Progression of AI Agents vs Human Experts: Analysis of Pick and Place Material Downtime Data for Proper Root Cause Analysis (RCA) and Guided Actions

Cameron Sobie, Ph.D., Andrew Scheuermann, Ph.D., Tim Burke, Ph.D., Gadi Meik, Shahar Re'em, Cristobal Zatarain

Arch Systems Inc.
CA, USA

ABSTRACT

Material-related downtime on pick-and-place (PnP) equipment is prevalent and costly, yet difficult to diagnose because symptoms span feeders, reels, splicing/presentation, nozzle pick dynamics, component variability, and operator actions. Expertise to correlate these heterogeneous signals is increasingly scarce. This paper evaluates whether large language model (LLM) agents can deliver repeatable, serialized root-cause analysis (RCA) and operator-ready actions when paired with domain data structure.

We present a practical data pipeline that transforms raw PnP event streams into expert-processed evidence: time-aligned alarms, reconstructed feeder/reel/part lineage, and contextual metadata. Using this evidence, we compare two prompting strategies: naïve versus expert-guided. We compare three data regimes: raw, processed, and processed+aggregated. Across multiple production days and experimental conditions, naïve prompting on raw streams yields generic advice lacking actionable specificity; expert prompting without processed evidence remains brittle. In contrast, the processed-data + expert-prompt configuration consistently produces precise RCA tied to serialized assets and clear corrective actions, achieving and exceeding human expert performance.

These results indicate a clear path to scalable AI-driven downtime reduction on modern P&P lines.

Keywords: Pick-and-place, material downtime, feeders, reels, downtime reasons, root-cause analysis, large language models, expert prompting, generative AI, maintenance, SMT.

INTRODUCTION

The electronics manufacturing sector faces a growing expertise gap as experienced personnel retire, with recent U.S. estimates projecting a net need for 3.8 million manufacturing employees by 2033 and as many as 1.9 million roles potentially going unfilled without targeted interventions [1]. At the same time, SMT front-end operations have become more complex (high mix, small batches), and material-related downtime on pick-and-place (PnP) equipment remains both frequent and difficult to diagnose because its symptoms span feeders, reels, vision systems, and component quality—often interacting in ways that overwhelm manual analysis.

A major recent goal of AI applied to operations is to be able to root cause and diagnose any downtime that occurs. In the case of SMT, this applies as a goal to being able to automatically label and explain 100% of downtimes that occur. To accomplish this goal requires a detailed and nuanced understanding of all of the types of downtime that can occur and enabling AI to properly address the right data sets for each subset of the problem. This work looks at the issue of diagnosing material downtime on PnP modules in particular. This is a different class of problem from other downtime types which can be simple. In this complex class, the correct identification of downtime reason only emerges after correlating feeder/reel/part genealogy, operator actions, and serialized production context—capabilities supported by modern Pick and Place platforms and feeder management workflows as used in this study [2, 3].

Building on our prior work demonstrating that expert-guided generative AI can match or exceed human diagnostics when reasoning over domain dashboards [4], this paper evaluates AI agents on a harder task: identifying and prioritizing root causes of material-related downtime directly from historical PnP event streams under three data-processing regimes: raw, processed, and processed + aggregated. We also explore two prompting strategies:

naïve vs. expert. We test whether domain expertise must be injected into both the data (structure) and the prompt (thinking) for LLMs to deliver operator-ready actions.

Our contributions are threefold: (i) we quantify the prevalence and composition of material downtime across real Pick and Place machine production days; (ii) we design an evaluation that scores LLM outputs against SME-established ground truth at the feeder/reel/part level with partial credit for near-matches; and (iii) we show that actionable performance requires the combination of expert-processed data and expert prompting. Conversely, we show that naïve prompting fails regardless of data quality, and raw data defeats both strategies, while expert prompting and processed data consistently achieves near Human expert scores. These findings align with broader evidence that LLMs are most effective in manufacturing when paired with domain frameworks and structured data contexts [4-7].

PROBLEM STATEMENT

Material-Related Downtime in SMT Operations

Material-related downtime represents a persistent and poorly characterized challenge in SMT. While industry practitioners recognize material issues as a significant source of production interruptions, the complexity of these failures and the volume of associated data have prevented systematic analysis and solution development. This section establishes the technical foundation of material-related downtime and quantifies its impact on SMT operations through the empirical analysis of production data.

The Hidden Cost of Material-Related Downtime

Material-related downtime in PnP operation can be related to any one of numerous possible causes such as malfunctions related to tape reels, specific parts, machine calibration, or feeder mechanisms. Unlike failures that manifest with clear symptoms such as attrition, material-related inefficiencies are qualitatively understood as a known issue but are challenging to quantify and prioritize against other production issues. The most straightforward failures involve material depletion when tape reels are consumed faster than operators can splice replacements. More complex failures involve vision system errors, feeder misalignment, nozzle calibration drift, or component quality variations that cause repeated pickup failures and eventual production halts. Without the right downtime tracking solution, these individual slowdowns go undetected.

The diagnostic complexity arises from the interconnected nature of these systems and the patterns that emerge across multiple feeders, part numbers, or production programs. These patterns remain invisible when analyzing individual error events but become significant when viewed systematically across larger datasets. A vision system error may indicate nozzle calibration issues, feeder mechanical problems, component quality variations, or tape reel manufacturing defects—requiring both statistical correlation detection and deep domain knowledge to distinguish meaningful signals from operational noise.

To establish the scope of material-related downtime in SMT operations, we analyzed production data from a contract manufacturing site over a one-month period covering multiple production lines. During the month, over 40,000 panels were produced during over 360 line-hours of working time. Our analysis methodology defines material downtime as instances where pick-and-place module cycle times exceeding the line's takt time coincide with material-related error codes and subsequent feeder or material manipulations by operators. Overlapping module downtimes are removed to ensure accurate downtime characterization for the entire line.

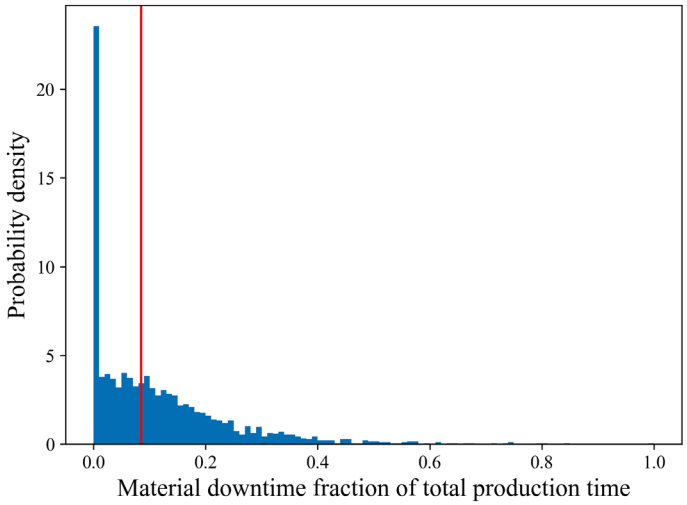

Figure 1: Distribution of material-related downtime fraction for each combination of production day and program. The vertical red line indicates the median material-related downtime fraction of 8.5%.

Figure 1 shows the distribution of daily material downtime percentages per program, calculated as the ratio of material downtime duration to total scheduled production time. The data shows that material downtime consistently accounts for a substantial fraction of operating time with a median of 8.5%, with some production days experiencing downtime fractions exceeding 20%. These numbers represent real money - for large manufacturers operating dozens of SMT lines, material downtime of this magnitude translates into millions of dollars in preventable losses annually.

The large spike at 0% is dominated by short production sessions that do not last long enough to incur material-related downtime, and as such can be considered to be right-censored. While this particular data is derived from specific Fuji PnP machines, the authors have analyzed data across multiple machine vendors and observed comparable rates of material downtime independent of machine vendor.

Next, we investigated the frequency of the machine-reported errors during the downtimes experienced in Figure 1, and show the distribution of attributed failure reasons in Figure 2.

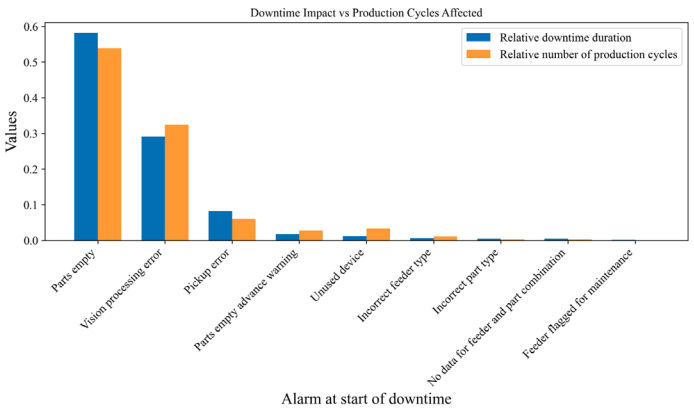

Figure 2: Breakdown of material downtime by reported error type/ cause

An analysis of the underlying causes as reported by the PnP machine is shown in Figure 2 demonstrates the complexity of material-related failures. Material depletion (Parts empty) account for approximately 55% of all material-related downtime, followed by vision system errors with 30% of downtime, followed by pickup errors and the remainder of reasons having minimal impact. However, these categories often interconnect—for example, a mechanically misaligned feeder may generate vision system errors that appear unrelated until analyzed systematically. The material downtime metric reported above was measured during serial production rather than machine setup (see Methodology). The pick-and-place lines analyzed in this study did not have automatic look-ahead feeder switching enabled during the study period, but were enabled with automated feeder/part interlocking and ERP variation.

This quantitative analysis reveals several critical insights. First, material downtime represents a larger fraction of total production time than commonly recognized in industry discussions. In many cases, more than 10% of all production time on the front-end SMT line is spent in downtime caused by material-related issues, representing a substantial area for cost savings. Second, the root causes span multiple subsystems and require correlation analysis to identify systematic issues. Third, the volume of associated data—including error messages, corrective actions, and component genealogy information—exceeds human analytical capabilities during normal production operations.

The absence of systematic solutions for material downtime analysis creates a significant operational challenge. Traditional approaches rely on experienced operators to manually correlate error patterns and identify recurring issues, a process that is becoming increasingly unsustainable. This gap between problem complexity and available analytical tools motivates the development of AI-powered solutions capable of processing large datasets and generating actionable insights for material downtime reduction.

METHODOLOGY

Dataset Description

To prove our approach works with real-world complexity, we analyzed production data from one month of production data from a contract manufacturing facility operating pick-and-place equipment. The dataset encompasses approximately 400k messages spanning production actions, feeder actions, and alarm events per analyzed production day. The messages are highly granular, describing the start and end of production on each module, serialized feeder loading and unloading, reel changes, and alarms of all levels. Creating a dataset appropriate for domain expert analysis requires numerous transformations to establish the feeder state at the time of an alarm, determine which corrective actions were taken to resolve an alarm, and quantify the resulting line downtime. The complexity of this analysis requires both domain knowledge and expert data engineering skills, which places it beyond reach for most production sites.

To evaluate LLM performance across different data processing paradigms, we created three distinct dataset versions representing varying levels of preprocessing and aggregation.

Raw Dataset: The unprocessed production data contains the complete, rich stream of machine messages, error codes, operator actions, and timing information generated by the PnP modules. This dataset includes production start/stop events, alarm notifications, feeder manipulation records, and component supply tracking. The raw format preserves all available information but requires significant transformation and domain expertise to interpret relationships between events and identify meaningful patterns.

Processed Dataset: The intermediate processing level correlates related events into structured production episodes per program, per day. Each row represents a single production event and contains nested structures linking the triggering error conditions, affected equipment (feeders, reels, part numbers), operator corrective actions, and resolution outcomes. This processing eliminates timing ambiguities and establishes clear cause-and-effect relationships while preserving the detailed information needed for root cause analysis. The processed dataset typically contains approximately 150 structured downtime events per production day; nevertheless, additional aggregations would be required for a human expert to be able to propose follow-up actions to reduce future downtime.

Processed and Aggregated Dataset: The highest processing level synthesizes patterns across multiple downtime events to identify systematic issues. This expert-level view aggregates corrective actions by feeder identifiers, reel serial numbers, and part numbers to reveal recurring problem sources. The aggregation focuses on the top 25 contributors to total downtime, ranked by cumulative impact. This dataset format mirrors the analytical approach typically employed by manufacturing engineers conducting systematic troubleshooting investigations.

LLM Analysis Framework

Our experimental framework evaluates LLM performance using two distinct prompting strategies that represent different levels of domain expertise integration. This approach allows us to isolate the impact of expert knowledge versus raw LLM capabilities in manufacturing data analysis.

Expert Prompting Strategy: The expert prompt incorporates comprehensive SMT domain knowledge, including detailed context about pick-and-place operations, common failure modes, diagnostic approaches, and industry best practices for material downtime resolution into an AI SME model. The prompt guides the model through a structured analytical framework that mirrors the thought processes of experienced manufacturing engineers. In the case of raw data, the prompt is additionally orchestrated with an attempt to guide the AI to perform the data engineering steps in addition to the manufacturing domain knowledge to use the resultant data set. This includes specific instructions for correlating error patterns, prioritizing root causes based on downtime impact, and formulating actionable recommendations appropriate for front-line operators. The expert prompts for all levels of data processing end with the instructions for the output data format, which is shared with the naive prompting strategy:

> *Provide the top five concrete actions to take to reduce material downtime from this data and one sentence describing commonalities. Limit your reply to 100 words.*

Naive Prompting Strategy: The naive prompt is as follows:

> *An SMT production line is experiencing excessive material downtime. This dataset describes the production line downtime on an SMT line. It contains the production data, alarms, and the feeder/reel actions of the machine when downtime occurred. Provide the top five concrete actions to take to reduce material downtime from this data and one sentence describing commonalities. Limit your reply to 100 words.*

It presents the data analysis task without domain-specific guidance or structured analytical frameworks telling it precisely how to use the different structures of data. This approach relies primarily on the LLM's pre-trained knowledge and general analytical capabilities, serving as a control condition to evaluate the necessity of expert knowledge integration.

Both naive and expert prompting strategies request identical output formats: identification of the primary systematic issues contributing to material downtime, ranked by impact, and specific corrective action recommendations for each identified issue.

Note on prompt sizes with data payloads

Prompt sizes, in terms of input tokens, are dominated by the data in all three cases. For raw data, the tokens can exceed 2M tokens given the sheer number of production messages produced in typical operation whereas the processed and aggregated datasets require fewer than 10k tokens. It is worth noting that with modern LLM tools the data is not simply ingested, processed by the foundational model, and an output produced, but rather the LLM is intrinsically agentic - it writes and executes processing code to understand attached data and provide insights without specific prompting. Therefore, the input token count serves to characterize the cost rather than the input size for zero-shot processing.

Evaluation Methodology

We employ a multi-faceted evaluation approach combining automated matching algorithms with expert human assessment. Ground truth solutions are established through detailed manual analysis of each production day's data by SMT experts with 20+ years of SMT experience. In contrast to our previous work on attrition root cause analysis, the goal is to identify areas worthy of deeper manual investigation to reduce material downtime rather than prescribing precise follow-up actions.

LLM Root-cause Action Scoring: For quantitative evaluation, we measure precise correspondence between LLM outputs and ground truth solutions with regards to feeder, reel, and part number overlap as the typical resolution action for their associated downtimes.

Partial Match Scoring: Recognizing that manufacturing analysis often involves identifying component combinations rather than individual elements, we implement a partial matching system. When LLM outputs identify one or two of the possible three related elements (feeder/reel/part combinations), they receive 33% or 67% of the available score for that identification.

Each response is given a score from 0-5 by taking the best five of the seven expert recommendations and one point for providing the correct evaluation of systemic issues for a score from 0-6.

Experimental Matrix

Our experimental design implements a systematic evaluation across multiple dimensions to isolate the effects of data processing level and prompting strategy. The complete experimental matrix evaluates six distinct conditions:

- 3 Data Processing Levels × 2 Prompting Strategies = 6 Experimental Conditions
- 3 Independent Production Days to show repeatability
- Total: 18 Individual Experiments

Each experiment follows identical procedures. The designated dataset and prompt combination are submitted to ChatGPT-5 via the standard chat interface, with responses captured for systematic evaluation. This approach ensures consistent experimental conditions while leveraging the most current LLM capabilities available for manufacturing analysis applications.

RESULTS & ANALYSIS

Overall Performance Matrix

The experimental results demonstrate clear performance differentials across data processing levels and expertise provided to the models. Figure 3 presents the comprehensive performance matrix showing LLM accuracy scores across all experimental conditions.

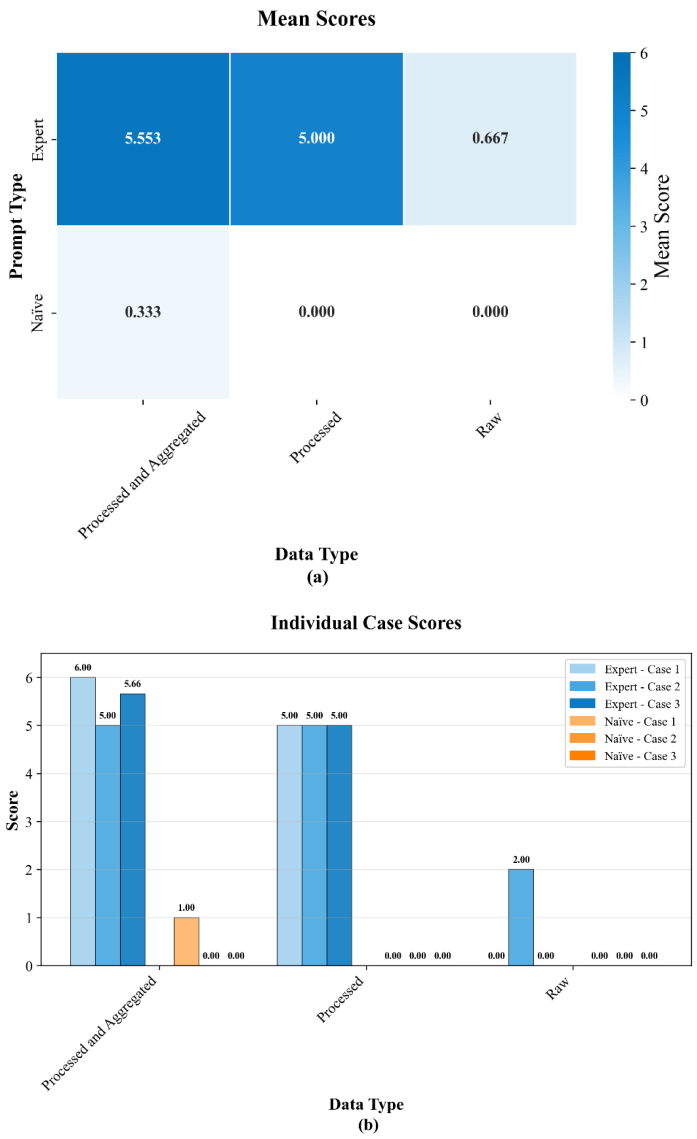

Figure 3: Detailed performance comparison of mean performance (a) and breakdown by individual case (b)

Figure 3 illustrates the performance scores by prompt expertise level and data refinement level. These results demonstrate that in order to adopt generative AI, domain expert prompting combined with expert processing of data is necessary. This isn't just better performance - it's the difference between wasted time and the immediate operational improvements.

Figure 3a clearly shows that both expert prompting as well as expert data analytics are necessary to achieve actionable, expert-level insights for a complex problem such as determining root-causes for PnP material downtime. Naive prompting fails across the board providing generic results with essentially added no value. The importance of expert data synthesis is also clear - the expert prompt described all necessary data transformations to create the Processed and Processed and Aggregated datasets from the Raw dataset, yet the LLM was unable to provide actionable insights. However, the recommendations closely resembled the focused, concrete actions of the other expert prompts as opposed to the generic responses from the naive prompts, which is potentially worse than providing no recommendation at all. We interpret this result as a caution for those seeking to guide generative AI to perform strictly numerical analysis tasks, transforming raw to processed data, vs treating already processed data as human readable input to then generate guidance with.

The minimal performance difference between processed and aggregated datasets when using expert prompting (5.0 vs. 5.6) is an interesting result. This indicates that LLMs with appropriate domain guidance can effectively synthesize patterns from detailed event data without requiring extensive pre-aggregation. In layman terms it could be said that the LLM has a large attention span so it can read over the larger set of processed messages and still distill good guidance. Conversely, naive prompting shows consistently poor performance regardless of data quality, highlighting the critical importance of domain expertise integration.

Discussion comparing the level of data fed to the LLM here vs human readable dashboards

In this study we probe raw, processed and aggregated data highlighting the depth to which current LLMs are able to synthesize from complex datasets with appropriate expert prompting. The transformation from raw PLC bytestream data at the heart of a machine into information ready for human interpretation follows numerous steps, roughly summarized as:

- PLC data
- Raw (smart machine) data
- Processed data
- Aggregated data
- Dashboard

In our prior work on attrition in Pick and Place machines, we showed the effectiveness of expert prompting with dashboards designed for domain experts [4] . We extended this result here by showing that the right analysis with modern AI-based tools have the capability with large context windows and innate agentic behaviour to derive expert-level guidance from data sources substantially closer to the machine. This new result has pragmatic consequences for applying AI solutions to industrial challenges because designing and implementing the combined aggregated data and dashboard is often a labor-intensive task requiring iterations between multiple teams and stakeholders, as well as the skill requirements of data engineers and analysts.

Comparative Analysis: Expert vs Naïve Prompting

To illustrate the practical differences between prompting strategies, we present detailed output comparisons from a representative production day analysis using processed dataset format.

Example Output - Expert Prompting, Processed Data (Score: 5/6):

Downtime is dominated by repeated feeder adjustment issues, especially for certain feeders and reels.

Top 5:

1. Part XXXXX310, feeder KT24L XXX937, reel XXXXX004

(15 cases, 514s).

2. Part XXXXX209, feeder KT8FL XXX075, reel XXXXX021 (14 cases, 437s).

3. Part XXXXX040, feeder KT08C XXX982, reel XXXXX002 (10 cases, 608s).

4. Part XXXXX010, feeder KT08C XXX471, reel XXXXX007 (9 cases).

5. Part XXXXX310 with feeder KT24C (multiple reels, 15+ cases).

Most errors are Parts E or Vision, typically resolved by feeder adjustment, suggesting feeder wear or misalignment.

Example Output - Naive Prompting, Processed Data (Score: 0/6):

1. **Prioritize feeder adjustments** – Most downtime stems from feeder-related errors; improving calibration and maintenance schedules will have the highest impact.

2. **Replace unreliable feeders** – Feeder changes account for \~9% of downtime; identify and swap out repeat offenders.

3. **Address "Parts E" errors** – Standardize handling procedures and root-cause analysis for this error category.

4. **Reduce "Vision" errors** – Calibrate cameras/sensors and retrain operators on alignment practices.

5. **Investigate "unknown" resolutions** – Improve logging and classification of unresolved downtime to ensure future prevention.

The LLM response to the expert prompt provides clear, actionable insights with references to specific serialized factors and correctly identified that most material downtime was resolved not through replenishment but rather feeders were simply removed and replaced with no reel change or part quantity change, indicating likely a physical adjustment to the system. The LLM response to the naive prompt provides broadly valid points, but no clear actions for an operator, nor does it leverage the detailed information available in the expert-processed dataset. Without expert prompting it was not able to learn the latent structure of the data and still utilize it well.

Performance Analysis by Data Processing Level

Raw Dataset Performance: Both expert and naive prompting approaches struggled with raw production data and in most cases yielded no relevant insights and a score of 0. The unprocessed message stream is likely too large for meaningful pattern recognition. Expert prompting showed no improvement over naive approaches and failed to identify specific systematic issues or provide actionable recommendations. The AI model appears to simply read the error codes and made suggestions informed by the foundational model's internal knowledge, rather than interleaving these with the other two datasets to reconstruct the state of the line and identifying downtimes and causes. This is likely a consequence of the sheer amount of raw data which is not appropriate for this level of prompting which neither expert nor naïve prompting can overcome. This demonstrates that even sophisticated domain guidance in the prompt cannot overcome inadequate data preprocessing for complex analytical tasks. Across the three independent production days evaluated (see Experimental Matrix), the raw dataset totals approximately 1.2 million messages (~400,000 messages per production day), while the processed dataset totals approximately 450 structured downtime events (~150 per production day). We note that the bottleneck described here is the LLM's finite context window and attention budget within a single zero-shot query rather than a limitation of "big data" processing capacity in the traditional distributed batch/stream-processing sense.

Processed Dataset Performance: Expert prompting vs naïve prompting achieved substantial performance improvement on processed data, successfully identifying 80-90% of ground truth problem areas and providing specific, actionable recommendations. The structured event correlation enabled effective pattern recognition and root cause prioritization. Naive prompting showed no improvement and continued to generate generic recommendations without specific equipment or part number identification.

Processed and Aggregated Dataset Performance: Expert prompting maintained high performance on aggregated data, demonstrating that LLMs can work equally effectively with expert-level summaries. The slight average performance improvement over processed data may reflect the enhanced clarity of systematic patterns in pre-aggregated format; however, the performance difference between the processed and aggregated datasets is not statistically significant. Naive prompting performance remained limited, suggesting that expert data processing alone cannot substitute for domain expertise in prompting.

Note on growing misunderstanding about error code reading: It is a common misnomer and understandable 'hope' in the industry, that complex manufacturing problems are represented in simple, single error codes found in live machine data. If only it were so, a generative AI could easily find the error code, match it to a manual and perform the root cause and guidance. While it is true that looking up the definition of an error code is simple for generative AI, the problem is that this is essentially never the real problem to be solved. The individual error codes present in the data stream are symptoms of the problem but not causes. There is typically a complex chain of error codes present in the data where the hard problem is correlating different data sets to determine what to look at. In fact, this simple error code lookup is a primary reason for the lack of proper results both prompting approach have on the raw data.

Cost-Benefit Considerations: Raw-data prompts required upward of 2 million input tokens, while processed and processed+aggregated prompts required fewer than 10,000 tokens - a token-cost differential exceeding 200x. Combined with the performance differential above (raw data scoring at or near 0 in nearly all conditions vs. near-expert scores for processed/aggregated data under expert prompting), this represents a strongly unfavorable cost-benefit tradeoff for raw-data approaches, independent of the specific dollar-per-token rate in effect at any given time. The LLM evaluation here strictly uses the foundational models as opposed to commercial agentic tooling, so

the input and output token counts reported here closely track the actual cost incurred.

Critical Success Factors Analysis

The experimental results highlight two essential requirements for effective LLM-based manufacturing analysis. First, the synergistic relationship between data structure and domain expertise is the determining success factor for actionable AI-based insights for industrial challenges. Second, the combinations of expert prompting with either processed or aggregated datasets achieved scores at near-expert levels, representing practical deployment readiness. These conditions successfully identify specific problem areas and generate actionable corrective recommendations. On the other hand, lacking expert data analytics or expert prompting produces generic results with no added value. This reinforces the results of our previous work, in which an AI SME model has two key advantages over an off-the-shelf LLM model - the "thinking advantage" of an expert's thought process, and a "data advantage", having access to expert-synthesized datasets.

Data Structure Requirement: The performance differential between raw and processed datasets demonstrates that appropriate domain-specific data preprocessing is critical to enable LLM effectiveness on industrial problems. Raw manufacturing data contains too much operational noise and lacks the event correlation structure necessary for pattern recognition. However, the minimal performance difference between processed and aggregated formats indicates that extensive pre-aggregation is not required when expert prompting is employed. This is a substantial gain in terms of data processing efforts - slicing and aggregating a complex manufacturing dataset into a human-readable form can require substantial effort, one that we show can be foregone when using an expert-prompted LLM.

Domain Expertise Requirement: The consistent performance gap between expert and naive prompting across all data processing levels reveals that domain knowledge integration into the prompt is equally as critical as its integration in data analysis and preprocessing. Expert prompting transforms general LLM capabilities into manufacturing-specific analytical tools capable of identifying systematic issues and formulating appropriate corrective actions. Naive prompting failed to achieve useful performance regardless of data quality.

Combined Effect: The interaction between data structure and domain expertise proves multiplicative rather than additive. Expert prompting with appropriately structured data achieves performance scores at or exceeding 5.0, while any combination involving either raw data or naive prompting fails to go beyond 0 in all but one case. This demonstrates that both the "data advantage" and "thinking advantage" are necessary conditions for successful AI-powered manufacturing analysis.

These findings establish clear implementation guidelines for manufacturers seeking to deploy LLM-based material downtime analysis. Success requires both sophisticated data preprocessing capabilities and deep integration of manufacturing domain expertise into AI prompting strategies.

Limitations and Future Work: This study has several limitations that motivate future research. Feeder maintenance logs and cycle-count records were not available for the production days analyzed in this study; correlating downtime patterns with feeder maintenance history is a natural and valuable extension of this framework. Second, comparing material downtime patterns on lines with look-ahead feeder switching enabled is a valuable direction for future research.

CONCLUSION

This research demonstrates that manufacturers can identify and provide clear resolution actions to operators for complex material downtime challenges today using expert-guided AI analysis . Our approach works on real production data and can generate immediate ROI through reduced downtime and improved line efficiency, whereas generic AI approaches failed to deliver. Extending our previous work on AI dashboard interpretation and root-cause analysis, we have demonstrated that LLMs can effectively analyze complex manufacturing challenges such as material-related downtime in SMT manufacturing. Synthesizing the highly granular data from the studied Pick and Place machines into appropriately structured data and incorporating expert domain guidance into prompts enabled LLMs to achieve near-expert level performance at the fingertips of anyone in the factory at any time. Expanding on our earlier research showing that AI could achieve expert-level performance in real-time dashboard interpretation for immediate actions, this work addresses a more complex challenge of synthesizing patterns across vast historical datasets to identify systematic material issues—an analytical domain where traditional dashboards do not exist.

Our experimental results demonstrate that the combination of data preprocessing and expert prompting enables LLMs to achieve near-expert performance in identifying systematic material issues and generating actionable corrective recommendations. This extends the proven "data advantage" and "thinking advantage" framework from live dashboard analysis to historical pattern recognition, establishing AI capabilities in areas of SMT manufacturing that have remained beyond the reach of conventional approaches.

The critical finding of this work is that effective AI-powered manufacturing analysis requires both a "data advantage" and a "thinking advantage" operating in concert, consistent with our previous dashboard reading research. Neither sophisticated data preprocessing nor expert prompting alone can achieve acceptable performance—the combination proves multiplicative rather than additive in effect. This principle, now validated across both real-time dashboard analysis and historical data synthesis, has significant implications for AI implementation strategies in manufacturing environments.

Expanding AI Capabilities Beyond Dashboard Analysis: This work demonstrates how the proven framework from our AI dashboard reading research can be extended to manufacturing challenges that

lack existing visualization solutions. Material downtime analysis requires complex data processing to reveal systematic patterns—a task that is presently beyond the scope of both industry experts and traditional dashboard approaches. By applying expert-guided LLM analysis to appropriately structured historical data, electronics manufacturers can now access AI-powered insights in operational domains that were previously beyond the reach of automation.

Comprehensive AI Manufacturing Intelligence: Together with our dashboard reading capabilities, this research establishes a comprehensive approach to AI-powered manufacturing intelligence. Real-time dashboard AI analysis provides immediate insights to operators, reducing their mental load and allowing them to remain focused on line efficiency as well as minimizing the need to call for experts to resolve line issues. Applying AI analysis to more complex challenges such as material downtime enables systematic identification of major recurring issues at a system level beyond the scope of a single line. The combination provides manufacturers with both reactive and proactive AI capabilities, significantly expanding the scope of expert knowledge that can be preserved and scaled through intelligent systems.

Addressing the Knowledge Crisis: This technology offers manufacturers a path to preserve and scale institutional knowledge as experienced personnel retire. Rather than losing decades of troubleshooting expertise, organizations can embed this knowledge into AI systems capable of continuous analysis and recommendation generation. The ability to maintain expert-level analytical capability independent of individual personnel represents a significant operational resilience improvement.

Future Directions: The success of expert-guided LLM analysis in material downtime, combined with our proven dashboard reading capabilities, establishes a foundation for comprehensive AI-powered manufacturing intelligence. Future research can extend this framework beyond the PnP to multi-machine quality process diagnostics, front-to-back SMT manufacturing, and to industry verticals beyond electronics. The demonstrated ability to analyze both real-time dashboard data and complex historical patterns suggests that AI systems can eventually provide complete analytical coverage from immediate problem-solving to strategic operational improvements.

Implications for the SMT Industry: For SMT manufacturers specifically, this work demonstrates that material downtime—traditionally understood as a challenge but prohibitively difficult to analyze and improve —can be systematically analyzed and addressed with a combination of expert data analytics and expert LLM prompting. This research establishes the foundation for a new generation of manufacturing intelligence systems that preserve expert knowledge while scaling analytical capabilities beyond traditional constraints. The demonstrated effectiveness of LLM-based analysis, when properly implemented with expert-level domain understanding, data analytics, and LLM prompting, offers manufacturers a concrete tool for addressing operational challenges at all levels of complexity.

REFERENCES

[1] Deloitte & The Manufacturing Institute. (2024, April 3). US manufacturing could need as many as 3.8 million new employees by 2033 (1.9 million possibly unfilled). https://www2.deloitte.com/… Deloitte

[2] Fuji. (n.d.). NXT Programming Manual (Fujitrax Verifier): Parts-out warnings, dynamic alternate feeders, and feeder replacement. (Manual excerpt). smtbox.com+1

[3] Fuji. (n.d.). NXT2 Quick Reference: Parts supply and alarm pictograms (parts-out, vision, pickup). (Quick-ref excerpt). Scribd

[4] Scheuermann, A., Burke, T., Zatarain, C., Meik, G., Re'em, S., & Sobie, C. (2025). Scaling factory expertise: Comparing AI dashboard vision and human diagnostics on FUJI systems (white paper). Arch Systems. https://archsys.io/hub/papers/benchmarking-human-and-ai-agent-experts-on-manufacturing-dashboard/

[5] CIRP Annals Editorial Board. (2024). Artificial intelligence in manufacturing: State of the art, perspectives, and future directions. CIRP Annals, 73(2), 723–749. https://doi.org/10.1016/j.cirp.2024.04.101 ScienceDirect

[6] Finkel, P., Wurster, P., & Radler, R. (2024). Large language models in production. Industry 4.0 Science, 40(6), 48–55. https://doi.org/10.30844/I4SE.24.6.48 industry-science.com

[7] Chen, Y., Xie, H., Ma, M., Kang, Y., et al. (2024). Automatic root cause analysis via large language models for cloud incidents. In EuroSys '24 Proceedings. https://doi.org/10.1145/3627703.3629553 (abstracted source) ResearchGate

APPENDIX

Appendix A: Raw Data Example

Panel production
Line name: Line 1
Machine ID: XXXX
Lane number: 1
Program name: Program A
Time: 2025-08-07 22:01:34+00:00
Panel number: 1
Action: completed
Cycle time (s): 18
Component loading actions:
Line name: Line 1
Machine ID: XXXX
Time: 2025-08-07 22:01:34+00:00
Action: reel_loaded

Unloaded components: None
Changed components: None
Loaded components: "[{'date_code': '', 'feeder_id': '', 'lighting_class': '', 'lot_number': '', 'new_quantity': 905, 'new_reel_id': 'XXXXX009', 'part_number': 'XXXXX226', 'slot_number': 10, 'stage_number': 1, 'supply_method': {'method': 0, 'method_name': 'feeder'}, 'vendor': ''}]"
Feeder errors:
Line name: Line 1
Machine ID: XXXX
Time: 2025-08-07 22:01:34+00:00
Device state: 15
Device state name: Parts Out Warning
Part number: XXXXX296
Slot number: 11
Stage number: 1

Appendix B: Processed Data Example

The following bullet points correspond to columns in the dataset and are followed by an anonymized example for a single row.
Machine identifier: 111
Start time: 2025-08-07 22:01:34+00:00
Error state during downtime: PartsE
Feeder actions:
"[{""part_number"":""XXXXX160"",""slot_number"":""11"",""stage_number"":""1"",""feeder_id"":""XXXXX312"",""reel_id"":""XXXXX003"",""resolution"":""feeder_changed""},{""part_number"":""XXXXX160"",""slot_number"":""11"",""stage_number"":""1"",""feeder_id"":""XXXXX312"",""reel_id"":""XXXXX003"",""resolution"":""feeder_adjustment""}]"
Feeder action count: 2
Downtime (s): 23

Appendix C: Processed and Aggregated Data Example

Error state: Vision
Resolution: feeder_adjustment
Part number: XXXXX310
Feeder ID: XXXXX937
Reel ID: XXXXX004
Downtime (s): 360
Downtime count: 12
Part total downtime count: 30
Part downtime (s): 1416
Feeder total downtime count: 14
Feeder downtime s: 492
Reel total downtime count: 30
Reel Downtime (s): 1416