Quality Management & AI Operations Portfolio

Brian Geiser | Quality Management | AI Operations | Conversation Analytics | QA Governance

Johnstown, PA | bgeiser90@gmail.com

Portfolio note

The examples are anonymized and representative. They demonstrate my approach to interaction evaluation, AI QA governance, calibration, reporting, root cause analysis, action planning, and operational improvement without proprietary employer data, client data, member data, customer data, or private customer information. Dashboard values marked illustrative are sample operating metrics, not client-specific results. Exact values in impact examples have been generalized to protect client confidentiality.

Portfolio at a glance

Quality management, AI QA governance, conversation analytics, and CX improvement.	AI-supported QA validation accuracy above 95%.
QA productivity improved by approximately 2x.	Manual review/reporting effort reduced by more than 50%.
Expanded automated QA coverage across a majority of reviewed interactions.	Contributed to double-digit CSAT improvement.
Closed-loop model connecting QA, AI Ops, Training, Knowledge, Product, Operations, and Leadership.	Designed for Quality Manager, AI Quality, CX Analytics, and AI Operations-adjacent roles.

Positioning statement

This portfolio is aligned to Quality Manager and AI Operations Manager roles. It shows how I connect customer interaction review, AI-supported QA validation, calibration governance, root cause analysis, AI operations controls, and leadership reporting into a closed-loop operating model. The goal is not simply to generate more QA scores; it is to identify customer-impacting risk, assign clear ownership, and drive measurable improvement.

Best aligned roles

Quality Manager	AI Operations Manager	AI Quality Manager	CX Quality Manager
Conversation Analytics Manager	QA Governance Lead	AI Evaluation Lead	Implementation Manager - AI QA / CX Analytics

What this portfolio demonstrates

- Evaluation of chat, voice, AI-agent, and human-assisted support interactions for accuracy, customer trust, policy adherence, escalation handling, resolution quality, and risk controls.

- Translation of QA findings into structured insight reports that leadership, operations, training, knowledge, product, and AI operations teams can act on.

- Human-in-the-loop AI QA governance, including validation, guardrails, calibration, exception handling, reason codes, and action routing.

- Practical scorecards, calibration routines, dashboard concepts, impact examples, and improvement loops for high-volume customer support environments.

Relevant background

Quality and Customer Experience leader with experience across contact center QA, AI-enabled quality monitoring, AI operations controls, AQM implementation support, call/chat evaluation, calibration, compliance monitoring, root cause analysis, dashboard reporting, and cross-functional improvement. Tools and environments include Snowfly AQM, Observe.AI / ObserveAI, CallMiner, Salesforce, LivePerson, Power BI, Tableau, Excel/Google Sheets, and remote BPO/customer support operations.

Portfolio contents

Section	What it demonstrates
1. AI Support QA Scorecard	Evaluation structure for AI, agent-assisted, and human support interactions, including outcome quality, risk controls, failure signals, and reason codes.
2. Customer Interaction Case Study	Signal detection, root cause analysis, validation logic, and action planning for an anonymized AI chat with human handoff.
3. Quality Signal Insight Report	Leadership-ready reporting that converts QA trends into business impact, owner assignment, recommended action, and success measures.
4. Calibration & Rubric Governance	Scoring consistency, reviewer alignment, disputed criteria review, rubric maintenance, and AI/human validation governance.
5. AI Quality Governance Model	Human validation, AI QA controls, exception handling, failure monitoring, action routing, and closed-loop quality review.
6. Dashboard Mockup & Metrics	Operating metrics, targets, KPI interpretation, and owner/action tracking for CX quality and AI quality monitoring.
7. Anonymized Transformation Snapshot	Before/action/after impact story connecting quality transformation to generalized measurable outcomes.
8. Selected Impact Examples	Public-safe proof points and measurable outcomes across AI QA coverage, accuracy, productivity, risk visibility, CSAT, and cost efficiency.
9. First 90 Days	How I would onboard, assess, validate, improve, and establish a repeatable quality operating rhythm in a new environment.

1 AI Support QA Scorecard

Use case: Evaluate AI chat, AI voice, agent-assisted, or human support conversations for quality, customer trust, and operational risk. The scorecard supports blended human/AI handoff experiences and separates outcome quality from checklist completion.

Dimension	Evaluation standard	Example failure signals	Weight
Issue understanding	The interaction identifies the customer's actual need, intent, context, and prior steps.	Misread intent; answered only part of the issue; ignored history or previous attempts.	15%
Accuracy & policy alignment	The answer is correct, policy-compliant, current, and aligned to approved guidance.	Incorrect resolution; outdated policy; unsupported promise; hallucinated information.	20%
Resolution quality	The interaction solves the issue or moves it to the correct next step with ownership and timing.	No clear resolution; vague next steps; unnecessary repeat contact; unresolved intent.	20%
Escalation & handoff	Escalation occurs at the right time and includes useful context for the receiving agent/team.	Late escalation; failed transfer; weak handoff summary; customer repeats information.	15%
Customer effort & clarity	The explanation is easy to follow, concise, and actionable.	Looping; unclear explanation; too many steps; generic or irrelevant response.	10%
Empathy & tone	The interaction shows ownership, professionalism, and customer-centered language.	Cold response; missed emotional cue; robotic or dismissive tone.	10%
Risk controls	Sensitive, compliance, refund, account, or safety issues are handled correctly.	PII risk; compliance miss; wrong account action; missed risk escalation.	10%

Operating principle

The score is not the final answer. The value comes from the reason code, failure pattern, business impact, owner, and recommended action. A 78% score with a clear root cause and action plan is more valuable than a 95% score that nobody can interpret.

Reason code examples

Reason code	Definition	Action path
POLICY_GAP	Guidance is unclear, conflicting, outdated, or difficult to apply.	Clarify SOP/KB, align Training and Ops, and update scorecard language if needed.
ESCALATION_MISS	The issue should have been escalated or handed off sooner.	Review trigger logic, routing rules, coaching guidance, and escalation criteria.
AI_CONTEXT_MISS	AI output missed prior context, customer intent, or conversation nuance.	Validate prompt/knowledge source, add examples, and flag for model/tool tuning.
WORKFLOW_DEFECT	Agent/AI followed the available process, but the process created customer friction.	Route to Product, Ops, Programs, or Knowledge for workflow redesign or policy simplification.

2 Representative Customer Interaction Case Study

This anonymized example demonstrates the review logic behind the scorecard: identify customer intent, isolate the breakdown, separate symptoms from root cause, validate the signal, and route action to the correct owner.

Interaction type	Anonymized AI chat with human handoff
Customer intent	Customer needed help resolving an account/payment issue and wanted confirmation that no duplicate charge would occur.
Primary signal	The AI repeated generic troubleshooting after the customer had already completed those steps.
Customer impact	Higher effort, repeated information, reduced trust in self-service, longer handle time, and higher DSAT/repeat-contact risk.

Breakdown observed

- The first AI response was topically relevant but not context-aware. It answered the category, not the customer's specific concern.

- The interaction showed a loop pattern: the customer restated the issue, but the AI repeated previous guidance instead of changing path.

- Escalation appeared to depend on explicit customer request rather than repeated failed-resolution signals.

- The handoff summary lacked attempted steps, risk/concern, and recommended next action, forcing the customer to repeat details.

Root cause analysis

Hypothesis	Evidence to validate	Recommended action
Weak failure-detection trigger	Repeated customer clarification did not trigger escalation or path change.	Add loop-detection QA flag for repeated customer restatement within the same issue category.
Knowledge article mismatch	AI selected generic guidance rather than account/payment-specific guidance.	Map knowledge source used against issue category and refresh retrieval examples.
Incomplete handoff summary	Human agent did not receive attempted steps, customer concern, or recommended next action.	Define required handoff fields: intent, steps attempted, risk/concern, and next action.

Action plan

Action	Owner	Measure of success
Create calibration sample set for this issue type so QA, Ops, Training, and Product agree on good resolution.	QA + Ops + Training	Reviewer agreement, dispute rate, coaching adoption.
Track weekly dashboard signals: loop rate, late escalation rate, transfer-summary completeness, repeat-contact risk, and DSAT mentions.	QA + AI Ops	Downward trend in loop defects and late escalation.
Route knowledge and workflow defects to accountable owners with due dates.	Knowledge + Product + Programs	Action closure rate and defect recurrence.

3 Quality Signal Insight Report

Audience: CX leadership, Operations, Training/Enablement, Knowledge, Product, AI Ops, and QA stakeholders. The purpose is to convert interaction reviews into a decision-ready signal: what changed, why it matters, who owns it, and how success will be measured.

Executive signal

Quality reviews show that customers are not primarily dissatisfied because agents lack empathy. The stronger pattern is resolution uncertainty: customers leave interactions unclear on next steps, ownership, timing, or whether the issue is truly resolved.

Signal	Trend observed	Likely impact	Recommended owner
Unclear next steps	Customers ask follow-up questions after a supposed resolution.	Repeat contacts, lower CSAT, longer handle time.	Training + Knowledge
Late escalation	Escalation happens only after customer frustration is explicit.	Higher DSAT, lower trust, avoidable customer effort.	Ops + QA + Product
Policy interpretation variance	Different agents apply the same guidance differently.	Inconsistent experience and coaching confusion.	QA + Enablement
AI summary gaps	Human agents receive incomplete context after handoff.	Repeat explanation, longer resolution, lower customer confidence.	Product + AI Ops

Recommended leadership actions

Priority	Action	Expected outcome	Measure
1	Refresh rubric language for resolution confirmation and next-step clarity.	More consistent evaluation and coaching.	QA variance, calibration agreement.
2	Update knowledge content with approved next-step phrasing by scenario.	Cleaner customer communication.	CSAT comments, repeat-contact rate.
3	Create escalation triggers for repeated restatement or unresolved intent.	Faster appropriate escalation.	Late escalation rate, containment quality.
4	Add human-handoff summary completeness to QA review.	Better transfers and lower effort.	Handoff defect rate, AHT, DSAT mentions.

Leadership communication approach

- Start with customer impact, not the QA score.

- Separate agent behavior, policy ambiguity, workflow defects, and AI/system behavior so owners are clear.

- Give leadership a decision-ready summary: signal, root cause, recommendation, owner, due date, and measure of success.

- Close the loop in the next WBR/MBR by showing whether the action changed quality outcomes.

4 Calibration & Rubric Governance Framework

Calibration is where QA becomes credible. The framework makes evaluation standards explicit, tests them against real conversations, and keeps rubrics aligned with policy, customer impact, and operational priorities.

Cadence	Activity	Output
Weekly	Reviewer calibration using 5-10 representative interactions.	Alignment score, disputed criteria, examples for coaching.
Biweekly	Rubric issue review with QA, Ops, and Training.	Scorecard wording updates, reason code refinements, coaching notes.
Monthly	Leadership quality readout.	Top trends, customer impact, root cause themes, action-plan progress.
Quarterly	Framework governance review.	Retire stale criteria, add new priorities, validate AI scoring accuracy.

Rubric governance checks

- Is each criterion observable in the interaction, or is it subjective?

- Does the scorecard evaluate customer outcome, not just agent behavior?

- Do reviewers agree on good, acceptable, and failing examples?

- Does the rubric reflect current policy, current product behavior, and current customer risk?

- Are AI-generated scores validated against human-reviewed examples before leadership treats them as operational truth?

Calibration scorecard

Metric	Target	Why it matters
Reviewer agreement	85%+ on key dimensions	Shows whether the rubric is being applied consistently.
Variance by criterion	Identify top 3 disputed criteria	Finds vague wording, unclear examples, or inconsistent interpretation.
AI/human agreement	Track by issue type and severity	Validates whether AI-assisted QA can be trusted for the use case.
Action closure rate	90%+ by due date	Prevents QA insights from becoming passive reporting.

Practical principle

If a scorecard cannot explain why performance changed, it is not finished. QA frameworks should create consistent evaluation and better decisions, not just more scores.

5 AI Quality Governance Model

AI support systems need quality controls after launch. This operating model treats AI output as a production process that must be monitored, validated, calibrated, and improved continuously.

1. Capture	2. AI QA Scoring	3. Human Validation	4. Calibration	5. Action Routing	6. Closed-Loop Review

Stage	Control	Owner group	Evidence
Interaction capture	Collect AI chat/voice, metadata, outcomes, escalations, CSAT/DSAT, and repeat-contact signals.	CX Ops / Data	Interaction sample and dashboard inputs.
AI QA scoring	Use AI-supported evaluation for broad coverage, signal detection, reason codes, and risk flags.	QA / AI Ops	Automated QA output, reason codes, risk flags.
Human validation	Review high-risk, low-confidence, disputed, or customer-impacting examples.	QA Leads	Validation sample, accuracy notes, exceptions.
Calibration	Compare AI output, human review, policy, and business priorities.	QA + Ops + Enablement	Calibration notes and rubric updates.
Action routing	Route findings to Training, Knowledge, Product, Engineering, Ops, or BPO partners.	Program owners	Action plan with owner and due date.
Closed-loop review	Measure whether changes reduce defects, DSAT drivers, escalations, and repeat contacts.	CX Leadership	WBR/MBR insight report.

Failure types monitored

Failure type	Example signal	Why it matters
Hallucination / unsupported claim	AI promises something policy does not support.	Customer trust and compliance risk.
Looping behavior	Customer repeats the same issue and AI repeats the same guidance.	Customer effort and DSAT risk.
Incorrect resolution	AI closes or contains issue without solving the actual intent.	Repeat contacts and operational risk.
Poor handoff	Human agent receives incomplete context.	Longer handle time and weaker experience.
Policy gap	AI/agent cannot provide consistent answer due to unclear guidance.	Training, content, and product improvement opportunity.

Governance mindset

Automation coverage does not prove AI quality. I look for accuracy, explainability, customer impact, exception handling, calibration alignment, and whether AI output drives better operational decisions.

6 Dashboard Mockup & Quality Operating Metrics

A useful dashboard answers three questions: What changed? Why did it change? What should we do next? The view below uses representative metrics for CX QA and AI quality monitoring. Values are illustrative and should be replaced with verified program data when available.

87.4%	**78%**	**12.6%**
QA score	**AI/human agreement**	**Late escalation rate**
Target: 90%+	Target: 85%+	Target: <7%
21%	**18%**	**64%**
Handoff defect rate	**Unclear next steps DSAT**	**Action-plan closure**
Target: <10%	Target: <10%	Target: 90%+

Metric interpretation

Metric	Current state	Target	Interpretation / action
QA score	87.4%	90%+	Review top defect categories and calibration variance.
DSAT driver: unclear next steps	18% of DSAT sample	<10%	Update KB guidance and coaching examples.
Late escalation rate	12.6%	<7%	Refine escalation triggers and monitor repeated-intent cases.
AI/human scoring agreement	78%	85%+	Review low-agreement criteria and refresh examples.
Handoff summary defect rate	21%	<10%	Add handoff summary criteria and route product feedback.
Action-plan closure	64%	90%+	Tighten owner/due-date tracking and leadership review.

Owner/action tracker

Priority	Quality signal	Owner	Next action	Due
1	Unclear next steps	Training + Knowledge	Publish approved next-step phrasing by scenario.	2 weeks
2	Late escalation	Ops + QA + Product	Validate repeated-intent trigger and escalation criteria.	3 weeks
3	AI summary gaps	Product + AI Ops	Define handoff summary fields and audit transfer payload.	4 weeks

7 Anonymized Transformation Snapshot

This snapshot shows how I describe quality transformation without exposing client names, program details, proprietary workflows, or exact client-specific data. The values below are generalized to protect confidentiality while preserving the scale of impact.

Transformation theme

Move quality from low-visibility manual review into an AI-supported, human-validated, closed-loop operating model that improves customer outcomes, reviewer productivity, and leadership decision quality.

Before	Actions taken	After / public-safe impact
- Manual QA sampling limited visibility into defect patterns. - Scoring and coaching could become checklist-driven instead of signal-driven. - Calibration variance made some findings harder to trust. - Defects were not always routed to the accountable owner. - AI/AQM outputs needed validation before leadership use.	- Supported AI/AQM quality monitoring and validation controls. - Built reason-code logic to separate agent behavior, policy gaps, workflow defects, and AI/system issues. - Established human-in-the-loop review for high-risk, low-confidence, and disputed outputs. - Connected QA findings to dashboards, calibration routines, owner-based action plans, and leadership readouts. - Routed recurring defects to Training, Knowledge, Product, Ops, AI Ops, or program owners.	- Maintained AI-supported QA validation accuracy above 95%. - Expanded automated QA coverage across a majority of reviewed interactions. - Improved QA productivity by approximately 2x. - Reduced manual review/reporting effort by more than 50%. - Contributed to double-digit CSAT improvement. - Helped move QA performance from below target to consistently above passing standards. - Reduced cost-per-call materially through quality/process improvements.

What this proves

- I can enter a quality or AI-supported operation, separate symptoms from root cause, and build a trusted operating rhythm.

- I do not treat AI-generated QA scores as automatically reliable; I validate accuracy, exceptions, and business impact before operationalizing them.

- I connect quality findings to measurable ownership: coaching, knowledge updates, escalation logic, workflow improvement, product feedback, and leadership review.

8 Selected Impact Examples - Anonymized

The examples below summarize public-safe impact themes from quality management, AI-enabled QA, calibration, and customer experience operations. Client names, program names, proprietary workflows, and exact client-specific data have been removed or generalized.

AI-enabled QA coverage

Supported AI/AQM quality monitoring efforts that expanded visibility beyond traditional manual sampling and enabled broader detection of customer-impacting defect patterns.

Expanded automated QA coverage across a majority of reviewed interactions while maintaining human validation controls.

QA accuracy and validation

Validated AI-supported quality outputs through human-in-the-loop review, calibration, and exception handling before treating automated results as operational truth.

Maintained AI-supported QA validation accuracy above 95% while reducing risk from overreliance on automation.

QA productivity

Shifted quality review from manual checklist scoring toward signal detection, reason-code analysis, exception review, and structured action planning.

Improved QA productivity by approximately 2x and reduced manual review/reporting effort by more than 50%.

Risk and compliance visibility

Strengthened review coverage for priority interaction types involving compliance, account actions, escalation risk, customer trust, and resolution quality.

Improved risk coverage for priority interaction types and created clearer paths for escalation, coaching, and corrective action.

Customer experience improvement

Connected QA findings to root cause themes such as unclear next steps, inconsistent policy interpretation, late escalation, weak handoff summaries, and avoidable customer effort.

Contributed to double-digit CSAT improvement by linking QA defects to coaching, knowledge updates, and operational action plans.

Performance improvement

Used calibration, scorecard refinement, coaching loops, and leadership reporting to move quality insights from passive reporting into accountable improvement routines.

Helped improve QA performance from below target to consistently above passing standards.

Cost and operational efficiency

Identified process friction, repeat-contact drivers, and avoidable effort patterns that increased support cost and reduced customer confidence.

Reduced cost-per-call materially through quality, process, and reporting improvements.

Cross-functional action planning

Translated interaction-level findings into leadership-ready insight reports with owners, action paths, success measures, and follow-up cadence.

Improved ownership across Training, Knowledge, Product, Operations, AI Ops, and leadership review cycles.

9 First 90 Days & Portfolio Takeaway

This operating plan is designed for entering a new quality, AI QA, or CX analytics environment: learn the current system, validate what can be trusted, identify the highest-risk defects, and create a sustainable operating rhythm.

Timeframe	Focus	Deliverables
First 30 days	Learn workflows, scorecards, channels, tools, stakeholders, and current pain points.	QA process map, current-state risk list, sample review notes, quick-win recommendations.
Days 31-60	Validate rubric logic, calibration health, AI QA output, dashboard gaps, and ownership loops.	Calibration summary, reason-code structure, insight report template, dashboard improvement plan.
Days 61-90	Implement repeatable quality governance and decision-ready reporting.	Monthly quality readout, action-plan tracker, AI QA validation cadence, stakeholder feedback loop.

Relevant experience signal

- Experience with QA operations, calibration, interaction evaluation, compliance monitoring, root cause analysis, dashboard reporting, and cross-functional quality improvement.

- Exposure to AI-enabled quality monitoring and conversation intelligence environments, including AQM implementation support and validation of quality signals.

- AI operations orientation: validation logic, failure monitoring, exception review, issue routing, stakeholder readouts, and closed-loop improvement.

- Practical operating model for connecting QA findings to training, knowledge, product, operations, AI operations, and leadership review cycles.

Portfolio takeaway

> My quality and AI operations approach connects AI-supported evaluation, human validation, calibration, root cause analysis, issue routing, and leadership reporting into a closed-loop operating model. The emphasis is on customer impact, accountable ownership, human-in-the-loop AI validation, and measurable improvement - not simply producing more scores.

Suggested live discussion topics

- Validating AI QA accuracy before leadership relies on automated scores.

- How I separate agent coaching issues from policy, workflow, knowledge, or AI-system defects.

- Building a 30/60/90 quality improvement plan for a new AI-supported contact center environment.