

From AI Pilot to Operating Discipline: A Dual-Loop Model for Throughput, Resilience, and Quality Control in Class II Medical Devices

Kenneth Xu

Doctorate Pan-European Studies, University of Pula
Los Angeles CA, USA
kxu124@gmail.com

Markus Mau and Nick Vyas

Doctorate Pan-European Studies
University of Pula, USA

Abstract

Small and mid-sized Class II medical device firms face a structural operating challenge. Regulatory evidence, supplier records, quality files, and launch decisions must be coordinated across markets, but compliance capacity often grows more slowly than operating complexity. This paper develops a dual-loop AI operating discipline for Class II diabetes device manufacturing. The model links a Regulatory Readiness Loop with a Supply Adaptation Loop through human review gates, controlled evidence retrieval, supplier early-warning scoring, and audit-ready quality-system records. The study uses a six-month case pilot at New Horizon Biotech, where a five-FTE AI Business Unit supported FDA 510(k) and EU MDR workflows without replacing expert judgment or formal quality controls. Results show that regulatory drafting cycle time decreased from 6.5 to 4.4 weeks, defect density fell from 2.8 to 0.9 errors per page, review iterations decreased from 4.2 to 2.5 cycles, and evidence retrieval time decreased from 12.5 to 1.5 hours. Supply outcomes also improved. Lead-time recovery decreased from 8.5 to 5.2 weeks, sole-source exposure declined from 42% to 34% of the bill of materials, and critical stock-outs decreased from two events to zero. The framework reached a median groundedness score of 0.94 by Month 6 and generated an estimated 82.4% six-month ROI. The paper contributes an operations management model for applying AI as a governed process-control layer in regulated manufacturing.

Keywords

Artificial Intelligence, Operations Management, Supply Chain Resilience, Quality Control, Medical Devices.

1. Introduction

Class II medical device manufacturers operate between two pressures. They must move quickly enough to compete in fast-moving diabetes technology markets, but they must also maintain strict regulatory and quality discipline. This tension is visible in products such as continuous glucose monitors, insulin pumps, and blood glucose monitoring systems. These products rely on software, electronics, sensors, adhesives, suppliers, labeling controls, performance evidence, and post-market monitoring. A delay in any one area can affect market access, inventory planning, and patient availability.

The core problem addressed in this paper is the scale-up paradox. As a smaller Class II device firm expands across markets, the amount of regulatory, quality, and supply chain work can grow faster than the number of professionals available to manage it. Larger firms can add regulatory staff, consultants, and enterprise platforms. Smaller firms often cannot. They need operating models that increase coordination and decision speed without weakening compliance.

AI can help, but only under strict operating discipline. In regulated manufacturing, AI cannot be used as an uncontrolled text generator or informal decision tool. Regulatory statements must be traceable to source evidence. Supplier-risk decisions must be explainable. Quality records must remain controlled. Human reviewers must approve claims, changes, and risk decisions. For these reasons, this paper treats AI as a governed process-control layer rather than a replacement for expert judgment.

This study develops and evaluates a dual-loop AI model at New Horizon Biotech, a small Class II diabetes device manufacturer. The model connects regulatory readiness and supply adaptation. The Regulatory Readiness Loop, or R-Loop, supports evidence retrieval, predicate comparison, drafting support, and EU MDR General Safety and Performance Requirements mapping. The Supply Adaptation Loop, or S-Loop, supports supplier early-warning scores, alternate-source readiness, and total landed cost analysis. The case evaluates whether a five-FTE AI Business Unit can improve operating performance while preserving quality-system accountability.

1.1 Objectives

This paper has four objectives. First, it develops an AI-enabled operating model for linking regulatory readiness and supply chain resilience in Class II device operations. Second, it evaluates whether AI-assisted workflows can reduce regulatory drafting cycle time by at least 25% while lowering or maintaining defect levels. Third, it evaluates whether supplier early-warning scoring improves supply recovery and reduces critical stock-out risk. Fourth, it validates whether the model remains auditable and financially viable through groundedness scoring, human review, and ROI analysis.

Table 1 summarizes the objectives, questions, measures, and evidence sources. The table also shows that the study does not define success by speed alone. In regulated operations, a faster process only has value if it also improves quality, traceability, and decision control.

Table 1. Research objectives, questions, measures, and evidence sources

Objective	Operational Question	Primary Measures	Evidence Source
Objective 1	Can regulatory and supply functions be linked through one AI-enabled operating model?	R-Loop, S-Loop, gates, human review controls	Framework design and pilot workflow
Objective 2	Can AI reduce regulatory latency without reducing quality?	Drafting time, defects per page, review cycles, retrieval time	Baseline and pilot logs
Objective 3	Can supplier early-warning improve resilience?	Lead-time recovery, sole-source exposure, stock-outs	Supplier and planning records
Objective 4	Can the model remain auditable and financially viable?	Groundedness, red-team tests, ROI, net value	Governance and financial review

2. Literature Review

Medical device operations sit at the intersection of regulatory control, quality management, and supply chain execution. FDA quality-system requirements and ISO 13485-based practices require controlled records, design controls, supplier controls, corrective actions, and traceable decision-making (U.S. Food and Drug Administration 2024; International Organization for Standardization 2016). EU MDR adds further evidence obligations through technical documentation, clinical evaluation, post-market surveillance, and General Safety and Performance Requirements mapping (European Parliament and Council 2017). These requirements make documentation quality and evidence traceability central parts of operating performance.

Prior research on regulatory systems shows that fragmented requirements across markets can delay medical device commercialization. Amaral et al. (2024) explain that country-specific requirements may slow innovation when firms must adapt evidence packages to different approval systems. Ahmadi (2024) also argues that AI-driven predictive analytics can support medical device risk management by identifying early warning signs before hazards become severe. Together, these findings support the view that compliance should be managed as an operating system, not as a final paperwork step.

Supply chain resilience research adds a second foundation. Christopher and Peck (2004) emphasize visibility, collaboration, and agility as resilience capabilities. Pettit et al. (2010) frame resilience as the balance between vulnerabilities and capabilities. The National Academies of Sciences, Engineering, and Medicine (2022) further argues that resilience depends on better coordination of supply and demand, not only higher inventory. This logic is highly relevant to Class II diabetes devices because supplier qualification, component availability, and regulatory evidence are connected.

AI and digital transformation literature adds a third foundation. McKinsey & Company (2024) describes how generative AI can summarize large datasets and support regulated response drafting. KPMG (2024) emphasizes that digital supply chain value depends on data quality, system integration, and reliable operating processes. These points are important because AI cannot solve poor data discipline by itself. It can accelerate work only when source records are accurate, current, and governed.

AI governance literature also matters. The International Medical Device Regulators Forum (2022) highlights lifecycle documentation and terminology discipline for machine learning-enabled systems. NIST (2023) proposes an AI risk framework built around mapping, measuring, managing, and governing AI risks. These sources clarify why regulated firms must constrain AI outputs through source evidence, human approval, and monitoring.

The literature therefore identifies a practical gap. Existing studies explain the value of AI, regulatory intelligence, quality control, and resilient supply chains. Fewer studies show how a small or mid-sized Class II manufacturer can combine these ideas into a lean operating model. This paper addresses that gap by presenting a dual-loop model with measured pilot outcomes.

2. Methods

This study uses an applied case-study method with baseline comparison. The case is New Horizon Biotech, a Class II diabetes device manufacturer with Western regulatory obligations and a global supply base. The case method is appropriate because the study examines a live operating model where regulatory, quality, supply chain, and technology decisions interact.

The method has four steps. First, the study defines the dual-loop model. The R-Loop supports regulatory evidence retrieval, predicate comparison, drafting support, and GSPR mapping. The S-Loop supports supplier risk scoring, alternate-source readiness, and scenario modeling. Second, the study defines review gates. AI may retrieve evidence, suggest mappings, and flag supplier risks, but human reviewers must approve regulatory claims and operational decisions. Third, the study compares baseline manual workflows against pilot AI-assisted workflows. Fourth, the study validates the model through threshold tests, groundedness scoring, and financial analysis.

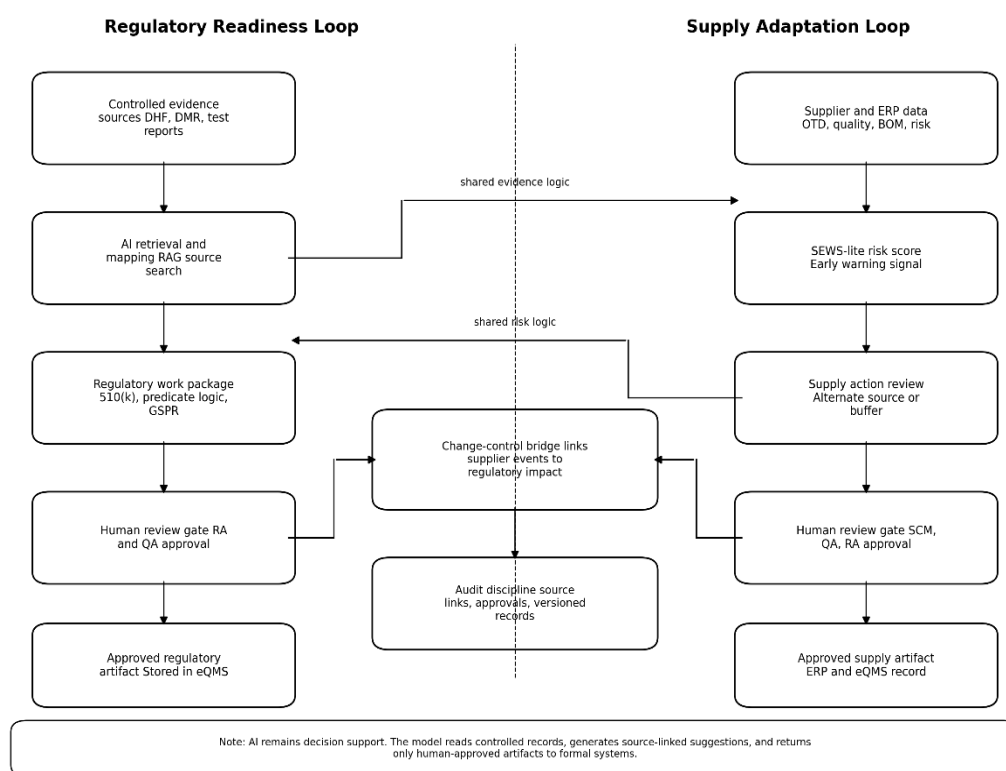


Figure 1. Dual-loop operating discipline for regulated Class II device operations

Figure 1 shows the logic of the proposed model. The AI layer does not replace ERP, eQMS, or expert review. Instead, it reads from controlled records, creates decision-support outputs, and returns approved artifacts to formal systems. This design keeps AI inside the quality-system boundary rather than allowing it to operate as an informal external tool.

The study tested three hypotheses.

H1: AI-assisted regulatory authoring will reduce drafting cycle time by at least 25% compared with the manual baseline, while maintaining or reducing defect density.

H2: Supplier early-warning scoring will improve lead-time recovery and eliminate critical stock-out events for monitored components during the pilot window.

H3: The five-FTE AI Business Unit will generate a positive six-month ROI when labor efficiency, inventory optimization, and launch acceleration are compared against pilot costs.

The study does not claim statistical generalization to all device firms. The data are operational performance records from one pilot. The analysis therefore uses effect-size comparison against defined thresholds, supported by governance and financial validation.

3. Data Collection

The pilot ran for six months and focused on Western regulatory and supply chain operations. The regulatory scope included FDA 510(k) preparation logic and EU MDR technical documentation support. China NMPA and Australia TGA activities were excluded to avoid language, data-localization, and third-party market access confounders.

The product scope included three Class II diabetes product conditions. P1 HorizonFlow was a mature insulin pump with legacy component risk. P2 HorizonGuard was a continuous glucose monitor undergoing EU MDR evidence mapping. P3 was a blood glucose monitoring system used to test new product and predicate-suitability logic. This portfolio allowed the pilot to test the model across legacy, high-volume, and new-product conditions.

The AI Business Unit had five FTEs: a product owner, a regulatory affairs systems lead, a data operations analyst, an AI solution engineer, and a supply chain analytics lead. The technology model followed a hybrid build-plus-buy approach. Internal build components supported retrieval against controlled Design History File and Device Master Record sources. External enterprise tools supported supply network simulation, multilingual drafting, and workflow scaling. Public consumer AI tools were excluded from proprietary workflows.

Table 2 explains the data model. The data were organized around regulatory throughput, supply resilience, governance, and financial performance. This structure was important because AI adoption in medical device operations cannot be judged by productivity alone.

Table 2. Data collection design and measurement logic

Data Category	Unit of Observation	Example Fields	Frequency	Purpose
Regulatory throughput	Submission or document package	Cycle time, errors per page, review cycles, retrieval hours, GSPR completeness	Event-based	Measures R-Loop speed and quality
Supply resilience	Supplier-week	SEWS score, delay days, dual-sourcing status, stock-outs	Weekly	Measures S-Loop response
Governance and auditability	AI output or review event	Groundedness, citation check, red-team result, approval status	Monthly and event-based	Measures AI trust and control
Financial performance	Pilot cost and value stream	Labor efficiency, inventory release, launch acceleration, net ROI	End of pilot	Measures financial viability

Before deployment, the team completed a data hygiene phase. This included standardizing supplier names, digitizing regulatory documents, cleaning duplicate vendor records, and preparing controlled repositories for retrieval. This work was not optional. AI quality depended on record quality. If records were inconsistent, the model could produce confident but incorrect risk conclusions.

5. Results and Discussion

5.1 Numerical Results

The pilot produced measurable improvements in regulatory throughput and supply resilience. Table 3 reports the primary baseline and pilot values. Improvement percentage is calculated as:

Improvement (%) = (Baseline Value - Pilot Value) / Baseline Value x 100

For GSPR completeness, the table reports the percentage-point increase. For critical stock-outs, the result is interpreted as event avoidance.

Table 3. Baseline and pilot results across regulatory, supply, governance, and financial measures

Metric	Baseline	Pilot	Improvement or Outcome
Drafting cycle time	6.5 weeks	4.4 weeks	32.3% reduction
Defect density	2.8 errors/page	0.9 errors/page	67.8% reduction
Review iterations	4.2 cycles	2.5 cycles	40.4% reduction
Evidence retrieval time	12.5 hours	1.5 hours	88.0% reduction
GSPR first-pass completeness	65%	95%	30 percentage-point increase
Lead-time recovery	8.5 weeks	5.2 weeks	38.8% faster
Sole-source exposure	42% of BOM	34% of BOM	8 percentage-point decrease
Critical stock-outs	2 events	0 events	100% avoidance
Median groundedness score	Not systematic	0.94 by Month 6	Within target range
Six-month ROI	Not applicable	82.4%	Positive ROI

Table 3 shows that the largest regulatory gain came from evidence retrieval. The reduction from 12.5 hours to 1.5 hours indicates that the R-Loop reduced search burden before drafting began. This matters because many regulatory delays do not begin with writing. They begin with finding the correct evidence, confirming its current status, and linking it to the correct regulatory claim.

The quality result is also important. Drafting time decreased, but defect density also fell. This suggests that the model did not merely accelerate low-quality output. It improved the review starting point by giving writers source-linked evidence and structured draft scaffolds. In practical terms, regulatory staff moved from search-and-create work toward verify-and-refine work.

The supply results show a similar pattern. Lead-time recovery improved by 38.8%, and critical stock-outs decreased from two events to zero. The reduction in sole-source exposure from 42% to 34% also suggests that supplier visibility translated into action. The S-Loop was not only a reporting dashboard. It supported alternate-source decisions before disruption became visible through missed delivery.

5.2 Graphical Results

The numerical results are summarized visually in Figures 2 through 4.

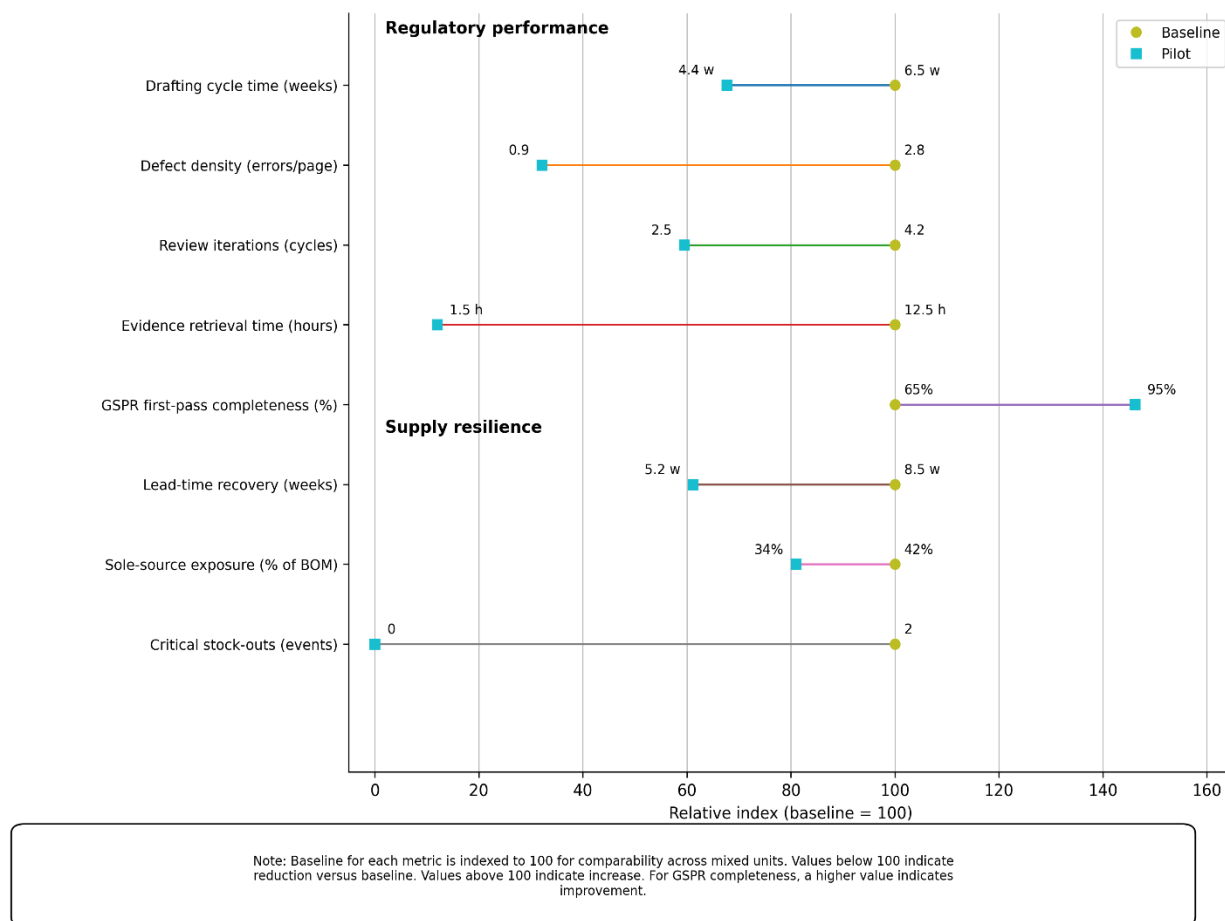


Figure 2. Baseline and pilot performance comparison across primary operating metrics

Figure 2 shows the operational pattern clearly. Every primary metric moved in the desired direction. Regulatory cycle time, defects, review iterations, and retrieval time all decreased. Supply lead-time recovery also improved. The chart helps readers see that the framework produced gains across both loops rather than only in documentation.

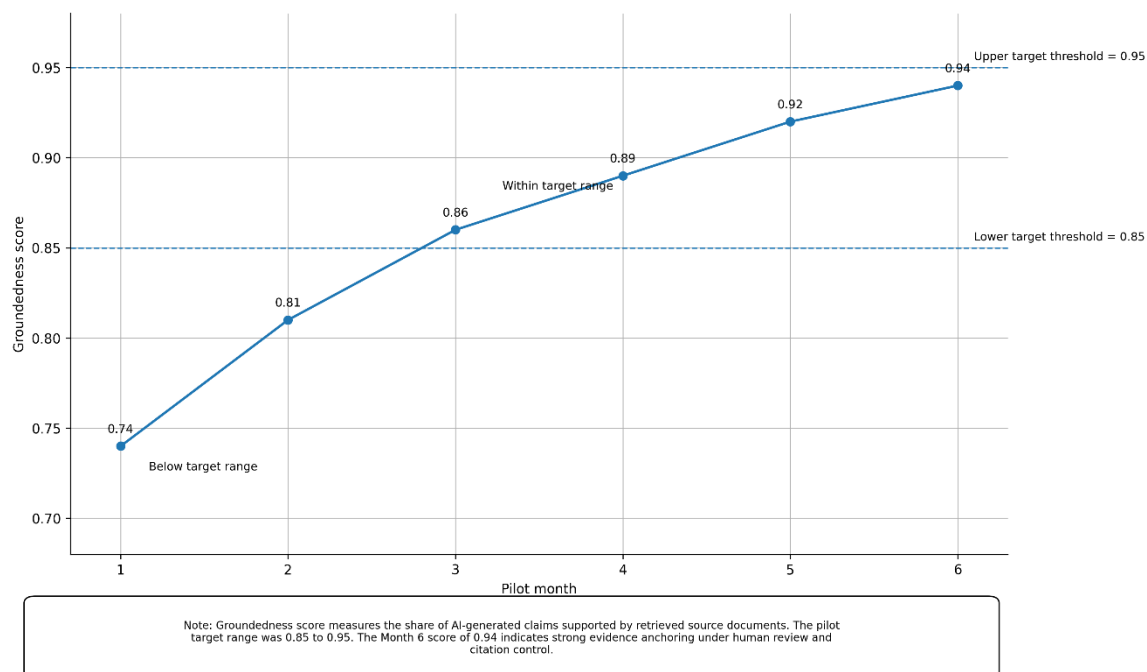


Figure 3. AI groundedness trend during the six-month pilot

Figure 3 explains why governance was central to the model. Groundedness measures the share of AI-generated claims supported by retrieved source documents. A score of 0.94 by Month 6 indicates that the system was strongly tied to controlled evidence. This result supports the argument that AI in regulated manufacturing must be measured for factual anchoring, not only output speed.

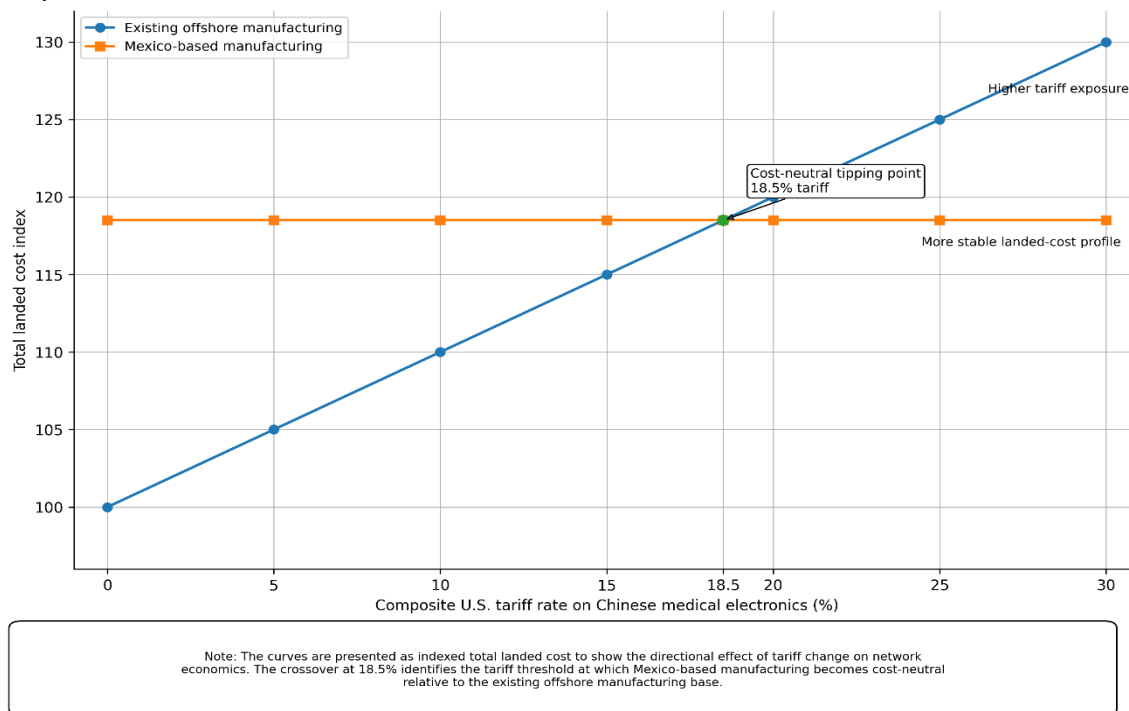


Figure 4. Total landed cost tipping point for Mexico-based manufacturing

Figure 4 converts tariff uncertainty into a managerial decision rule. The pilot simulation found that Mexico-based manufacturing becomes cost-neutral when the composite tariff rate reaches 18.5%. This does not automatically require relocation. It gives leadership a defined threshold for evaluating near-shoring. The figure also shows how AI-supported scenario analysis can move supply chain strategy from opinion to quantified decision support.

5.3 Proposed Improvements

The case results suggest five improvements for future deployment.

First, data governance should be treated as ongoing operating expense. The pilot found that supplier-name inconsistency and duplicate vendor records could distort risk calculations. A one-time data cleanup is not enough. Data quality declines unless ownership and funding remain permanent.

Second, the R-Loop should expand from premarket preparation into post-market surveillance. The same retrieval and evidence-mapping logic can support complaint trending, CAPA linkage, periodic safety updates, field-action readiness, and change impact review.

Third, the S-Loop should strengthen external risk indicators. Supplier scores should include delivery performance, quality history, financial health, logistics signals, cyber disruption alerts, component end-of-life notices, and regulatory events. The score should remain explainable, so planners can understand why action is needed.

Fourth, human review discipline should remain mandatory. The pilot reduced manual search and drafting burden, but it also introduced automation complacency risk. Citation forcing, side-by-side source review, disagreement tagging, and periodic gold-set audits should remain part of the operating model.

Fifth, the model should scale by governance maturity rather than tool count. Adding more products or regions should not mean adding disconnected AI tools. It should mean expanding reusable prompt libraries, approved source repositories, ownership rules, and regional review controls. Table 4 summarizes these improvements and the operating reason behind each one.

Table 4. Proposed improvements for broader deployment

Improvement	Operating Rationale	Expected Benefit
Treat data governance as ongoing OpEx	Supplier and document data degrade over time	More reliable AI outputs
Extend R-Loop into post-market surveillance	Evidence needs continue after clearance	Better lifecycle compliance
Strengthen S-Loop external indicators	Delivery data alone misses silent risks	Earlier disruption detection
Maintain human review controls	AI confidence can create complacency	Lower audit and submission risk
Scale through governance maturity	Tool sprawl creates new silos	Cleaner enterprise adoption

5.4 Validation

Validation used four mechanisms: threshold testing, baseline comparison, governance validation, and literature alignment.

The H1 threshold was satisfied. The target was at least a 25% reduction in regulatory drafting cycle time. The pilot achieved a 32.3% reduction. The quality condition was also satisfied because defect density declined from 2.8 to 0.9 errors per page. A speed-only result would not be acceptable in regulated documentation. The combination of faster drafting and lower defect density supports the R-Loop contribution.

The H2 threshold was also supported. Lead-time recovery improved by 38.8%, and critical stock-outs decreased from two events to zero. Sole-source exposure also decreased by 8 percentage points. These results show that the S-Loop improved resilience through earlier detection and action.

The H3 threshold was supported through ROI analysis. Gross value generated was estimated at \$1.55 million. Pilot investment was \$0.85 million. Net value was therefore \$0.70 million. ROI was calculated as:

$$\text{ROI} = (\text{Gross Value} - \text{Pilot Investment}) / \text{Pilot Investment} \times 100$$

$$\text{ROI} = (\$1.55 \text{ million} - \$0.85 \text{ million}) / \$0.85 \text{ million} \times 100 = 82.4\%$$

Governance validation focused on groundedness and hallucination prevention. The model used source-linked retrieval, red-team testing, low-temperature settings, negative constraints, citation forcing, and human approval. By Month 6, the model

reached a median groundedness score of 0.94. This supports the design principle that AI in regulated manufacturing should be constrained by evidence and review gates.

External validation comes from alignment with existing literature. The findings are consistent with the National Academies of Sciences, Engineering, and Medicine (2022), which emphasizes coordination of supply and demand as a resilience mechanism. They also align with NIST (2023), which emphasizes AI risk governance, and with IMDRF (2022), which emphasizes lifecycle control for machine learning-enabled systems. The results further support McKinsey & Company (2024) and KPMG (2024), which argue that AI value depends on data quality, governance, and operating-model integration.

6. Conclusion

This paper developed and evaluated a dual-loop AI operating discipline for Class II medical device operations. The study addressed a practical issue faced by small and mid-sized regulated manufacturers. Regulatory complexity, supplier risk, quality documentation, and market-access pressure often increase faster than available compliance headcount. Rather than proposing a large enterprise transformation, the study tested whether a five-FTE AI Business Unit could improve performance through a governed process-control layer.

The first objective was met through the design of the dual-loop model. The R-Loop supported evidence retrieval, predicate comparison, technical documentation, and GSPR mapping. The S-Loop supported supplier early warning, alternate-source readiness, and total landed cost analysis. Both loops were connected through human review and formal quality-system records.

The second objective was met through regulatory performance gains. Drafting cycle time decreased from 6.5 to 4.4 weeks, exceeding the 25% target. Defect density, review iterations, and evidence retrieval time also improved. These findings suggest that AI's strongest near-term value is controlled evidence organization and review acceleration, not autonomous authorship.

The third objective was met through supply chain improvement. Lead-time recovery improved from 8.5 to 5.2 weeks, sole-source exposure declined from 42% to 34% of the bill of materials, and critical stock-outs were avoided during the pilot window. The S-Loop also produced a useful strategic insight by identifying an 18.5% tariff threshold for Mexico-based manufacturing cost neutrality.

The fourth objective was met through auditability and financial validation. The model used human-in-the-loop review, groundedness scoring, citation forcing, and red-team testing to control AI risk. It also generated a positive six-month ROI of 82.4%.

The unique contribution of this paper is its practical operating model. It shows how a smaller Class II medical device firm can use AI without weakening quality controls or replacing professional judgment. The findings are bounded by the single-company case design, the six-month pilot window, and the Class II diabetes device scope. Future research should test the model across additional firms, product categories, post-market workflows, and regional regulatory settings.

References

- Ahmadi, S., AI-driven strategies for enhancing medical device regulatory compliance, *RF Quarterly*, vol. 4, no. 3, pp. 24-30, <https://www.raps.org/News-and-Articles/News-Articles/2024/9/AI-driven-strategies-for-enhancing-medical-device>, 2024.
- Amaral, C., Paiva, M., Rodrigues, A. R., Veiga, F. and Bell, V., Global regulatory challenges for medical devices: impact on innovation and market access, *Applied Sciences*, vol. 14, no. 20, article 9304, <https://doi.org/10.3390/app14209304>, 2024.
- Chatterjee, S. and Chaudhuri, R., Supply chain sustainability during turbulent environment: examining the role of firm capabilities and government regulation, *Operations Management Research*, vol. 15, no. 3, pp. 1081-1095, <https://doi.org/10.1007/s12063-021-00203-1>, 2022.
- Christopher, M. and Peck, H., Building the resilient supply chain, *The International Journal of Logistics Management*, vol. 15, no. 2, pp. 1-13, 2004.
- Davenport, T. H. and Prusak, L., *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA, 1998.
- Deloitte, 2023 global life sciences outlook, Deloitte Insights, Available: <https://www.deloitte.com/global/en/industries/life-sciences-health-care/perspectives/global-life-sciences-sector-outlook.html>, 2023.
- Eisenhardt, K. M., Building theories from case study research, *Academy of Management Review*, vol. 14, no. 4, pp. 532-550, 1989.

- European Parliament and Council, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, Official Journal of the European Union, vol. L117, pp. 1-175, 2017.
- International Medical Device Regulators Forum, Machine learning-enabled medical devices: key terms and definitions, IMDRF/AIMD WG/N67, Edition 1, Available: <https://www.imdrf.org/documents/machine-learning-enabled-medical-devices-key-terms-and-definitions>, 2022.
- International Organization for Standardization, ISO 13485:2016 Medical devices: quality management systems: requirements for regulatory purposes, ISO, Geneva, Switzerland, 2016.
- KPMG, Supply chain trends 2024: the digital shake-up, KPMG International, Available: <https://kpmg.com/xx/en/our-insights/ai-and-technology/supply-chain-trends-2024.html>, 2024.
- McKinsey & Company, Generative AI in the pharmaceutical industry: moving from hype to reality, McKinsey & Company, Available: <https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality>, 2024.
- National Academies of Sciences, Engineering, and Medicine, Building Resilience into the Nation's Medical Product Supply Chains, The National Academies Press, Washington, DC, <https://doi.org/10.17226/26420>, 2022.
- National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, U.S. Department of Commerce, <https://doi.org/10.6028/NIST.AI.100-1>, 2023.
- Nonaka, I. and Takeuchi, H., The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, New York, NY, 1995.
- Pettit, T. J., Fiksel, J. and Croxton, K. L., Ensuring supply chain resilience: development of a conceptual framework, Journal of Business Logistics, vol. 31, no. 1, pp. 1-21, <https://doi.org/10.1002/j.2158-1592.2010.tb00125.x>, 2010.
- U.S. Food and Drug Administration, Medical devices; Quality System Regulation amendments, Federal Register, vol. 89, no. 21, pp. 7496-7525, Docket No. FDA-2021-N-0507, 2024.
- Yin, R. K., Case Study Research and Applications: Design and Methods, 6th ed., SAGE Publications, Thousand Oaks, CA, 2018.

Biography

Kenneth Xu is a PhD researcher in Doctorate Pan-European Studies at the University of Pula. His research focuses on AI-enabled regulatory operations, medical device supply chain resilience, and quality-system transformation in regulated manufacturing. He has professional experience in the medical device industry, including regulatory documentation, product lifecycle support, global submissions, technical communication, and cross-functional operational execution. His doctoral research examines how small and mid-sized Class II medical device manufacturers can use governed AI tools to improve regulatory throughput, supplier-risk visibility, and operational resilience without weakening auditability or human accountability.