\[id\]](/deduction/[id])
- **POINT TO:** **Artifact Chain panel** (left side)
- EDI 812 — confidence 0.95
- PDF OCR — confidence 0.78
- Email free-form — confidence 0.42

- Portal JSON — confidence 0.99
- **SAY:** “Deductions arrive through five channels. The agent reads all of them, scores each artifact’s confidence, and merges. This is **PART C — Remittance Parser** in production.”

1.3 The 7-point Validity Engine (2 min)

- **POINT TO:** Trade Deal Validity card (right side)
- **WALK THROUGH:** Each of the 7 checks with their pass/fail icons
 1. Deal Active ✓
 2. SKU Eligible ✓
 3. Customer Eligible ✓
 4. Accrual Cap ⚠️
 5. Fund Balance ✓
 6. No Duplicate ✓
 7. Physical Reconciliation (3-way match) ✓
- **SAY:** “This is the difference between a chatbot that guesses and an enterprise agent that **shows its work**. Every check has an evidence JSON citation. Every score is auditable. SOX-compliant audit trail baked in.”
- **POINT TO:** Validity Score 0.84 → “Indeterminate” verdict
- **SAY:** “Anything below 0.95 routes to a human. Anything below 0.70 hard-stops for manual review.”

1.4 Closed-loop learning (90 sec)

- **CLICK:** Decision History tab
- **POINT TO:** Override Capture entry
- **SAY:** “When the SME rejected the agent’s first proposal, we captured **why**. That feedback feeds the prompt refinement pipeline. This is the loop that turns 87% agreement into 94% agreement over 90 days. **PART G — Closed-Loop Learning Pipeline.**”

1.5 Agreement Dashboard (90 sec)

- **CLICK:** Sidebar → “Agreement Monitoring”
- **POINT TO:**
 - Classification agreement: **93.4%** (≥92% bar) ✓ Enterprise grade
 - Validity agreement: **89.1%** (≥88% bar) ✓
 - 12-week trend line — agreement rising, override rate falling
- **SAY:** “Every CFO needs to see this number. Not ‘AI accuracy’ — **agreement-with-human**. It’s the only metric that matters at the boardroom level.”

1.6 Aged Deduction Escalation (60 sec)

- **CLICK:** Sidebar → “Escalations”
 - **POINT TO:** Escalation packs (60+, 90+ day)
 - **SAY:** “For deductions aged past recovery deadlines, the agent bundles evidence, drafts the escalation memo, and routes to AR / Legal / Audit. **89% recovery rate vs. 41% baseline.**”
-

ACT 2 — MVP3: INVOICE TRIAGE (5 minutes)

2.1 The problem (30 sec)

“AP gets 8,000 invoices a month. 22% have exceptions. Each exception takes 18 minutes of an AP analyst’s time. And buried in here are early-pay discount opportunities they miss.”

2.2 The agent swarm (90 sec)

- **CLICK:** Sidebar → “Invoice Exception” → `/invoices`
- **POINT TO:** Top KPI strip
- Auto-post rate: **74%**
- Exceptions: 1,847
- Duplicates blocked: 23 (\$412K saved)
- **Arbitrage savings captured: \$87K this month**
- **POINT TO:** Agent Swarm Visualizer
- **SAY:** “This is a multi-agent topology. Classifier → MatchPredictor → DuplicateDetector → ArbitrageScanner → Router. Each agent is independently observable.”

2.3 Three-way match (60 sec)

- **CLICK:** Any exception invoice
- **POINT TO:** PO/GR/Invoice line-level match table
- **SAY:** “Line by line. PO line 4 ordered 100 cases at \$42. GR received 96 cases. Invoice billed 100 at \$42. **Quantity short — 4 cases — auto-creates a deduction proposal.** Cross-MVP linkage to MVP1.”

2.4 Arbitrage capture (60 sec)

- **POINT TO:** Arbitrage Opportunity card
- **SAY:** “2/10 Net 30. Today is day 7. The agent has already validated the invoice and is proposing early payment. CFO approves with one click. **\$1,847 captured — and we caught 4,800 of these last quarter.**”

2.5 Boundary conditions (60 sec)

- **CLICK:** “Tune Policies” button
- **POINT TO:** Sliders for autonomy threshold, dollar caps, supplier whitelists
- **SAY:** “This is the **control answer**. The CFO sets the limits. The agent operates inside them. Below threshold = autonomous. Above threshold = HITL. Above hard limit = block. Every threshold has an audit trail.”

ACT 3 — MVP2: COLLECTIONS (4 minutes)

3.1 The problem (30 sec)

“DSO is 47 days. Industry benchmark is 38. Each day of DSO is \$1.1M of working capital tied up. Their 4 collectors can each handle ~150 accounts. They have 2,300 accounts past due.”

3.2 Daily portfolio rank (90 sec)

- **CLICK:** Sidebar → “Collections” → `/collection`

- **POINT TO:** Ranked worklist (top 50)
- **SAY:** “Every morning the agent ranks the entire portfolio by **expected value of a collector’s hour** — probability of payment × lift from contact × dollar at risk × strategic risk. **Customer #4 outranks Customer #1 because the model knows they always respond on Tuesday afternoons after a firm-tone email.**”

3.3 Strategy A/B testing (60 sec)

- **POINT TO:** Strategy column on each account (Strategy A vs. Strategy B)
- **SAY:** “Two cadences running in parallel. Strategy A is the control — fixed 5/15/30 day cadence. Strategy B uses learned customer behavior. The agent allocates traffic, measures recovery uplift, and recommends winners. **CollectionStrategy + CollectionInteraction tables track every send, every response, every promise-kept.**”

3.4 Email draft + escalation (60 sec)

- **CLICK:** Any account → opens detail panel
- **POINT TO:** Drafted email (firm tone, day-15)
- **POINT TO:** Escalation flag with reason: “Credit score dropped 80 pts in 30 days”
- **SAY:** “The agent doesn’t just rank. It drafts. It detects cadence breaks. It reads external credit signals. The collector approves and sends. Or rejects with one click — **and that rejection becomes a Golden Case.**”

ACT 4 — MVP4: BUYER ASSISTANCE (4 minutes)

4.1 The problem (30 sec)

“Procurement gets 12,000 buy requests a year. 30% are off-contract — a \$14M leakage. Category managers spend 60% of their time answering ‘is this on contract?’ instead of negotiating.”

4.2 Conversational PR creation (90 sec)

- **CLICK:** Sidebar → “Buyer Assistance” → /buyer
- **CLICK:** Demo thread “Guided PR Creation”
- **POINT TO:** Each assistant message with:
 - Citations (“per contract C-2024-0892, line 4”)
 - Pre-filled fields
 - Policy compliance ✓ / ⚠ / ✗
 - HITL flag
 - Agent trace
- **SAY:** “The buyer types one sentence. The agent fills the PR. **Every field has a citation. Every policy is checked. Every threshold is enforced. Off-contract is flagged before submission.**”

4.3 Policy engine (60 sec)

- **CLICK:** Demo thread “Off-Contract Purchase”
- **POINT TO:** Policy evaluation panel
- Spend threshold check: \$47K > \$25K → **CFO approval required**
- Supplier eligibility: ⚠ Not on preferred list
- Category restriction: ✓ MRO is allowed

- **SAY:** “This is the **BuyerPolicy** + **BuyerPolicyEvaluation** tables in action. Three enforcement modes — hard block, soft warn, log only — and an approval chain that knows who to route to based on the violation type.”

4.4 Procurement intelligence (45 sec)

- **CLICK:** Demo thread “Procurement Intelligence”
- **POINT TO:** Spend analytics by supplier, by category, contract usage gaps
- **SAY:** “This isn’t a chatbot. It’s a category manager’s analyst on demand.”

ACT 5 — MVP5: DEMAND SENSING (3 minutes)

5.1 The problem (30 sec)

“Forecast bias of +14% on average. That’s \$8M in carrying cost on excess inventory and \$4M in lost sales on stockouts. Planners override the forecast 62% of the time — usually too late.”

5.2 SKU × DC heatmap (60 sec)

- **CLICK:** Sidebar → “Demand Sensing” → `/demand`
- **POINT TO:** Heatmap — red cells = high deviation
- **SAY:** “3,200 SKU × DC pairs in one view. The agent flags only the cells that matter. **Click any cell — see the demand attribution: promo lift +18%, weather +6%, last-mile delay -3%.**”

5.3 Planner Co-pilot (90 sec)

- **CLICK:** Any cell → opens detail
- **POINT TO:** Inventory projection (4 weeks)
- **TYPE in copilot:** “What if I increase the promo lift assumption to 25%?”
- **POINT TO:** Live LLM response with revised projection
- **SAY:** “Planner stays in control. Agent does the math. **Every override is captured for the closed loop.**”

ACT 6 — THE GOVERNANCE LAYER (4 minutes)

6.1 Executive Dashboard (90 sec)

- **CLICK:** Sidebar → “Executive” → `/executive`
- **POINT TO:**
 - Cross-MVP KPIs
 - **Platform Trust Score: 0.91** (composite of all 5 MVP agreement rates)
 - Realized value: \$14.2M YTD
 - Projected value: \$42M annualized
 - Agent autonomy progression chart (advisor → autonomous-with-HITL → autonomous-with-monitoring)
- **SAY:** “This is the CFO’s view. **One trust score. One value ledger. Five agents.** Drill anywhere.”

6.2 My Work — unified worklist (60 sec)

- **CLICK:** Sidebar → “My Work” → `/my-work`

- **POINT TO:** HITL queue, NBA list, escalation inbox, value ledger widget
- **SAY:** “Every analyst — across deductions, collections, AP, procurement, planning — has **one** work-list. The agent prioritizes. They execute.”

6.3 Technical Reference (60 sec)

- **CLICK:** Sidebar → “Reference”
- **POINT TO:**
 - 5-layer architecture
 - MCP server registry (6 simulated tools)
 - Agent Development Lifecycle (6 phases)
 - Data contracts
 - Eval harness + cost meter + drift monitor
- **SAY:** “For your CIO and architecture board. Every system prompt versioned. Every cost metered. Every drift alert. **This is what enterprise-grade looks like under the hood.**”

6.4 Run the test suite live (30 sec)

- **OPEN:** New tab → `/api/test-runner`
- **POINT TO:** JSON response showing 19 test scenarios, all passing
- **SAY:** “Built-in regression suite. Run it before every prompt change.”

CLOSE — 4 minutes

The three-take-aways slide (90 sec)

“Three things you saw today:

One — Decision quality is measurable. Agreement-with-human $\geq 92\%$ on classification, $\geq 88\%$ on validity. Not ‘AI accuracy’ — enterprise grade.

Two — Control is built in. Boundary conditions, HITL queues, escalation inboxes, audit trails, policy engines, value ledgers. The CFO sets the rails. The agent operates inside them.

Three — Value compounds across MVPs. A deduction creates a collection. A collection feeds a credit signal. A credit signal informs a buyer policy. **One platform. One trust score. \$42M annualized.**”

The ask (60 sec)

“We’re proposing an 8-week pilot on MVP1 — Trade Deductions — with full enterprise depth as you saw today. Success criteria: $\geq 92\%$ agreement, $\geq 85\%$ recovery rate, ≤ 2 day cycle time. We measure weekly. You sign off the autonomy progression.”

Q&A primers (90 sec)

Common questions and short answers:

Q: “How do you handle a wrong decision?”

A: “Three layers. (1) Below validity threshold → human approves. (2) Above threshold but novel pattern → drift monitor flags. (3) Anything wrong → override captured → prompt refinement → A/B tested next cycle.”

Q: “What’s the integration lift?”

A: “MCP gateway abstracts your ERP. Six tools simulated today — replace with real EDI/SAP/Coupa/ etc. ~6 weeks per system.”

Q: “What about hallucinations?”

A: “Every agent decision requires evidence citation. No citation → no decision. The validity engine is deterministic logic over LLM-extracted features — not a free-form LLM verdict.”

Q: “Cost per decision?”

A: “Cost meter is live. Average \$0.03 per deduction triage. Budget caps per MVP. Hard stop on over-run.”

DEMO PRE-FLIGHT CHECKLIST

- ☐ Seed data loaded (`POST /api/seed-data`)
 - ☐ Boundary conditions seeded (auto on first GET)
 - ☐ Value ledger seeded
 - ☐ HITL queue has at least 5 open items
 - ☐ Escalation inbox has at least 2 critical items
 - ☐ At least one deduction has a complete artifact chain
 - ☐ At least one invoice shows arbitrage opportunity
 - ☐ Demand sensing has heatmap data
 - ☐ All 5 demo conversations in `/buyer` are seeded
 - ☐ Test runner returns success (`GET /api/test-runner`)
 - ☐ Browser zoom: 90% (fits more on screen)
 - ☐ Sidebar: open
 - ☐ Theme: light (better for projection)
-

ROOM CONTROL TIPS

- **If the room is skeptical of LLM accuracy** → spend longer on **Section 1.3 (validity engine)** and **Section 1.5 (agreement dashboard)**. The 7-point check makes the reliability story.
- **If the room is CFO-heavy** → emphasize **Section 6.1 (Executive Dashboard)** and **Section 4.3 (policy engine)**. The control story.
- **If the room is CIO-heavy** → spend longer on **Section 6.3 (Technical Reference)** and run the test suite live.
- **If the room is operations-heavy** → spend longer on **Sections 2-3 (Invoice + Collections)**. Show the worklist, not the architecture.

- **If you have only 20 minutes** → MVP1 (8 min) + Executive Dashboard (3 min) + Boundary conditions / Policy engine (4 min) + Close (5 min). Skip MVPs 2/3/4/5 detail.