

Research Article

DMAIC, Augmented: Generative AI Strengthens Lean Six Sigma Execution

Oleg Smirnov* 

TYM Business Consulting, Miami, United States

Abstract

This article examines how generative artificial intelligence (AI) assistants can strengthen Lean Six Sigma (LSS) execution across every phase of the Define-Measure-Analyze-Improve-Control (DMAIC) methodology. The objective is to give LSS practitioners field-tested patterns, quantified cycle-time data, and a governance framework for integrating AI assistants into quality-improvement projects without compromising methodological rigor. The investigation draws on recent industry deployments using Anthropic's Claude assistant and a construct called skills: versioned playbooks that bundle written instructions, scripts, reference data, and validation rules into reusable artifacts. Methods include direct measurement of cycle-time deltas across Define, Measure, and Control tasks; four worked Analyze-phase examples spanning hypothesis testing, Measurement System Analysis, Statistical Process Control, and regression interpretation; and analysis of a multi-skill portfolio deployed at a small aviation services client. Results demonstrate substantial efficiency gains: project-charter drafting time fell from 90 to 120 minutes to roughly 6 minutes; monthly payroll consolidation fell from 90 minutes to under 30 seconds (a 99.4 percent reduction) with the first-pass error rate dropping to zero; cross-system benefits reconciliation fell from 4 hours to under 60 seconds, with monthly completion improving from 5 of 12 cycles to 12 of 12. The aggregate skills portfolio at one client achieved approximately 84 hours of administrative cycle-time savings per year. The article identifies six required governance controls (human review, version control, controlled prompts, sensitive-data boundaries, hallucination mitigation, and auditability) and aligns them with ISO 9001: 2015 and the NIST AI Risk Management Framework. The conclusion is that generative AI does not alter DMAIC methodology; it lowers the cost of executing each phase well, expanding the share of LSS that is practiced in fact rather than only in theory. The findings support Master Black Belts and quality champions in deliberately piloting AI-enabled tooling, prioritizing Control-phase tasks first, and treating skill development as a formal deliverable with version control, ownership, and quality gates.

Keywords

Lean Six Sigma, DMAIC, Generative Artificial Intelligence, Quality Management, Process Improvement, Standard Work, AI Governance, Skills Automation

*Correspondence: Oleg Smirnov (olegsmirnovjr@gmail.com)

Received: 16 May 2026; **Accepted:** 26 May 2026; **Published:** 26 October 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Lean Six Sigma (LSS) practitioners agree that the methodology is sound [1, 2, 7]. The recurring obstacle is the human labor required to operate it: charters drafted late, data collection plans abbreviated under deadline, and control plans documented at closure, then quietly retired [9]. Generative artificial intelligence (AI) does not change the Define, Measure, Analyze, Improve, and Control (DMAIC) process. It changes the cost of executing each phase well [10, 11]. Industry surveys document a rapid expansion of generative AI in enterprise workflows in recent years [6, 14], and Quality 4.0 research has begun mapping the implications for quality management [8, 12, 15]. This article reports field-tested patterns from recent deployments, with measured cycle-time deltas, four worked Analyze-phase examples, and a governance framework drawn from existing quality-management practice. The featured assistant is Anthropic's Claude, and the most consequential capability is its ability to run skills: versioned playbooks that turn ad hoc analyst work into reproducible standard work [13, 16].

Client examples in this article have been anonymized, and operational details have been generalized where appropriate to preserve confidentiality while retaining the logic of improvement.

2. Approach and Data Sources

This article is a practitioner case report rather than a controlled study. The evidence comes from Lean Six Sigma deployments led by the author during [YEAR] and [YEAR] at three organizations: a small aviation services firm, which is the primary site and the source of all recurring-task measurements, with four production skills in continuous monthly use; an administrative-process project at a [TYPE OF ORGANIZATION], which supplied the Define-phase example; and a procurement pilot at a [TYPE OF ORGANIZATION], which supplied the Improve-phase examples. The Analyze-phase examples are drawn from individual projects across these deployments.

Manual baseline cycle times were [recorded by the performing analyst over the cycles immediately preceding each skill deployment / taken from client timesheet records]. Post-deployment runtimes were captured from script execution logs. First-pass error rates were established by reconciling each delivered workbook against source-file totals, which is the same check the skills now perform automatically. Dollar figures for deduction errors are taken directly from the audit reports. Where a single project supplied an observation, such as the FMEA candidate count or the SOP reviewer-edit ratio, the figure is reported as an illustration of the pattern rather than as a population estimate. All figures are rounded, and the observation window for the production skills is the first nine months after deployment.

The assistant used throughout was Anthropic's Claude, operating with file and code execution tools and with skills: self-contained folders that bundle instructions, scripts, and reference files that the assistant loads on demand [16]. The patterns described are not specific to this vendor. Any assistant that can execute code, read and write files, and load a versioned instruction package can implement them. Claude is named because it is the instance that was measured.

3. What an AI Assistant Offers the LSS Practitioner

Three affordances matter for LSS work. First, plain-language reasoning allows a Yellow Belt to describe a process gap and receive a structured problem statement for revision. Second, tool execution lets the assistant open spreadsheets, run Python, query databases, and produce deliverable files; the chat window is the entry point, not the workspace. Third, a skill package bundles a recurring task as a written instruction set, plus scripts, reference data, and validation rules. A Master Black Belt (MBB) authors the skill once; every Belt below can invoke it by name. This is the closest software analog to executable standard work that LSS practitioners have had at scale [3].

4. Applying Generative AI Across the DMAIC Phases

4.1. Define

The Define phase is where projects begin to drift. Claude removes the writing tax around scoping artifacts. In a recent administrative-process project, a 47-minute sponsor interview transcript, three prior scorecards, and a stakeholder email thread were provided as inputs. Claude returned a redline-ready project charter (problem statement, scope, in-scope/out-of-scope, baseline metric, goal statement, team, and constraints) in approximately 6 minutes; the comparable manual draft typically required 90 to 120 minutes of Black Belt time. For Voice of the Customer (VOC) synthesis, a separate batch of 184 open-text survey responses was clustered into eight themes, with verbatim customer quotes preserved for each theme in under 10 minutes. Prior practice required a Green Belt for most of a working day. Scoping decisions still belong to the MBB; the assistant simply shortens the path from raw inputs to a defensible draft.

4.2. Measure

The Measure phase is where the largest, most repeatable

cycle-time gains have been observed. A small aviation services client illustrates the pattern. The monthly payroll review required five separate insurance deduction reports (medical, dental, health savings account, vision, and voluntary life) to be pulled from the payroll system as distinct files with differing headers and footer totals. Manual consolidation averaged 90 minutes per cycle with a measured first-pass error rate of approximately 4 percent when the work was rushed. A purpose-built skill was developed that ingests the five files, identifies each category from file content rather than filename, builds a unified employee-by-category matrix, validates each category sum against the source-file footer, and emits a single workbook with anomalies flagged for human resolution. Cycle time dropped from 90 minutes to under 30 seconds (a 99.4 percent reduction). First-pass error rate fell to zero across observed routine cases during the first nine months of operation because the script refuses to deliver unreconciled output. The annualized administrative effort saved on this single task was approximately 18 hours. The same pattern generalizes: anywhere the Measure phase requires a recurring multi-source data pull with a fixed schema, a skill is a candidate.

4.3. Analyze

The Analyze phase rewards specificity. Four worked examples illustrate where Claude adds leverage and where the Belt's judgment remains decisive.

Hypothesis testing. Given a CSV of cycle-time observations split across two process variants, Claude will identify the appropriate test (two-sample t-test for normal data with equal variance, Welch's t when variances differ, Mann-Whitney U when normality fails), generate the Python or Minitab steps, and report the p-value and confidence interval. In a recent staffing-mix analysis, this collapsed a 45-minute exercise into 8 minutes. The Belt still owns the practical significance call: a statistically significant 4-second mean difference may have no operational meaning.

Measurement System Analysis (MSA). For a Gauge R&R study with three operators and 10 parts measured twice, Claude can compute the ANOVA-method variance components, the percent contribution of repeatability and reproducibility, and the number of distinct categories (ndc). It will produce the standard interpretation rubric (less than 10 percent acceptable, 10 to 30 percent marginal, greater than 30 percent unacceptable). The caution: the assistant cannot tell you whether the measurement procedure is operationally appropriate. That requires a Gemba check.

Statistical Process Control (SPC) anomaly detection. Given a time series of subgroup means, Claude can construct X-bar and R charts, apply the Western Electric or Nelson run rules, and produce an annotated list of out-of-control signals. In one operational data set of 412 observations, this surfaced two assignable-cause shifts that the Belt's prior visual inspection had missed. The investigative work of identifying the assignable cause remained a team activity.

Regression interpretation. Claude will fit a multivariate regression and report coefficients, p-values, and R-squared. It will, on occasion, propose causal narratives that do not survive scrutiny. Treat regression outputs as hypotheses, not conclusions, and route them through Five Whys before changing the process. This is the single most common failure mode observed when Belts begin using AI assistants in Analyze.

4.4. Improve

In Improve, the assistant shifts from analyst to drafter. Pilot designs, Failure Mode and Effects Analysis (FMEA) tables, change-management communications, training one-pagers, and Standard Operating Procedures (SOPs) are documents, and documents are where AI offers reliable leverage. Provided with the process map and a scoring rubric, Claude generated an FMEA candidate list of 34 failure modes for a procurement pilot, compared with the 19 the team had enumerated in their initial workshop. Of the 15 additional candidates, the team accepted 9 as legitimate (severity-weighted), rejected 4 as duplicates, and flagged 2 as out of scope. The team's role in debating severity, occurrence, and detection scoring was unchanged; the longer initial list improved the debate.

SOP drafting follows a similar pattern. For the same procurement pilot, the assistant produced a first-draft work instruction (purpose, scope, responsibilities, prerequisites, numbered steps, decision criteria, exception handling, references) in approximately 11 minutes from a tagged process map and three sample transactions. The comparable manual draft typically required 3 to 4 hours of Black Belt time. Across the eight SOPs generated in the pilot, the reviewer-edit ratio averaged 22 percent of words changed in the first revision pass: meaningful, but well below the cost of authoring from scratch.

4.5. Control

The Control phase is where skills earn their place. A second example from the aviation services client demonstrates the pattern. Benefits enrollment (*EmployeeNavigator*) and active payroll deductions (*QuickBooks*) had drifted apart over time, with active enrollments lacking deductions and stale deductions persisting after coverage drops. The monthly cross-system reconciliation required approximately 4 hours and was performed inconsistently, completed in roughly 5 of 12 months in the prior year. A skill was developed to run the audit on demand. It extracts both rosters, reconciles them by name with an alias map for known mismatches, compares enrolled coverage against active deductions, and produces a four-tab report with findings color-coded by severity: Critical for active payment errors, High for likely-stale records or termination conflicts, Medium for amount discrepancies, and OK for systems-in-agreement. Runtime is under 60 seconds. Monthly completion is now 12 of 12 since deployment. The first three audit cycles surfaced \$1,847 in cumulative deduction errors that would have continued indefinitely under the prior cadence.

The MBB recommendation that follows is simple: identify the three or four recurring Control-phase tasks most prone to lapsing in your current deployment and build skills for them first.

5. Skills as Standard Work

The deeper observation is what a skill represents in LSS terms. A skill is standard work, made executable. It is poka-yoke applied to a process step, because the script refuses to

emit a silently incorrect output. It is tribal knowledge captured while it is still available rather than lost when the experienced employee departs. A laminated SOP above a workstation may document standard work; a skill can help ensure consistent execution. Each validated skill becomes a deployment asset that compounds across projects: the current portfolio for the aviation services client comprises four production skills, with aggregate cycle-time savings of approximately 84 administrative hours per year (Table 1) and no observed silent failures over the first nine months of operation.

Table 1. Production skills portfolio at the aviation services client. All four skills run monthly; annualized savings assume twelve cycles per year.

Skill (DMAIC phase)	Manual base-line per cycle	Skill runtime per cycle	Hours saved per year
Payroll deduction consolidation, five carrier reports (Measure)	90 min	Under 0.5 min	18
Benefits enrollment to payroll deduction reconciliation (Control)	240 min	Under 1 min	48
Carrier invoice posting to accounting system, dental plan (Control)	[xx] min	[x] min	[xx]
Carrier invoice posting to accounting system, supplemental coverages (Control)	[xx] min	[x] min	[xx]
Total			84

Skill development itself is not without cost. In the deployments described above, each production skill required an estimated 8 to 16 hours of Master Black Belt time to specify, build, test against representative inputs, and document for handoff. An additional 2 to 4 hours per quarter is typical for maintenance: schema changes, edge-case handling, and minor logic updates as the underlying process evolves. The implication for return-on-investment (ROI) calculations is direct. A skill that saves 90 minutes per month repays its build cost within roughly six to twelve months and produces compounding returns thereafter; skills that target less-frequent tasks should be evaluated against the same threshold before commitment. Practitioners should also factor in the learning curve for prompt and instruction design itself, which is typically 20 to 40 hours of deliberate practice for an experienced Black Belt to reach competence. The pattern is familiar to LSS readers: tool investments compete for the same capacity as project work, and the candidate-skill backlog should be prioritized using the same effort-impact lens applied to any improvement queue.

6. Governance Matters More Than Ever

The benefits described above are real. So are the risks, and quality professionals will rightly insist on governance proportional to them. Six controls have proven necessary in field deployments.

1. Human review of every AI-generated artifact. Charters,

FMEAs, SOPs, and audit reports pass through a named human reviewer before downstream use. The reviewer functions as a QC gate in its own right, not as a substitute for one.

2. Version control for skills. Each skill is stored in a versioned repository with semantic versioning (major.minor.patch), a changelog, a documented owner, and an explicit retirement date.

3. Controlled prompts and instructions. Prompts and skill instructions are treated as controlled documents and validated on the same cadence as SOPs within a quality management system (ISO 9001: 2015, clause 7.5) [4].

4. Sensitive data boundaries. In regulated environments (HIPAA, SOX, GxP, ITAR), data must not leave approved boundaries. Enterprise AI deployments now offer zero-retention modes and on-premises options that should be selected explicitly.

5. Hallucination mitigation. Risk is reduced by retrieval grounding (the assistant operates over named source documents rather than open-ended generation) and by requiring citations on factual claims.

6. Auditability. The prompt, model version, input data hash, and output are captured for any artifact entering a regulated process.

The NIST AI Risk Management Framework (AI RMF 1.0) [5] provides a usable starting structure for organizations that have not yet established an AI governance program.

Adoption barriers are real and largely cultural. Belt skepti-

cism, executive overestimation, and the temptation to skip human review when the assistant appears confident are the three most common failure patterns observed across deployments. Treat them with the same change-management rigor applied to any process change.

7. Deployment Considerations for Champions

Three practical recommendations follow. First, do not roll out AI-enabled tooling across all Belts simultaneously; pilot inside a single project with a Black Belt prepared to learn the tool deliberately. Second, prioritize Control-phase tasks for the first three skills. The instinct is to begin with Define or Measure; in practice, Control is where lapses cost the most and where standardization produces the cleanest return. Third, treat skill development as a formal deliverable. Inventory, version, and assign owners. The MBB who enters the next deployment with a portfolio of validated skills enters with the leverage that the prior generation of practitioners did not have.

8. Limitations and Future Research

Several limitations bound these findings. First, the quantified results come from a single small client and a handful of individual projects. They illustrate a pattern rather than estimate an effect size, and the magnitudes should be expected to vary with process complexity, data quality, and the maturity of the organization's existing documentation. Second, the comparisons are before-and-after measurements with the same staff. No control condition was available, and some of the improvement may reflect the process documentation that skill building itself forces rather than the AI tooling as such. Third, the author designed and built the skills and also measured their performance, so observer bias cannot be excluded; the reconciliation checks and execution logs reduce this risk but do not remove it. Fourth, the first-pass error rate of zero rests on a small number of routine monthly cycles and should be read as no observed failures, not as a demonstrated capability. Fifth, the findings are tied to one assistant at one point in a rapidly moving technology. Model updates can change behavior, which is precisely why the version-control and auditability controls in Section 6 matter.

Future work should replicate the Control-phase pattern across multiple organizations and industries, compare alternative assistants on identical skill specifications, and track maintenance effort over a multi-year horizon so that the return-on-investment estimates in Section 5 can be tested rather than assumed. A controlled comparison in which matched Belts complete the same Define and Improve deliverables with and without an assistant would also establish whether the cycle-time gains persist once the novelty of the tool has faded.

9. Conclusion

DMAIC remains DMAIC. The methodology, the toolset (Pareto, Ishikawa, FMEA, SPC, control plans, Value Stream Mapping), and the discipline they require have not changed. What has changed is the marginal cost of executing each tool well. A capable AI assistant, paired with a portfolio of governed skills, converts the slow and easily skipped portions of LSS into work that is consistently performed and measurably tracked. The opportunity for our community is to apply the same engineering rigor to AI deployment that we apply to every other process change, and to expand the share of the methodology that is practiced in fact rather than only in theory.

Abbreviations

AI	Artificial Intelligence
ANOVA	Analysis of Variance
DMAIC	Define, Measure, Analyze, Improve, Control
FMEA	Failure Mode and Effects Analysis
GxP	Good Practice (Manufacturing, Clinical, Laboratory)
HIPAA	Health Insurance Portability and Accountability Act
ISO	International Organization for Standardization
ITAR	International Traffic in Arms Regulations
LSS	Lean Six Sigma
MBB	Master Black Belt
MSA	Measurement System Analysis
ndc	Number of Distinct Categories
NIST AI	NIST Artificial Intelligence Risk Management Framework
RMF	Project Management Professional
PMP	Quality Control
QC	Return on Investment
ROI	Standard Operating Procedure
SOP	Sarbanes-Oxley Act
SOX	Statistical Process Control
SPC	Voice of the Customer
VOC	

Author Contributions

Oleg Smirnov: Conceptualization, Data curation, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] American Society for Quality. (n.d.). Six Sigma Definition - What is Lean Six Sigma? Retrieved from <https://asq.org/quality-resources/six-sigma>
- [2] George, M. L. (2002). *Lean Six Sigma: Combining Six Sigma Quality with Lean Production Speed*. McGraw-Hill.
- [3] Womack, J. P., & Jones, D. T. (2003). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation* (Rev. ed.). Free Press.
- [4] International Organization for Standardization. (2015). *ISO 9001: 2015, Quality management systems: Requirements*. ISO.
- [5] National Institute of Standards and Technology. (2023). *AI 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>
- [6] McKinsey & Company. (2024). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. McKinsey Global Survey.
- [7] Snee, R. D. (2010). Lean Six Sigma: Getting better all the time. *International Journal of Lean Six Sigma*, 1(1), 9-29. <https://doi.org/10.1108/20401461011033130>
- [8] Antony, J., Sony, M., McDermott, O., Furterer, S., & Pepper, M. (2023). How does performance vary between early and late adopters of Industry 4.0? A qualitative viewpoint. *International Journal of Quality & Reliability Management*, 40(1), 1-24. <https://doi.org/10.1108/IJQRM-05-2021-0134>
- [9] de Mast, J., & Lokkerbol, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics*, 139(2), 604-614. <https://doi.org/10.1016/j.ijpe.2012.05.035>
- [10] Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. <https://doi.org/10.3386/w31161>
- [11] Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. <https://doi.org/10.48550/arXiv.2303.10130>
- [12] Sony, M., Antony, J., & Douglas, J. A. (2020). Essential ingredients for the implementation of Quality 4.0: A narrative review of literature and future directions for research. *The TQM Journal*, 32(4), 779-793. <https://doi.org/10.1108/TQM-12-2019-0275>
- [13] Anthropic. (2024). *Building effective agents*. Anthropic Engineering. Retrieved from <https://www.anthropic.com/research/building-effective-agents>
- [14] World Economic Forum. (2023). *The Future of Jobs Report 2023*. World Economic Forum.
- [15] Antony, J., McDermott, O., & Sony, M. (2022). Quality 4.0 conceptualisation and theoretical framework: A systematic literature review and future research directions. *The TQM Journal*, 34(5), 1169-1188. <https://doi.org/10.1108/TQM-06-2021-0190>
- [16] Anthropic. (2025). *Equipping agents for the real world with Agent Skills*. Anthropic Engineering. Retrieved from <https://www.anthropic.com/engineering/equipping-agents-for-the-real-world-with-agent-skills>

Biography

Oleg Smirnov is Head of U.S. Operations at TYM Business Consulting, a Miami-based consulting and accounting firm serving clients across the United States and Canada. He is a Project Management Professional (PMP) and Lean Six Sigma Master Black Belt with deployment experience across Fortune 500 companies and technology startups. His work includes Kaizen Leader curriculum development at Tesla, process improvement initiatives at PepsiCo and Equinox, and ongoing Lean Six Sigma deployments focused on administrative and operational optimization for small businesses. He holds a BA from West Virginia University and is pursuing an MS in Accounting and Financial Management at the University of Maryland Global Campus. His current research interest is the use of generative AI tooling as executable standard work within quality management systems.