Constituted from the Foundation

Systems Fail When You Don't Listen

Jonathan Douthit
B.S. Mechanical Engineering, Minor in Robotics • MBA Candidate
Capital Project Engineer & Production Supervisor, Armstrong World Industries
Co-Founder, Seraphim AI

Co-authored with Claude (Anthropic) • March 2026

I am not a machine learning researcher. I am a mechanical engineer, an operations leader, and a builder. For five years I have designed, installed, and maintained critical infrastructure in manufacturing plants — systems where failure means an entire facility goes dark. I am also co-founding Seraphim AI, an agentic platform built on biblical principles for faith-based and underserved communities. These worlds may appear distant from the work of Anthropic's Societal Impacts team, but the distance is shorter than it looks.

The Societal Impacts team studies how AI systems behave when they meet the unpredictable realities of human use. They build tools like Clio to surface patterns that top-down evaluation cannot anticipate. They ask whether Claude's constitutional values hold under adversarial conditions. They measure the gap between what a system is designed to do and what actually happens when people interact with it.

I have spent my career navigating that same gap — between how systems are designed and how people interact with them physically, financially, and relationally. This paper presents two infrastructure case studies and one AI project that converge on a single thesis:

> *Systems fail when they are designed without the people closest to the work. This is true whether the system is a 30,000-pound gearbox, a compressed air network, or an AI agent.*

This principle is not original to me. It is the foundation of the participatory design tradition, from the Scandinavian labor-union technology projects of the 1970s to Toyota's Genchi Genbutsu — "go and see for yourself." It is the lesson of the NUMMI plant in Fremont, California, where the worst factory in America became the best in one year using the same workers and the same equipment, simply because management began listening to the floor. And it is the animating conviction behind Seraphim AI, where constitutional design begins not with the preferences of engineers, but with the needs of the communities the system is meant to serve.

I. When Systems Fail: The Florida Gearbox Project

In 2023, I was assigned to lead a critical infrastructure replacement at an Armstrong World Industries ceiling tile plant in Florida. The project involved replacing the gearbox and drive system on an 8-deck, 500-foot dryer — the first stage in the plant's manufacturing process. Every product line in the facility depended on this dryer. If it failed, the entire plant was dead. Four fabrication lines, hundreds of workers, and a full order book — all bottlenecked through a single piece of equipment.

I was the third engineer assigned. Two had already failed.

The Failure Mode Was Human, Not Technical

The first engineer — I will call him Steve — ran the original design. Steve was technically competent but operated in a purely top-down mode. He did not solicit input from the operators who ran the dryer daily or the mechanics who would maintain the new system. In one incident that crystallized the problem, he told an operator directly: *"Your opinion doesn't matter. I'm the engineer on the job."*

The resulting design reflected this philosophy. It was technically specified but had never been validated by the operators who ran the dryer or the mechanics who would service the new system. The consequences of that omission would not become fully visible until after installation — but they were built into the design from the start.

The second engineer was a younger colleague who was only on the project for a few months before finding a better opportunity closer to his family. No fault of his own — but his departure created a continuity gap that compounded the design problems already in place.

The participatory design literature has a name for what went wrong here. The Scandinavian tradition calls it the failure to integrate the "epistemology of the floor" — the tacit knowledge held by workers whose daily proximity to the system gives them information that no engineering model can fully capture. Toyota's production philosophy formalizes this as *Genchi Genbutsu*: go to the actual place, observe the actual work, build consensus with the people who do it. At the NUMMI plant in Fremont, California, Toyota demonstrated that the same workforce GM had written off as the worst in America could produce the highest-quality vehicles in the GM system — simply by giving them the authority and the obligation to stop the line when something was wrong. The problem at NUMMI was never the workers. It was a management philosophy that refused to listen. The problem in Florida was the same.

Rebuilding from the Ground Up

I arrived in August with a hard deadline: the plant ran a 10-day-on, 4-day-off continuous production schedule, and the only feasible installation window was a 10-day outage between Christmas and New Year's. If we missed it, the entire plant would stay down until a resolution was found. There was no fallback.

Starting from scratch at a plant I had never worked at, I ran a competitive bid process across four contractor categories: mechanical (rigging, welding, installation), roofing, laser alignment, and electrical. The mechanical contractor I selected was a newer company — one that had been, as I described it, "faithful in little things." They had proven themselves on routine maintenance and smaller engineering jobs in the $1,000 to $100,000 range, and the plant's engineering team felt confident they were ready for capital-tier work. I also pulled in an electrical engineer, a civil engineer who could anticipate foundation and structural issues with the crane placement, and a designer for AutoCAD drawings of the monorail system. In total, four crews of 20 to 30 workers.

The installation itself required a 400-ton crane — so large it had to be assembled on site using a smaller crane. The old gearbox weighed 30,000 pounds; the new system, 12,000 pounds. Because there was no way to maneuver the equipment through the building, we removed a section of the roof and installed the gearbox from above.

The Closest Call — And What It Revealed

During drive tuning — disconnecting the motor from the gearbox so it could spin freely and be optimized for electrical savings — a seal burst and oil flooded the area. We lost half a day. Fortunately, we had two spare gearbox assemblies on site. Having spares for critical infrastructure was standard practice at Armstrong, a discipline rooted in the same principle I learned during an internship at Johnson Controls: always design for more capacity than required, because redundancy is the difference between a recoverable setback and a catastrophic failure.

We swapped the assembly immediately. I called the mechanical contractor: *"Can you get me another crew for tonight?"* They mobilized six additional workers for an emergency night shift. I slept four hours, came back to supervise, bought pizza for the crew, and we finished on schedule. The project was delivered on time, on budget, within the 10-day outage window. The plant's critical operations continued without interruption.

But the seal failure was not an isolated incident — it was a symptom. The tight timeline had demanded that I install Steve's original design largely as specified; there was no time for a full redesign before the outage window. Once the system was operational and I could observe it under real conditions, the deeper engineering flaws surfaced: chronic oil leaks, excessive vibration, temperatures exceeding safe thresholds, and components positioned where they could not be practically serviced or replaced. The design had been technically specified but never validated by the people who would maintain it.

The Redesign: Empirical Data over Engineering Ego

With the plant running and the immediate crisis resolved, I turned to fixing what Steve's top-down process had gotten wrong. I gathered empirical data — vibration readings, temperature measurements — and walked the plant floor to find analogous systems that were already working well in other parts of the facility. I brought this data and these examples to a collaborative redesign effort involving senior mechanical engineers at Armstrong, the plant's maintenance department, and the vendor. The resulting design was not my design. It was a design constituted through the input of everyone who would build, maintain, and operate the system. The flaws that had been invisible to an engineer who refused to listen were obvious to the mechanics and operators who lived with the equipment every day.

II. When Systems Are Built Right: The Pennsylvania Compressed Air Overhaul

If the Florida project illustrates what happens when systems are designed without the people closest to the work, the Pennsylvania project illustrates what happens when they are designed *with* them.

Compressed air is a manufacturing plant's third utility, after electricity and gas. It drives valves, labeling machines, conveyors — nearly every process in the facility. At Armstrong's Pennsylvania plant, the compressed air system consisted of two separate networks with six aging compressors, most over 30 years old. One had already caught fire and been dead for a year; replacement parts were no longer manufactured and could only be sourced on eBay from other plants that were tearing down. The system was failing, and when it failed completely, the plant would stop.

I designed the replacement: a consolidated system with one 600-horsepower main compressor, one 300-horsepower trim compressor, and one 300-horsepower unit for redundancy — approximately 1,200 horsepower total, unified into a single network with modern controls. But the design that made it into the

plant was not the design I would have produced alone.

Charles: The Epistemology of the Floor in Practice

Charles was a maintenance mechanic who had been at the plant for over 30 years. He was the kind of person who held a maintenance department together without anyone noticing — greasing bearings no one else knew how to service, keeping aging equipment alive through institutional knowledge that existed nowhere in writing. I asked his supervisor for two hours of his time, then walked the entire compressed air system with him.

Charles showed me things I would not have found in the engineering drawings. He showed me which pipes were hot lines and which were cold. He explained the cooling system bottlenecks and the piping restrictions that limited airflow. He showed me a compressor that had been dead for a year because a critical component was no longer manufactured. And he showed me the constraint that would have derailed the entire project if I had missed it.

The Bidirectional Filter: A Design Decision Born from Listening

One of the plant's newer production lines — the flex line — used European equipment manufactured in Spain, all specified to run on oil-free air. This created an isolated system: the American compressors, which had trace oil in the lines, could not back-feed into the European equipment. Charles knew this because he had worked with both systems for years. The European and American engineering philosophies differed in ways that created practical incompatibilities invisible to anyone who had not operated them daily.

I brought this constraint to the compressor company we were working with. The solution was a bidirectional filter — a component that allowed oil-free air to flow into the American system (as it always could) while also filtering oil from the American compressors to safely supply the European equipment. This single design decision, born entirely from listening to a 30-year mechanic, eliminated the system isolation and created true network-wide redundancy.

The project was delivered using formal 5-why root cause analysis and structured lessons-learned processes. But the most important analytical tool was a conversation with the person closest to the work.

The Contrast

The difference between the two projects was not technical complexity. Both involved critical infrastructure that the plant depended on. Both required significant capital investment. Both carried serious consequences for failure. The difference was constitutional: the Pennsylvania project was *constituted* on a foundation of ethical project management and meaningful collaboration with the people closest to the work. The Florida project's original design was constituted on the assumption that the engineer's expertise was sufficient and the operator's experience was irrelevant. The results were predictable.

This is the same dynamic the Societal Impacts team studies when they analyze how Claude's values hold up in real-world interactions. A system's constitution — whether it is a set of engineering specifications or a set of alignment principles — is only as robust as the process that produced it. If that process excludes the people the system is meant to serve, the constitution will fail under pressure.

III. Constitutional Design for the Underrepresented: Seraphim AI

Anthropic's research on "algorithmic monoculture" has documented a measurable gap in whose perspectives are represented in frontier AI systems. Evaluations using the OvertonScore metric — which measures how fully a model covers the range of legitimate viewpoints on contentious issues — show that Claude and its competitors currently score between 0.35 and 0.41, covering less than half of the perspectives in public discourse. Esin Durmus's research on cultural representativeness has shown that LLMs carry structural biases toward Western, secular, and urban value systems. The Societal Impacts team treats this as a research problem. I am treating it as a building problem.

Seraphim AI is an agentic platform oriented toward faith-based and underserved communities — populations that are largely absent from the usage data that frontier AI companies currently study. I co-founded it because I saw a gap between how alignment is discussed in the AI research community and how it is experienced by communities for whom values are not abstract principles but the organizing framework of daily life.

Architecture: A Constitutional Hierarchy Grounded in Scripture

Seraphim's architecture is a retrieval-augmented generation (RAG) system built on a ChromaDB vector store, hosted on AWS, and currently using Gemini Flash for cost-optimized inference, with plans to transition to a locally hosted model. The system was built using Claude Code. But the technical stack is less important than the constitutional design.

The training data is layered in a deliberate hierarchy. At the base, I processed all 66 books of the Bible — every one of the 1,189 chapters — through a structured devotional framework asking questions such as: *What does this passage mean for us as humans? What does it say about God? What does it say about our need for a Savior?* I then used Claude Sonnet to compose 10-to-15-page syntheses of each book, incorporating historical context, commentaries, common misuses of the text, and different denominational viewpoints. Individual verses were indexed separately at the highest retrieval weight.

The resulting system enforces a strict constitutional hierarchy:

1. Verses (highest weight) — Nothing in the system can contradict what Scripture explicitly states. This is the immovable foundation.
2. Chapter summaries — Devotional-level interpretation that cannot contradict individual verses but provides contextual understanding.
3. Book syntheses — Historical, theological, and denominational context that cannot override chapter-level or verse-level truth.
4. Base LLM (lowest weight) — The general knowledge of the language model, treated as what Christian theology calls *general revelation*: useful knowledge about the world that is filtered through the *special revelation* of Scripture before reaching the user.

This hierarchy is structurally analogous to Anthropic's own 4-tier priority framework for Claude's constitution, where Safety overrides Ethics, which overrides Anthropic Guidelines, which overrides Helpfulness. Both systems solve the same fundamental problem: when values conflict, there must be a deterministic logic for deciding which value prevails. In Claude's case, safety is the immovable foundation. In Seraphim's case, it is Scripture. The architectural pattern is the same; the normative content differs. I arrived at this design independently, through the practice of building for a specific community, before I encountered the formal literature on Constitutional AI.

An Alignment Test from the Field

Shortly after deploying the initial chatbot, one of the elders at my church conducted what was essentially a red-team evaluation. He was concerned about whether the system would hold firm on biblical truth or bend under the pressure of culturally contested questions. Specifically, he probed Seraphim on the question of homosexuality — whether the Bible defines it as sin.

Seraphim provided a clear answer grounded in Scripture: homosexuality is defined as sin in the Bible. But the system did not stop at the doctrinal claim. It spoke what Christians call "truth in love" — explaining that God meets people where they are, that transformation comes through relationship with Christ rather than as a prerequisite for approaching Him, and that the process of sanctification is one of humility and surrender, not behavioral compliance. The constitutional hierarchy held: the verse-level truth was not overridden by the LLM's general training, nor was it softened to align with secular cultural expectations. The system delivered doctrinally faithful output with pastoral sensitivity — because both qualities were built into the constitution from the foundation.

This is the same question the Societal Impacts team asks about Claude: *Do the model's constitutional values hold under adversarial conditions?* I have now tested this question in my own system, with a real stakeholder, on a real issue where the cultural pressure to soften the answer is significant. The answer is yes — when the constitution is built on firm foundations with clear hierarchical authority, the system holds.

Why This Matters for Societal Impacts Research

The communities Seraphim is designed to serve — faith-based populations, rural congregations, people in the Global South — are not well represented in the datasets that frontier AI companies use to study alignment. I know these communities because I have served them directly: construction work on an agricultural research center in Tanzania during Ramadan alongside Muslim co-workers, food bank and clothing drive ministries in Chicago, community garden work in a food desert in rural Pennsylvania, and ongoing support for local shelters, women's centers, and prison ministries. These are not abstract populations to me. They are people I have worked alongside, and their perspectives on what "aligned" AI looks like may differ substantially from the perspectives currently represented in global opinion datasets.

Seraphim is my attempt to build from the floor up — to constitute an AI system on the values and needs of the community it serves, rather than designing for them from a distance. It is participatory design applied to alignment.

IV. Conclusion: The Builder's Perspective

I am an operator and a builder. I think in systems, measure outcomes, and iterate based on data. I have managed $8.5 million in concurrent capital projects. I have led a production crew to an 18.2% productivity increase and a 27.9% reduction in scrap through bottom-up pattern recognition in operational data. I am co-founding a startup where sociotechnical alignment is a design constraint from the initial project charter, not an afterthought.

I am not the most credentialed applicant for this role. I have not trained in formal research conventions. I have not built evaluation pipelines or worked with tools like Clio. These are real gaps, and I name them honestly.

But I bring something that may be harder to find: five years of ground truth from inside the systems being transformed by AI. The Anthropic Economic Index tracks how AI is used professionally. I have been on the other side of that equation, watching how human workers in manufacturing adopt, resist, reshape, and sometimes break the systems deployed around them. I have navigated the gap between design intent and real-world performance — the same gap this team studies in Claude's behavior. And I have begun building an AI system where the question of whose values are represented is not a research finding but a founding principle.

The lesson of the floor is simple and repeatable: systems constituted without the people closest to the work will fail. Systems constituted *with* them will hold. I have seen this in a 30,000-pound gearbox in Florida, a compressed air network in Pennsylvania, and a chatbot grounded in Scripture. What has happened will happen. The pattern is the same. The question is whether we build on firm foundations or repeat the failures of those who did not listen.

I believe this perspective has value for a team studying how AI systems interact with the complex realities of human society. I am ready to learn the tools, the methods, and the conventions. I am not above 80-hour weeks, swing shifts, or holidays to get where I need to be. I have worked full-time since I was sixteen and paid my way through college without debt. I will not be the most credentialed applicant, but I will be among the most disciplined.

"What has been will be again, what has been done will be done again; there is nothing new under the sun." — Ecclesiastes 1:9

References

1. Bai, Y. et al. "Constitutional AI: Harmlessness from AI Feedback." Anthropic, 2022.

2. Anthropic. "Claude's New Constitution." anthropic.com, January 2026.

3. Anthropic. "Clio: Privacy-Preserving Insights into Real-World AI Use." December 2024.

4. Huang, S. et al. "Values in the Wild: Discovering and Analyzing Values in Real-World Language Model Interactions." Anthropic, April 2025.

5. Anthropic. "The Anthropic Economic Index." February 2025.

6. Durmus, E. et al. "Towards Measuring the Representation of Subjective Global Opinions in Language Models." Anthropic, 2024.

7. Roper, R. et al. "Benchmarking Overton Pluralism in LLMs." arXiv:2512.01351, 2025.

8. Nygaard, K. and Bergo, O.T. The Norwegian Iron and Metal Workers Union (NJMF) Project, 1971–1973.

9. Bødker, S. et al. "The UTOPIA Project: Training, Technology and Product in Quality of Work Perspective." 1981–1986.

10. Toyota Motor Corporation. "Toyota Production System: Vision & Philosophy." global.toyota.com.

11. Adler, P. "The NUMMI Case: Democratic Taylorism and the Transformation of the GM-Fremont Plant." 1992.

12. Sundin, A. and Sjögren, D. "The Details Are Not the Details: They Make the Design. And They Make the Systemic Effects." Helseplattformen case study, 2024.