FROM DATA TO INSIGHT: A GENERATIVE AI AGENT ACCELERATING DECISION MAKING

Alberto Y. Zaldivar Jr.
Jerald M. Constantino
Jhon Vincent A. Gupo

Factory Analytics and Operations Excellence – Global Operations
Western Digital Corp., 109 Technology Ave., SEPZ, Laguna Technopark, Biñan, Laguna, Philippines 4024
Alberto.Zaldivar@wdc.com, jerald.constantino@wdc.com, jhon.vincent.gupo@wdc.com

ABSTRACT

Data analytics plays a vital role in modern manufacturing, where companies like Western Digital generate thousands of data points daily. However, a significant portion of this data remains underutilized due to manual, time-consuming analysis processes that require expert knowledge to extract actionable insights. While Generative AI platforms like ChatGPT offer strong analytical capabilities, they are not suitable for handling confidential enterprise data.

To address this, a domain-specific Generative AI system was developed and deployed within a secure environment. The system integrates natural language interaction with automated agents for exploratory data analysis, anomaly detection, and insight generation.

The solution shows strong potential to reduce manual effort by up to 80%, consistent with efficiency gains reported in prior studies. It also enhances decision-making speed and expands access to insights across operational teams. Finally, this deployment represents a critical step toward democratizing AI-assisted data exploration, with efforts underway to enable adoption across all departments.

1. 0 INTRODUCTION

In the era of Industry 4.0, data analytics has become a critical enabler in manufacturing, supporting optimization, root cause analysis, and rapid decision-making across the production lifecycle. As processes become increasingly complex and data volumes grow exponentially, the ability to transform raw data into meaningful insights is essential for maintaining competitiveness and operational efficiency. In semiconductor assembly, where precision, yield, and turnaround times are tightly coupled with process stability, timely and accurate interpretation of data holds direct implications on product quality and throughput.

In large-scale manufacturing environments such as Western Digital Corporation (WDC), vast volumes of process and quality data are generated daily from equipment logs, sensor networks, and quality inspections. However, this data remains underutilized due to the complexity of extracting actionable insights. Managers, engineers, and decision-makers are often hindered by the need for specialized statistical expertise, manual analysis efforts, and delays in identifying key anomalies or process deviations. This bottleneck in data-to-decision flow slows down corrective actions, reduces operational agility, and impacts yield and productivity.

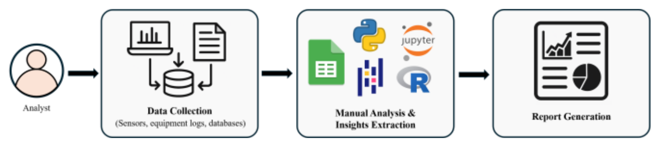


Fig. 1. Manual Data Analysis Workflow. The current workflow involves data collection, manual analysis using spreadsheets and scripts environment, and report generation.

Currently, engineers and analysts are conducting their work through manual processes as depicted in Fig. 1. For instance, analysts often rely on spreadsheet-based tools such as spreadsheet or scripting languages like Python and R to perform data cleansing, compute summary statistics, and generate visualizations. These tasks, though routine, are time-consuming and prone to human error, especially when dealing with large-scale datasets from heterogeneous sources. Anomaly detection typically requires manual threshold-setting or retrospective analysis using static control charts, making real-time responsiveness infeasible. Root cause analysis, when anomalies are detected, often involves iterative hypothesis testing or laborious correlation checks across multiple variables.

This labor-intensive workflow imposes significant delays between data collection and actionable insight generation. Furthermore, it limits the accessibility of advanced analytics to a small group of experts, thereby excluding operators, process engineers, and other stakeholders who could benefit from real-time data intelligence. The lack of standardized automation and intuitive tools inhibits the widespread adoption of data-driven practices, reducing the effectiveness of continuous improvement initiatives.

However, advancements in technology offer promising solutions to these challenges. For example, previous attempts to streamline data analysis have involved Business Intelligence (BI) dashboards[1] and statistical process control systems[2]. While useful, these tools often require domain expertise and manual configuration, limiting their accessibility to non-specialists and reducing their effectiveness in uncovering hidden patterns or real-time anomalies. Recent efforts incorporating machine learning (ML) have shown promise, but many still lack interpretability and ease of use, especially in fast-paced industrial environments[3]. While recent advancements in generative AI offer promising capabilities for automating data analysis, widely known platforms such as ChatGPT, Gemini, and others are not suitable for handling confidential or internal company data due to privacy and security concerns.

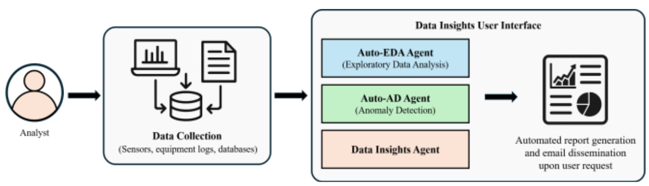


Fig. 2. Automated Data Insights Workflow. The solution streamlines data-to-decision flow via an AI-driven user interface, featuring Auto-EDA, Auto-AD, and Data Insights agents.

Thus, there is a growing need for domain-specific, privately hosted Generative AI (GenAI) systems that can safely operate on internal datasets without exposing them to third-party services. To address these limitations, the proponents introduced Data Insights system, an AI-powered analytical assistant tailored for semiconductor assembly data. As illustrated in Fig. 2, the system combines multi-agents for automated exploratory data analysis (Auto EDA), anomaly detection (Auto AD), and Data Insights with natural language interaction to bridge the gap between raw data and decision-making. By automatically generating descriptive statistics, custom visualizations, trend analyses, and root-cause explanations, the agent empowers users across roles and expertise levels to rapidly derive insights and take informed action.

The remainder of this paper is organized as follows. Section 2 reviews related literature on automated data analysis in manufacturing. Section 3 describes the methodology behind the development of the Data Insights system, including its architecture and core components. Section 4 presents results and discussion, featuring visual outputs and use cases from semiconductor assembly. Section 5 concludes the study by highlighting key findings, while Section 6 offers recommendations for future enhancements and integration into smart manufacturing systems.

2. 0 REVIEW OF RELATED WORK

In recent years, ML-driven analytics and advanced AI models have been leveraged to accelerate data-driven decision making in semiconductor assembly, with studies reporting dramatic improvements in throughput, yield, and efficiency[4, and 5]. For example, Hung et al.[6] developed a three-phase data-science framework for semiconductor assembly that uses LASSO and stepwise regression to select critical process parameters and then apply multiple ML models (neural networks, support vector regression, and gradient boosting) to predict the extent of die delamination (a leading cause of chip failure) from pre-bonding sensor data. Their empirical study on actual fab data showed that this framework provides "effective delamination prediction" and supports rapid troubleshooting (thus shortening cycle times and improving yield). In a related defect-detection application, Shen and Lee[7] use deep learning to improve wafer bin map recognition (WBMR) in assembly. They embed a convolutional network (Inception CNN) in a defect classification pipeline that includes autoencoder–based data augmentation and region classification, enabling the system to distinguish systematic defect patterns (for example, tool or recipe issues) from random noise. By adjusting the anchors with a revised Jaccard index and retraining on augmented samples, the WBMR accuracy was significantly improved, demonstrating how ML can automate pattern recognition in packaging inspection and thereby speed up quality-control decisions. Similarly, Wang and Chiu[8] tackle wire-bonding quality by predicting the "shear force" (bond adhesion) before bonding takes place. They automatically extract six features from probe-mark images and use PCA-transformed features in a random-forest classifier; this model identifies bad chips with 97.92% accuracy before bonding. By catching low-shear-force die in advance, their approach avoids unnecessary bonding steps, saving processing time and cost and hence improving overall packaging throughput. These examples illustrate how ML-based vision and classification systems can make detection faster and more accurate than traditional manual or rule-based inspection.

Beyond defect inspection, ML also enables predictive maintenance and process optimization to speed decisions. For instance, Pradeep et al.[9] apply ML to equipment and wafer-processor sensor data to predict impending wafer or tool failures; their models (including random forests, SVMs, and gradient boosts) achieve high accuracy (>93% for wafer-failure prediction). By forecasting failures, the system recommends maintenance before breakdowns, thereby reducing unplanned downtime and boosting equipment utilization (in one study wafer fab tools ran only ~44% of the time before PdM). In effect, the data-driven model reduces equipment failure by enabling predictive maintenance and increases productivity. A broader survey by Chen et al.[4] confirms that ML models are being deployed in advanced process control (APC), virtual metrology, and in-line quality control to accelerate feedback and improve yield, noting that ML-driven automation is rapidly gaining popularity and advancing at a swift pace in semiconductor manufacturing. Likewise, Yu et al.[10] demonstrate that integrating cross-fab data with cloud-based ML tools can drastically shorten new-product introduction (NPI) cycle time: by automating yield management and test-time analysis, ML analytics helped expedite yield learning, cut test time, and enhance quality for a safe product launch. In their study of lead-frame assembly, higher throughput was achieved by using ML-guided statistical yield tools on hard/soft bin data and first-fail test parameters, which expedited yield learning, test time reduction (TTR) and quality enhancement during production ramp-up. These findings indicate that ML-driven "decision-support" at the system level fusing sensor, test, and yield data can compress analysis time and provide real-time recommendations for process adjustments.

Across these application domains from predictive maintenance and scheduling to defect detection and yield management, researchers report that ML methods consistently enable faster, more accurate decisions. For example, Cao et al.[11] presents a real-time surface-defect detector for chip packages using a YOLOv7 deep network enhanced with attention modules; compared to a baseline network, their ML model runs ~21.6% faster (and with far fewer parameters) while improving detection accuracy by 1.39% on a custom packaging-defect dataset. This sort of speed-up is representative: ML systems eliminate cumbersome hand-tuned feature extraction, automatically optimize across many variables, and operate continuously, so that decisions (for example, flagging a fault or triggering a maintenance alert) happen nearly instantaneously once sufficient data are collected. In each case above, the use of supervised or deep learning on historical assembly data has translated to dramatic gains. The surveyed literature emphasizes that ML-driven models and analytics provide real-time monitoring, predictive maintenance, and adaptive control that minimize downtime and waste while optimizing output. In summary, studies from the past five years show that embedding ML into semiconductor assembly whether in backend packaging lines or test and NPI flows can slash decision latency and error rates. By converting streaming fab data into automated predictions and prescriptions, ML empowers shop-floor operators and engineers to act more quickly and confidently, accelerating semiconductor assembly decision cycles and improving product yield and quality.

3.0 METHODOLOGY

This paper implements a Generative AI-based system that automates statistical insight extraction to address manual analysis limitations in large-scale manufacturing. The system supports natural language queries for performing exploratory analysis, anomaly detection, and experiment design without coding. This section outlines the system architecture, and its core modules (Auto-EDA, Auto-AD, and Data Insights) designed to accelerate decision-making in semiconductor assembly.

3.1 System Overall Architecture

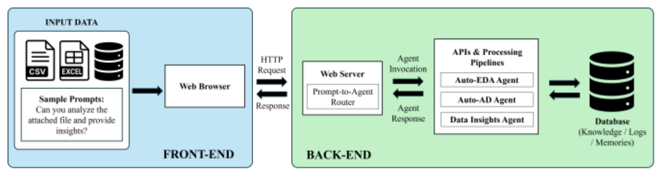


Fig. 3. Architecture of the AI-Assisted Insight Generation System. The architecture illustrates the interaction between the user interface, web server, agent classifier, modular agent pipelines, and the underlying database.

Fig. 3 illustrates the overall system architecture of the proposed AI-assisted insight generation framework. The process begins with user interaction via a web browser, where structured data files (such as CSV or Excel) and a natural language prompt are submitted. These inputs are transmitted to a web server, which functions as the central controller. A Prompt-to-Agent Router within the server interprets the prompt to determine the user's intent and dispatches the request to the appropriate agent pipeline. For instance, a query such as, "Perform data analysis using the attached dataset and provide an insight," would trigger both the Auto-EDA and Data Insights agents.

Table 1. Description of Processing Agents

Agent name	Purpose / Task
Auto-EDA	Performs automated exploratory data analysis, including summary statistics, visualizations, and data profiling.
Auto-AD	Detects anomalies or outliers in the dataset using statistical or machine learning-based methods.
Data Insights	Generates contextual insights or natural language summaries from the data, based on the user's prompt.

As shown in Table 1, available agents include Auto-EDA, Auto-AD, and a general-purpose Data Insights agent. Each agent executes its specific task and may interact with a backend database that stores knowledge logs, memory, or previously generated insights. The results are then returned to the front-end for user interpretation. This modular architecture is extensible, allowing new agents to be integrated with minimal modification to the routing logic. Detailed descriptions of each processing pipeline are provided in the following sections.

3.2 Data Insights Agent Pipeline

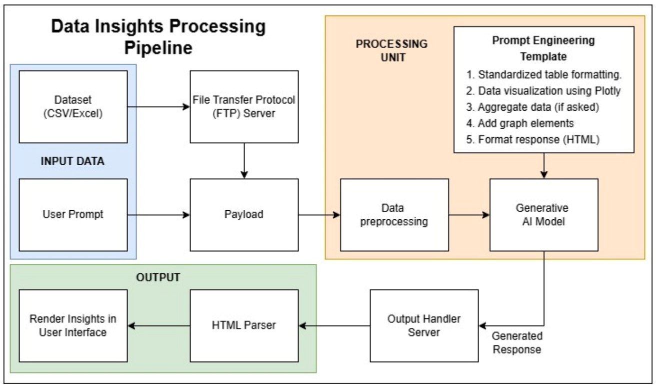


Fig. 4. System Architecture of the Data Insights Processing Pipeline. The architecture illustrates the end-to-end workflow for generating insights from structured tabular data using a GenAI model.

As illustrated in Fig. 4, users submit two types of input: a structured dataset (in CSV or Excel format) and a natural language prompt. The dataset is uploaded and staged via a File Transfer Protocol (FTP) server, while the user prompt is directly passed to the system. These components are combined into a unified API payload for further processing.

Subsequently, the Processing Unit begins with a data preprocessing module that performs tasks such as schema validation, data type checking, formatting standardization, and data cleaning. A prompt engineering template then guides the construction of the final prompt issued to the GenAI model. The template includes directives for standardized table formatting, data visualization using Plotly, optional data aggregation, adding graph elements (such as legends, axes, and axis labels) and HTML response formatting.

Finally, the GenAI model receives the processed input and generates an HTML-formatted response. This response is forwarded to an Output Handler Server, which acts as a temporary storage and transfer layer. The output is then parsed by an HTML parser and rendered in the user interface, allowing users to view interactive visualizations and textual summaries.

This modular architecture supports scalable, low-effort insight generation from raw manufacturing data, enabling domain experts and non-experts alike to engage with analytical outputs without requiring programming.

3.3 Automated Exploratory Data Analysis Agent Pipeline

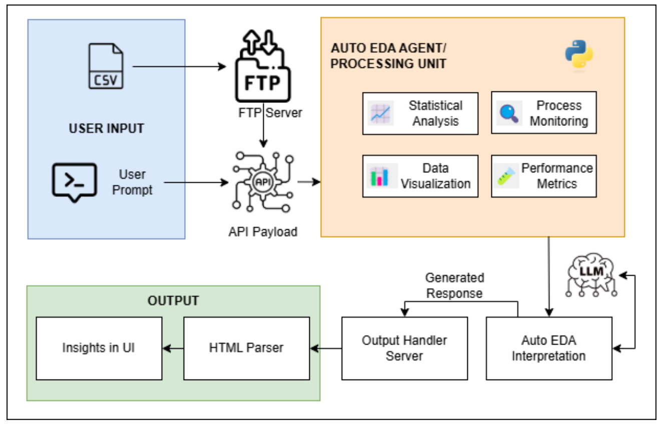


Fig. 5. System Architecture of the Auto Exploratory Data Analysis Pipeline. The architecture illustrates the end-to-end workflow for generating automated Exploratory Data Analysis from structured tabular data while utilizing large language model for generating insights.

As illustrated in Fig. 5, the Auto Exploratory Data Analysis (Auto EDA) module follows a structural framework like the data insight pipeline, consisting of four main sections: User Input, Processing Unit, Generated Response, and Output. The distinction lies in the Processing Unit, which is composed of four key components.

First, Statistical Analysis computes essential statistical measures such as mean, median, standard deviation, and skewness. It also analyzes the distributions of Key Process Output Variables (KPOVs) and Key Process Input Variables (KPIVs), enabling detection of data patterns and emerging trends.

Second, Data Visualization provides insightful graphical representations through histograms, box plots, scatter plots, time series plots, and heatmaps. These visual tools help illustrate data distributions and offer statistical summaries to enhance interpretation. Third, Process Monitoring focuses on evaluating critical manufacturing characteristics. This component assesses process capability and stability over time, aiding in performance tracking and continuous improvement.

Finally, Performance Metrics further evaluate the same manufacturing indicators such as AVT, Ni-Thickness, and

Flatness to quantify overall performance, ensuring reliable benchmarking and supporting long-term optimization efforts.

3.4 Automated Anomaly Detection Agent Pipeline

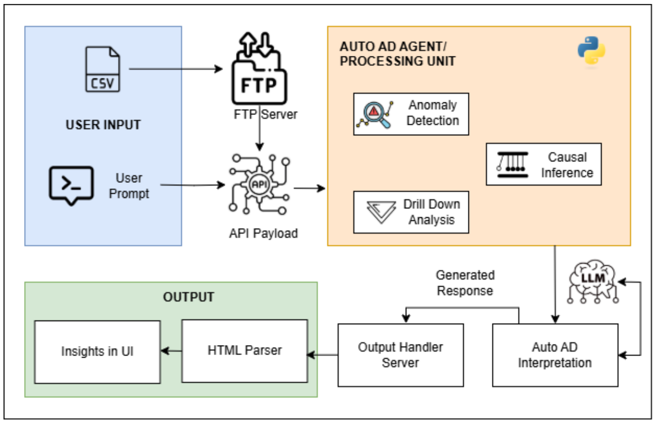


Fig. 6. System Architecture of the Auto Anomaly Detection Pipeline. The architecture illustrates the end-to-end workflow for generating automated Anomaly Detection from structured tabular data while utilizing large language model for generating insights.

Like the Data Insight and Auto EDA modules, the Auto Anomaly Detection (Auto AD) system follows the same pipeline structure comprising User Input, Processing Unit, Generated Response, and Output as shown in Fig. 6. The primary distinction lies in the Processing Unit, which is tailored to support two key components: Anomaly Detection and Causal Inference. The Anomaly Detection component identifies data points that deviate significantly from expected patterns, capturing rare events, data inconsistencies, or hidden trends. To support interpretability, visualizations such as histograms are used to display the distribution of anomaly scores, highlighting outliers and skewness in the data. Box plots are employed to illustrate the range, median, and variability of anomaly scores, distinguishing between normal (0) and anomalous (1) instances. The Causal Inference component aims to determine which features have a genuine causal influence on the occurrence of anomalies, as opposed to mere correlation. This facilitates a deeper understanding of the root causes of anomalous behavior. Visual aids include a Feature Importance Chart generated using an isolation forest model, which ranks numerical features by their impact on anomaly detection. Additionally, a Top 5 Features vs. Index Plot visualizes how the most influential numerical features vary across data records, aiding in the identification of specific anomaly-driving points. For categorical variables, the top five influential features are selected based on statistical association tests such as Theil's U, providing insights into which categories are most strongly linked to anomalous events.

4.0 RESULTS AND DISCUSSION

This section presents the business impact of the proposed system, highlighting efficiency gains achieved through deployment. It also discusses test case results demonstrating the performance of each core agent, including Data Insights, Auto-EDA, and Auto-AD. For testing and public dissemination, a publicly available dataset from Kaggle[12] related to the semiconductor industry was used to demonstrate the system's capabilities.

4.1 Business Impact and Efficiency Gains

The deployment of the GenAI-powered Data Insights system resulted in substantial operational benefits within the WDC analytics environment. One of the most notable improvements was the significant reduction in manual effort required for routine data analysis tasks, including insights generation, anomaly detection, and visualization preparation.

In line with recent findings by Uhunoma[13], which demonstrated that GenAI could reduce work turnaround time by up to 80%, the implemented system achieved comparable efficiency gains. Specifically, time spent on insight generation was reduced compared to the baseline manual workflow, which previously involved labor-intensive scripting, spreadsheet-based processing, and iterative reporting cycles. These efficiency improvements stem from the automation of repetitive tasks and the elimination of manual scripting, data formatting, and revalidation steps that were traditionally required.

Beyond time savings, the system also contributed to improved decision latency, enabling faster detection of anomalies and more timely corrective actions. This enhancement directly supports engineering responsiveness to quality excursions and process drifts, thereby strengthening yield protection.

The adoption of a natural language interface represents a critical step in the ongoing democratization of data analytics, accelerating the transition from manual, expert-driven tasks to accessible, AI-assisted processes. With this capability, non-technical users (including production supervisors and quality engineers) can interact with structured data without requiring programming or statistical expertise. This has led to enhanced cross-functional collaboration and broader data utilization in day-to-day operational decision-making.

Finally, by deploying the system within a private and secure infrastructure, all internal manufacturing data remained confined within organizational boundaries, addressing concerns regarding data confidentiality and regulatory compliance. This secure deployment approach ensures the solution's suitability for continuous use in high-sensitivity domains such as product engineering, reliability monitoring, and process control.

4.2 Data Insights Test Cases Result

The Data Insights Agent was utilized to assess whether wafer processed through Etching equipment demonstrated statistically higher defect rates compared to those processed via Lithography or Deposition. The agent responded by generating comparative visualizations, including a violin plot and a box plot, representing the distribution of defect counts across these process stages.

Fig. 7. Defect Count Distribution by Process Stage. This violin plot visualizes the spread and frequency of defects across process stages, with Etching showing the widest distribution and highest concentration of defects.

The visual evidence in Fig. 7 confirms that wafers subjected to the Etching stage exhibited both a wider spread and a higher concentration of defects compared to those processed in Lithography or Deposition. The agent's descriptive summary highlighted these differences, citing both the median and variability of defect counts.

These findings validate the agent's utility in surfacing actionable insights through natural language interaction, enabling faster diagnostic assessments without manual scripting or statistical programming. The full visual outputs, including the box plot and descriptive statistics by stage, are provided in Appendix A for reference.

4.3 Exploratory Data Analysis Test Cases Result

Using the same wafer dataset provided to the Data Insight Agent, the data was subsequently passed to the EDA Agent for further exploration and insight generation. The agent responded with a comprehensive analysis of the wafer data, ranging from a general overview to detailed interactive visualizations.

The scatter plot presents individual data points as blue dots, each representing a combination of defect count and its corresponding temperature. A red line overlays the plot to indicate a linear regression fit, providing a straight-line approximation of the trend in the data. The associated R^2 value of 0.070 suggests a very weak linear correlation between defect count and temperature_C. Additionally, a blue dashed line represents a quadratic (second-order polynomial) fit, which also yields an R^2 value of 0.070. This indicates that even a more flexible curved model does not capture a strong relationship between the two variables. Overall, the low R^2 values from both fits imply that there is no significant trend or correlation between defect count and temperature in this dataset.

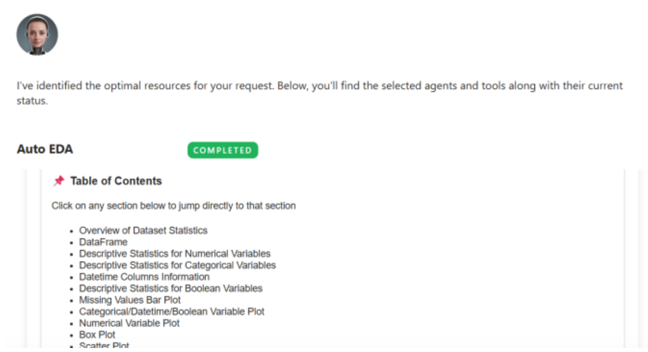


Fig. 8. Table of contents generated by the Auto EDA agent, covering an overview of the dataset up to interactive visualizations such as box plots and scatter plots, where users can select the x- and y-axes of the graphs.

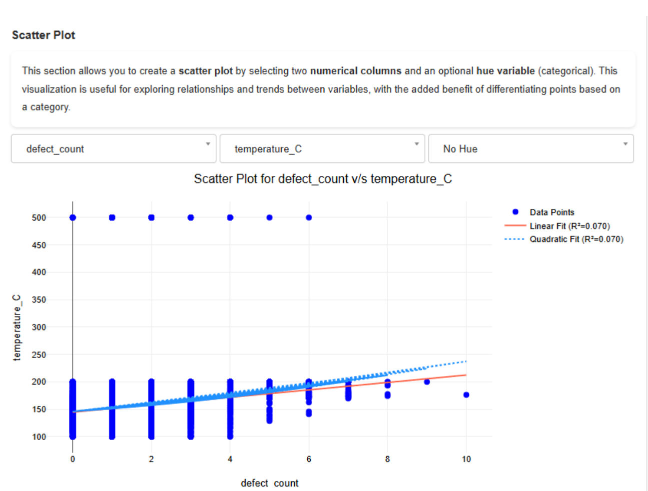


Fig. 9. Interactive scatter plot generated by the EDA agent. In this example, *defect_count* and *temperature_C* were selected to examine their correlation.

4.4 Automated Anomaly Detection Test Cases Result

Using the same wafer dataset previously provided to the Data Insight Agent and EDA Agent, the data was then passed to the Auto AD Agent for anomaly detection. The agent responded with two main sections: Anomaly Detection and Causal Inference, both of which focus on identifying anomalies within the data.

The bar chart in Fig. 10. concludes that among the categories, Etching shows the highest contribution, accounting for nearly 50% of the total anomalies detected. This suggests that the etching process is a significant source of irregularities within the dataset. Deposition and Implantation follow, contributing approximately 19% and 16% respectively. Lithography accounts for about 13%, while Etch (a possibly mislabeled or redundant category) contributes the least at around 3%. These insights highlight which stages of the process may require closer monitoring or process optimization to reduce the occurrence of defects.

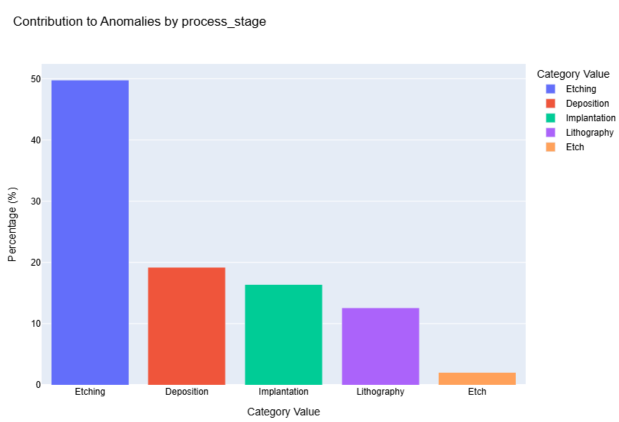


Fig. 10. Bar chart showing the contribution of each *process_stage* to anomalies, generated by the Auto AD agent.

5.0 CONCLUSION

This paper introduced a GenAI-based Data Insights system to streamline data analysis and decision-making in semiconductor assembly operations in WDC. By combining natural language interfaces with automated modules for Data Insight, Auto-EDA and Auto-AD agents, the system significantly reduced manual workload and enhanced insight accessibility across technical and non-technical users.

The implementation achieved a reduction of approximately 80% in manual analysis time, while ensuring secure, internal handling of sensitive manufacturing data. These results demonstrate the feasibility and effectiveness of deploying domain-specific GenAI solutions in high-volume, data-intensive industrial environments.

6.0 RECOMMENDATIONS

To extend the impact of the proposed system, several enhancements are recommended. First, integrating the platform with real-time data sources such as manufacturing execution systems or IoT-based sensors would enable continuous monitoring and immediate anomaly detection. Second, fine-tuning the underlying GenAI model using domain-specific data can improve contextual accuracy and the relevance of generated insights. Incorporating a user feedback mechanism is also advised to support iterative prompt optimization and continuous system learning. Additionally, enhancing the visualization layer to support interactive features, such as drill-downs and cross-variable comparisons, would provide users with deeper analytical capability. Finally, implementing access control and data governance policies will ensure secure and role-based access to both raw data and generated reports, thereby maintaining compliance with internal security standards.

7.0 ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Operations Analytics Excellence department for their full support and guidance throughout the development of this project. Furthermore, appreciation is also extended to the Information Technology team for providing the necessary platforms and technologies that enabled the successful implementation of the system.

8.0 REFERENCES

1. P. Vejjanugraha et al., “An Automated Data Analytics and Overall Equipment Effectiveness Visualization Technique for Assembly Line on Continuous Manufacturing System using Power BI,” in **2022 7th International Conference on Business and Industrial Research (ICBIR)**, Bangkok, Thailand, **2022**, pp. 562-567, doi: 10.1109/ICBIR54589.2022.9786497.

2. D. Noskievičová, T. Smajdorová and E. Tylečková, “Statistical Process Control in Big Data Environment,” in **2020 21th International Carpathian Control Conference (ICCC)**, High Tatras, Slovakia, **2020**, pp. 1-6, doi: 10.1109/ICCC49264.2020.9257251.

3. L. M. Choong and W. Kuang, “Advanced Data Analytics and Supervised Machine Learning in Optics Engineering Analysis,” in **2021 International Conference on Computer & Information Sciences (ICCOINS)**, Kuching, Malaysia,

2021, pp. 157-159, doi: 10.1109/ICCOINS49721.2021.9497207.

4. Y.-L. Chen et al., "Exploring Machine Learning for Semiconductor Process Optimization: A Systematic Review," in **IEEE Transactions on Artificial Intelligence**, pp. 1–21, **2024**, doi: 10.1109/tai.2024.3429479.

5. S.-Y. Hung, C.-Y. Lee, and Y.-L. Lin, "Data science for delamination prognosis and online batch learning in semiconductor assembly process," in **IEEE Transactions on Components Packaging and Manufacturing Technology**, vol. 10, no. 2, **2019**, pp. 314–324, doi: 10.1109/tcpmt.2019.2956485.

6. S.-Y. Hung, C.-Y. Lee, and Y.-L. Lin, "Data science for delamination prognosis and online batch learning in semiconductor assembly process," in **IEEE Transactions on Components Packaging and Manufacturing Technology**, vol. 10, no. 2, **2019**, pp. 314–324, doi: 10.1109/tcpmt.2019.2956485.

7. P.-C. Shen and C.-Y. Lee, "Wafer Bin Map recognition with Autoencoder-Based data augmentation in semiconductor assembly process," in **IEEE Transactions on Semiconductor Manufacturing**, vol. 35, no. 2, **2022**, pp. 198–209, doi: 10.1109/tsm.2022.3146266.

8. C.-S. Wang and J.-E. Chiu, "Shear force classification before wire bonding based on probe mark 2-D images using machine learning methods," in **IEEE Transactions on Semiconductor Manufacturing**, vol. 35, no. 2, **2022**, pp. 210–219, doi: 10.1109/tsm.2022.3162401.

9. D. Pradeep, B. V. Vardhan, S. Raiak, I. Muniraj, K. Elumalai, and S. Chinnadurai, "Optimal Predictive Maintenance Technique for Manufacturing Semiconductors using Machine Learning," in **2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)**, **2023**, pp. 1–5, doi: 10.1109/icct56969.2023.10075658.

10. H. Yu et al., "Expediting manufacturing launch using integrated data with AI/ML analytic solutions," in **IEEE Journal of the Electron Devices Society**, **2024**, p. 1, doi: 10.1109/jeds.2024.3522126.

11. Y. Cao, Y. Ni, Y. Zhou, H. Li, Z. Huang, and E. Yao, "An auto chip package surface defect detection based on deep learning," in **IEEE Transactions on Instrumentation and Measurement**, vol. 73, **2023**, pp. 1–15, doi: 10.1109/tim.2023.3347799.

12. A. Pyne, "Wafer Dataset," **Kaggle**, [Online]. Available: https://www.kaggle.com/datasets/avijit15/wafer-dataset. [Accessed: June 13, 2025].

13. E. Uhunoma, "From hours to minutes: The impact of generative AI on work turnaround time and productivity," **United Int. J. Res. Technol.**, vol. 6, **2025**, pp. 98–108.

9.0 ABOUT THE AUTHORS

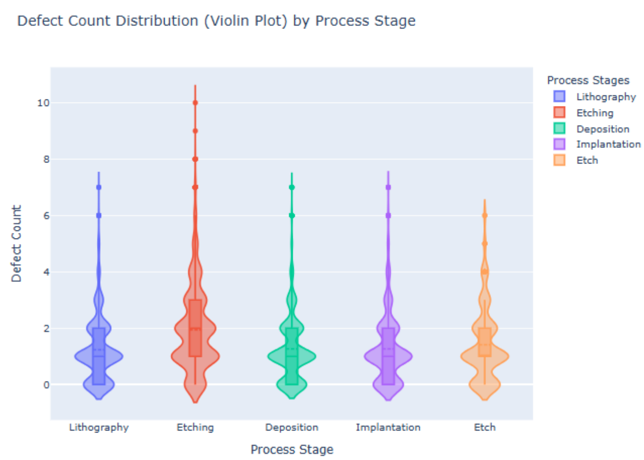
Alberto Zaldivar is a Technology Lead at Western Digital, with a total of 20 years of experience. He holds a Bachelor of Science in Computer Engineering from Adamson University.

Jerald Constantino is an Associate Data Scientist at Western Digital Corporation Philippines, specializing in machine vision, machine learning, and generative AI projects. He holds a Bachelor of Science in Computer Engineering from Batangas State University - TNEU and is currently studying for a Master of Science in Computer Engineering at Mapúa University.

Jhon Vincent Gupo is an Associate Data Scientist at Western Digital Storage Technologies for Operation Excellence. He is also a recognized Digital Twin Champion. He holds a bachelor's degree in computer science at Laguna State Polytechnic University – Los Baños.

10.0 APPENDIX

Appendix A – Expanded Visualizations for Defect Count Analysis using Data Insights Agent



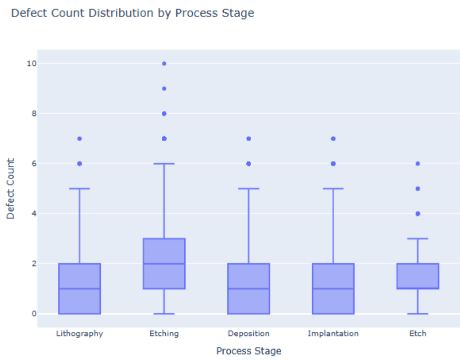
Defect Distribution Density Analysis

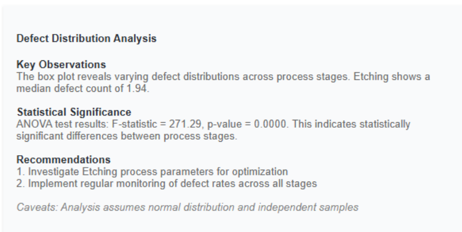
Key Observations
The violin plot shows the complete distribution shape of defect counts, revealing asymmetric patterns.

Distribution Characteristics
Each process stage shows distinct defect count distributions with varying spread and central tendencies.

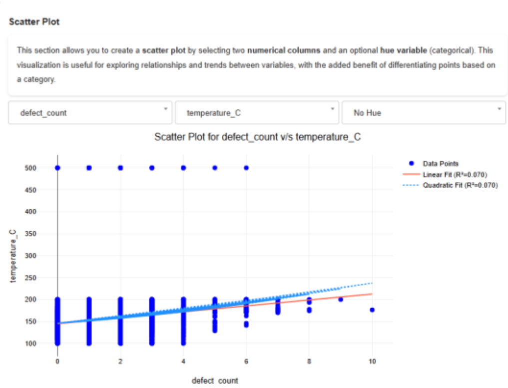
Recommendations
1. Focus quality control efforts on processes with higher defect density
2. Standardize processes to reduce variability in defect counts

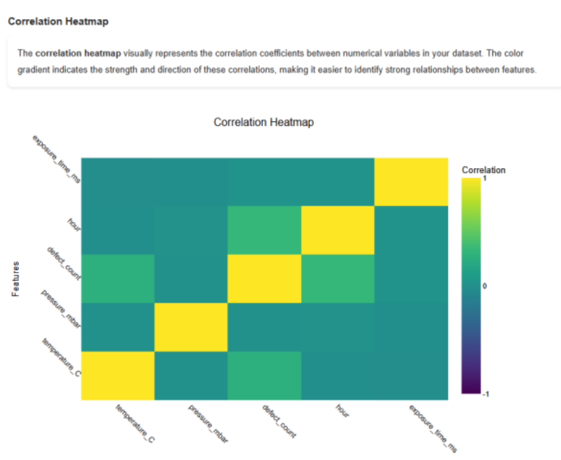
Caveats: Density estimation may be affected by sample size differences between stages





Appendix B – Expanded Visualizations of Scatter Plot and Correlation Heatmap which is a result of Auto EDA Agent





Appendix C – Expanded Visualizations of the result provided by Auto AD Agent showcasing Anomaly Detection and Causal Inference section