2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED P&ID COMPLIANCE CHECKING VIA LLM-BASED CODE GENERATION

Heejun Youn[1], Miyoung Uhm[1], Juho Han[1], Justin S. Lee[1], H. David Jeong[2], and Ghang Lee[1,3]
[1]Yonsei University, Seoul, Korea, Republic of South Korea
[2]Texas A&M University, College Station, TX, USA
[3]Technical University of Munich, Munich, Germany

Abstract

Manual P&ID review is time-consuming, while existing image-based or hard-coded automation methods lack efficiency. We propose a hybrid framework that extracts P&ID data from DXF files and converts natural language rules into FOL-based Python verification modules. By restricting the LLM to code generation, the framework ensures deterministic rule verification via formal logic. Evaluated on four industrial drawings (140 objects, 587 checks), our approach achieves a 97.5% F1-score, compared to 55.9% for the image-based LLM baseline. This enables non-experts to automate design, conditional, and engineering rules, advancing QA in digital twin environments.

Introduction

Piping and Instrumentation Diagrams (P&IDs) are critical engineering documents in plant facility design, containing information on equipment, piping systems, instrumentation, and process flows. Quality assurance of P&IDs is essential for plant safety, operational efficiency, and compliance with industry standards. However, the current review process still relies heavily on experienced engineers who must manually cross-reference thousands of drawings against multiple design specifications. This process is labor-intensive and vulnerable to human errors, particularly when dealing with complex and project-specific design rules.

The recent emergence of Large Language Models (LLMs) has opened new possibilities for bridging the gap between natural language specifications and executable code. Yet the direct application of LLMs for P&ID verification raises serious concerns: hallucinations during inference and the non-deterministic nature of outputs pose significant risks in safety-critical engineering contexts where reproducibility and accuracy are required. These limitations highlight the need for a hybrid approach that leverages the flexibility of LLMs for code generation while delegating the actual verification to deterministic reasoning engines.

Existing approaches to automated P&ID quality assurance fall into three broad categories, each with notable limitations. Image-based methods using computer vision achieve high symbol recognition accuracy but cannot verify compliance with semantic design rules and are sensitive to drawing quality (Alimin et al., 2025; Kim et al., 2025). Rule-based systems provide deterministic validation but require substantial expert effort for rule encoding and lack flexibility when adapting to project-specific requirements (Balhorn et al., 2025). Machine learning approaches such as Transformer-based semantic analysis show promise on large datasets but raise concerns about interpretability and portability across heterogeneous engineering environments (Gómez-Vega et al., 2025). A detailed review of these studies is provided in Section 2.

To address these gaps, this study proposes an automated P&ID verification framework that combines DXF-based data extraction with First-Order Logic (FOL)-based rule verification, restricting the role of LLMs to code and parser generation rather than direct validation. The core contributions are as follows:

1) We present a hybrid verification framework that extracts engineering objects directly from DXF-format CAD files through layer analysis and pattern recognition and transforms natural language design rules into executable Python verification functions via FOL formalization and LLM-assisted code generation. By confining LLM involvement to the code generation phase, the framework ensures that safety-critical verification outcomes remain deterministic and reproducible.
2) We formalize nine categories of Annotation Rules (AR) and Basic Design Rules (BR) into FOL expressions that serve as unambiguous, machine-verifiable specifications. This neuro-symbolic approach enables users with limited coding expertise to modularize complex design rules without manual hard-coding.
3) We validate the framework on four industrial P&ID drawings comprising 140 engineering objects and 587 rule checks. The proposed approach achieves an overall extraction F1 of 97.5% with a 100% rule compliance pass rate, compared to 55.9% F1 for the image-based LLM baseline. Equipment and Line extraction maintain 100% F1 across all four drawings without parser modification, demonstrating generalization beyond a single test case.

Related Work

Rule-Based Verification Systems

Schulze Balhorn et al. (2025) proposed a systematic approach that converts P&IDs into DEXPI-based structured data or graph representations to detect design rule violations through explicit pattern matching. Their method demonstrated clear advantages in identifying specific violation patterns and maintaining compatibility with industry standards but requires substantial effort for rule definition and lacks flexibility in processing natural language specifications. Additionally, scalability is limited when adapting to diverse project requirements.

Machine Learning-Based Approaches

Gómez-Vega et al. (2025) introduced Smart P&ID, which extracts text and attribute data from databases and CSV files, and transforms piping information into token sequences processed by Transformer-based models such as DistilBERT. Their approach showed high accuracy in detecting semantic inconsistencies in large datasets and demonstrated promising adaptability to rule changes. The proposed method faces challenges in interoperability and portability to other engineering environments, as it is heavily dependent on specific Smart PID data structures and does not fully account for legacy CAD systems or image-oriented PDF formats commonly used in the field.

Image-Based Detection Methods

Alimin et al. (2025) employed computer vision techniques including YOLO, Faster R-CNN, and SIFT/ORB algorithms for symbol detection and geometric verification in P&IDs. While achieving over 90% recognition accuracy in symbol identification, these methods are limited to detecting visual inconsistencies and cannot verify compliance with semantic design rules. Furthermore, the performance of these approaches is highly sensitive to image quality and resolution, limiting their applicability in real-world scenarios where drawing formats and quality vary. Beyond these three categories, Additional related work in DEXPI-based interoperability, graph-based P&ID analysis, and digital twin-integrated validation will be addressed in the journal extension of this study.

Methodology

This study develops an automated rule verification system for P&ID drawings through a two-phase approach. First, we conduct a comparative test to evaluate data extraction performance between image-based LLM approaches and DXF-based direct parsing. Second, based on the extraction results, we implement a FOL-based rule verification framework that transforms natural language design specifications into executable verification modules. This section describes experimental design, data extraction methodologies, and rule verification architecture.

Rule Set Composition

P&ID design review rules can be categorized as follows:

- Annotation Rules (AR): Rules related to notation and compliance with international standards.
- Basic Design Rules (BR): Rules related to the presence of essential design information.
- Conditional Rules (CR): Rules related to conditional design requirements.
- Design Rules (DR): Rules related to compound design principles.
- Engineering Rule (ER): Rules related to engineering logic and calculations.

These rules verify the presence or absence of required information elements and check basic formatting compliance, without involving complex conditional logic, engineering calculations, or cross-drawing relationships. Since this study focuses on establishing baseline extraction and verification capabilities for AR and BR rules, selecting a drawing that comprehensively tests these rule categories provides the most robust validation of the proposed framework's fundamental capabilities.

For the test, nine design rules were implemented from a total of eleven rules defined across the AR and BR categories, excluding rules that involve inter-rule dependencies. The AR rules (AR-1 to AR-5) verify drawing number uniqueness, entity identification schemes, pipe nominal diameter, pipe number, and material class. The BR rules (BR-1 to BR-4) verify name,

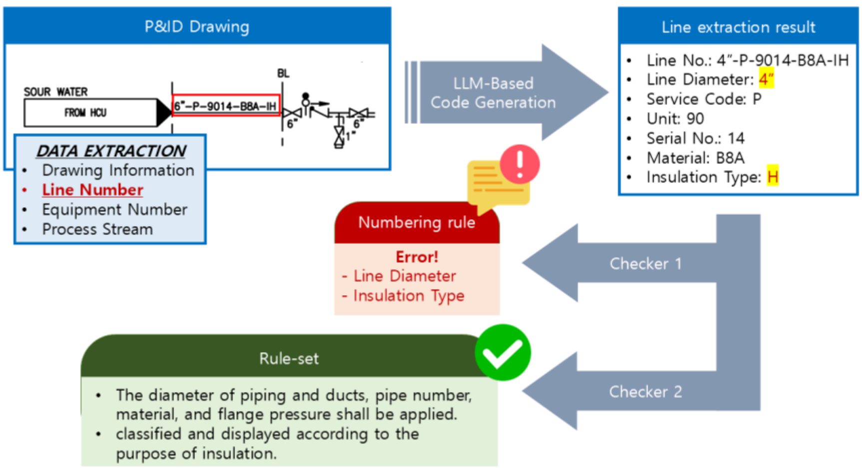

Figure 1. P&ID Compliance Checking via LLM-based Code Generation Framework

specification, type, and references to connected drawings (Table 1).

Table 1. Selected AR and BR rules

Class	Code	No.	Keyword
Annotation	AR	01	Drawing Number
	AR	02	Object ID
	AR	03	Line Diameter
	AR	04	Line Number
	AR	05	Line Material
	AR	06	Valve Failure mode
Basic Design	BR	01	Equipment ID
	BR	02	Equipment Spec.
	BR	03	Equipment Type
	BR	04	Process Flow
	BR	05	Manhole/Tray

These nine rules are applied to all extracted entities, including equipment, piping, instruments, and connected drawings.

Ground truth annotations were manually created. All equipment tags, line numbers, service codes, material classes, and instrumentation tags were extracted and cross-verified against the original design specification documents. Each extracted item is classified as True Positive (TP), False Positive (FP), or False Negative (FN) against the ground truth. Precision, Recall, and F1 Score are computed per category and macro-averaged across categories. These metrics are employed not as machine learning performance indicators but as standard information retrieval measures to quantify extraction completeness and correctness under a unified evaluation framework for both approaches.

P&ID Drawing

Seven P&ID drawings from an industrial Sour Water Stripper unit (A39002-500-A 101 through A107) were available for this study. Drawings A101–A103 are Symbol & Legend sheets defining project-wide naming conventions and serve as the reference for numbering rule verification. The remaining four drawings (A104–A107) are P&IDs, each containing equipment, piping lines, and process stream information subject to AR and BR rules (Table 2).

Table 2. Reviewable rules in each P&ID drawing

P&ID No.	Class	Adaptable Rules
A39002-500-A 101	Standard	1
A39002-500-A 102	Standard	1
A39002-500-A 103	Standard	1
A39002-500-A 104	Drawing	11
A39002-500-A 105	Drawing	9
A39002-500-A 106	Drawing	10
A39002-500-A 107	Drawing	11

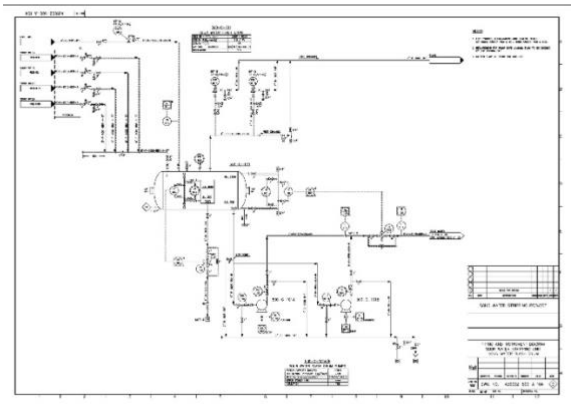

Figure 2. A39002-500-A 104 P&ID drawing

Among these, Drawing A104 (Sour Water Flash Drum) was obtained in both PDF and DXF formats to enable a fair comparison between the image-based LLM extraction (Approach A) and DXF-based parsing (Approach B). Drawings A105–A107 were obtained in DXF format only and were used together with A104 to evaluate the generalization of the proposed framework across drawings of varying complexity. The four P&IDs collectively contain 180 items

Table 3. Test data summary across four drawings

P&ID No.	Drawing Info	Equipment	Lines	Process Streams	Total
104	10	2	26	7	45
105	10	3	19	5	37
106	10	4	22	6	42
107	10	5	34	7	56
Total	40	14	101	25	180

Experiments

Approach A: Image-based LLM Extraction

The first approach utilizes a LLM equipped with vision-language capabilities to directly extract design information from P&ID drawings rendered as PDF images. This approach represents the prevalent image-based LLM methodology in recent automated P&ID interpretation research and serves as the baseline for comparing the performance of the proposed DXF-based automated review framework.

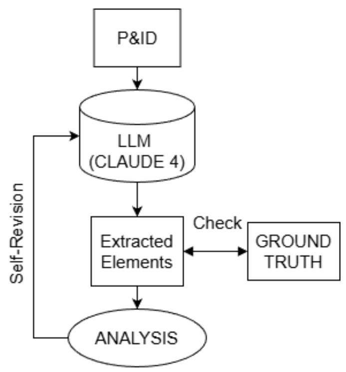

Figure 3. The flow of Image-based LLM Extraction

In this methodology (Figure 3), P&ID drawings are converted into high-resolution images and used as the primary input. Without the aid of dedicated CAD parsing or rule-based engines, the model simultaneously performs text and symbol recognition to extract engineering data, such as equipment, piping, and instrumentation. The interpretation relies solely on the model's visual reasoning and internal domain knowledge. To ensure structured outputs, a specific prompt restricts the response format to JSON, explicitly defining key data fields—including drawing numbers, equipment tags, and piping specifications—required for Annotation Review (AR) and Basic Review (BR).

Claude Sonnet 4.5 (Anthropic, 2025) was employed for this extraction task as a widely accessible vision-enabled LLM. The Opus model was excluded due to prohibitive token costs and significantly slower inference speeds for large image inputs. Since Approach A aims to establish a representative baseline rather than to maximize LLM performance, Sonnet 4.5 was deemed sufficient for this purpose. The model processes the drawing image and a structured prompt concurrently to return engineering data in a predefined JSON schema. The extracted data are then evaluated against the ground truth for a quantitative comparison with the proposed method.

However, this approach possesses inherent structural limitations. Since drawing legends cannot be explicitly fixed as external knowledge, consistency in symbol interpretation is not guaranteed. Furthermore, it lacks the capability to verify complex logical structures, such as Conditional Rules (CR) or Detailed Design Rules (DR). Most notably, the occurrence of hallucinations during the inference process and the non-deterministic nature of the outputs pose significant challenges for safety-critical design reviews that demand high reliability. By establishing Approach A as a baseline, this study aims to highlight these deficiencies and demonstrate the superior reliability of the proposed DXF-based framework utilizing FOL.

Approach B: DXF-Based Direct Parsing

The second approach leverages the original CAD file format (DXF) to extract engineering information directly from vector data without image conversion (Figure 4).

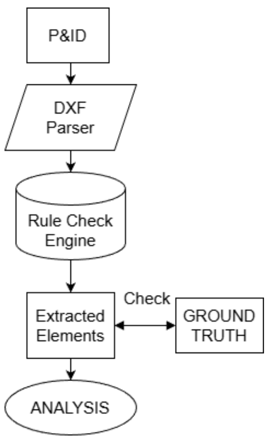

Figure 4. DXF-based data extraction pipeline

The DXF-based extraction approach offers several critical advantages over conventional image-based methods in terms of data integrity and analytical precision. By eliminating degradation from image conversion and vision interpretation errors, this method ensures lossless information preservation and facilitates spatial relationship analysis. Furthermore, it allows for direct access to the CAD organizational structure, including layer-based filtering, block definitions, and embedded metadata, thereby enhancing attribute richness. Beyond its technical accuracy, the DXF approach is highly scalable as processing time remains independent of image resolution. Most importantly, it yields deterministic results, ensuring reproducible extraction without the stochastic variation typically associated with image-based processing.

The role of LLMs in the framework is strictly limited to code and parser generation, not direct rule verification. This architectural choice mitigates hallucination risks while leveraging LLM flexibility.

The DXF-based data extraction process is structured into three distinct stages to ensure high fidelity and structural integrity.

- Stage 1: Entity Parsing. The ezdxf library for Python is employed to parse the DXF file, extracting all text-bearing entities while preserving their original Cartesian coordinates and layer associations. This stage captures raw textual content directly from the CAD environment without information loss.
- Stage 2: Pattern-Based Classification. Extracted text elements are categorized using domain-specific regular expressions derived from project-specific naming conventions formulated in accordance with identification of equipment tags, line numbers, and other standardized engineering identifiers.
- Stage 3: Coordinate-Based Attribute Association. Spatial analysis links equipment tags with their corresponding specifications through a deterministic search algorithm operating within a defined radius (default: 150 DXF units) around each identified tag. To ensure association accuracy, invalid patterns—including instrument tags, specification table headers, and control keywords—are filtered through rigorous exclusion rules. This mitigates data misinterpretation risks arising from spatial proximity of unrelated elements.

The verification engine consists of two checking modules.

Checker 1(Numbering Rules) verifies the formal integrity of the extracted information based on the numbering scheme defined in the Symbol & Legend sheet (A39002-500-A 101-103) of the P&ID drawing (Figure 5).

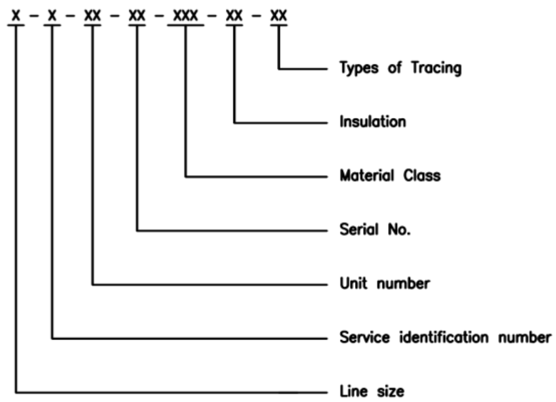

Figure 5. Line Identification

Checker 2 (AR & BR): This module examines whether mandatory information required by each rule is present and evaluates attribute completeness. Inspired by (Pan et al. 2023) neuro-symbolic framework, we adopt a similar three-stage paradigm—formulation, reasoning, interpretation—but adapt it specifically for P&ID verification. To establish unambiguous verification criteria, design rules are formalized using FOL.

FOL is essential for P&ID rule verification for three reasons. First, it eliminates ambiguity in natural language rules through precise semantic expression. For instance, the vague statement "Appropriate insulation shall be used for high-temperature piping." becomes:

$\forall x$ (Pipe(x) $\wedge$ HighTemperature(x) $\geq 65 \rightarrow$ Insulated(x).

Second, FOL systematically represents complex conditional rules and nested logical structures, enabling integration of multi-condition design rules (e.g., rules involving operating pressure, temperature, material, and fluid properties) into unified logical expressions. Third, formally expressed rules enable mechanical verification through automated theorem provers or model checkers, allowing pre-validation of logical consistency and conflict detection among rules.

Experimental Result

LLM Extraction Method Performance

To maximize extraction performance, the methodology underwent three iterations with progressive improvements:

Table 4. F1 Score Comparison by Data Categories

Category`	Test 1	Test 2	Test 3
Drawing info	100%	100%	100%
Equipment	100%	100%	100%
Lines	48%	48%	48%
Process streams	71.4%	71.4%	71.4%
Overall	55.9%	55.9%	55.9%

Table 5. Iterative Refinement Process

Class	Revision Detail	Result
Iteration 1 (Baseline)	Direct extraction using the base prompt without post-processing.	F1 55.88
Iteration 2 (Post processing)	Applied automated post-processing includes text normalization and format verification against P&ID naming conventions.	F1 55.88
Iteration 3 (Error Pattern Correction)	Analyzed common error patterns from previous iterations and incorporated specific correction rules including size extraction, service code verification, and missing entity detection.	F1 55.88

The image-based approach demonstrated several systematic failures:

1. Text recognition errors: Misrecognition of characters in low-contrast regions

2. Spatial relationship inference: Difficulty in correctly associating spatially disconnected text elements with their corresponding equipment or piping symbols, leading to attribute misassignment.

3. Overlapping element confusion: Inability to disambiguate text when overlapping with graphical symbols, leader lines, or dense annotations, resulting in incomplete or erroneous extraction.

4. Implicit attribute loss: Failure to extract attributes not explicitly labeled in text form, such as connection relationships inferred from line continuity or material specifications derived from line style conventions.

These limitations align with fundamental challenges of image-based approaches documented in prior studies (Alimin et al., 2025), demonstrating the inherent trade-offs between vision model flexibility and extraction reliability.

DXF-based Method Performance

The proposed DXF-based extraction and rule verification pipeline was applied to all four process drawings (A104–A107). Table.6 presents the extraction accuracy compared with the ground truth across 180 items. Drawing Info, Equipment, and Lines achieved 100% F1 across all four drawings. The three unmatched items occurred exclusively in A107 Process Streams, where F1 dropped to 53.3% (Precision 50.0%, Recall 57.1%), while A104–A106 maintained 100% in all categories.

Table 6. Extraction accuracy compared with Ground Truth

Category	Ground Truth	Extracted	Matched	F1
Drawing Info	40	40	40	100%
Equipment	14	14	14	100%
Lines	101	101	101	100%
Process streams	25	26	22	86.3%
Total	180	181	177	97.5%

Error analysis of A107 Streams identified 4 true positives, 4 false positives, and 3 false negatives. All errors originated from streams lacking explicit FROM/TO boundary markers in the DXF source: (1) anchor text absence preventing stream detection (1 case), (2) label-boundary mismatching causing line number misreading or false extraction (2 cases), and (3) incorrect direction assignment due to ambiguous spatial layout (1 case). These are inherent limitations of text-based spatial matching, not parser logic defects.

Table 7 presents the rule verification results. Checker 1 performed 119 numbering-format checks and Checker 2 conducted 468 AR/BR rule checks; all 587 items passed with zero errors. Thirteen warnings were issued for material class codes not registered in the rule-set list and missing equipment attributes in the DXF source. These warnings indicate incomplete source data rather than rule violations and therefore do not affect the pass/fail determination.

Table 7. AR/BR rule verification results across four drawings

Rule	Keyword	Checked Items	Warning
AR-1	Drawing Number	4 Drawings	0
AR-2	Object ID	140 IDs	0
AR-3	Line Diameter	101 Lines	0
AR-4	Line Number	101 Lines	0
AR-5	Line Material	101 Lines	10
BR-1	Equipment ID	14 Equipment	0
BR-2	Equipment Spec.	14 Equipment	2
BR-3	Equipment Type	14 Equipment	1
BR-4	Process Flow	25 connections	0

The AR rules (AR-1 to AR-5) verify formal integrity of drawing numbers, entity identification schemes, and piping attributes including nominal diameter, line number format, and material class. The BR rules (BR-1 to BR-4) validate the completeness of equipment information and process flow references to connected drawings. All nine rules passed across the four drawings, confirming compliance with the project naming conventions defined in the Symbol & Legend sheets (A39002-500-A 101–103).

Analysis of LLM Extraction vs. DXF parsing

In the context of P&ID information extraction, LLM–based approach demonstrated effectiveness in identifying textual elements and equipment tags within drawings. However, several structural limitations were observed in the extraction of piping and process-related data. First, due to the nature of image-based inputs, text misdetection and information loss occurred depending on drawing resolution and visual noise. In addition, probabilistic inference during numerical data extraction introduced the possibility of incorrect outputs. Furthermore, frequent false detections were observed when spatially adjacent elements visually interfered with each other, resulting in mixed or conflated data. These issues can be attributed to the inherent sensitivity of vision-based algorithms and probabilistic models to complex physical layouts and visual variability within engineering drawings.

- Data Integrity: By eliminating image conversion, information loss and visual interpretation errors are fundamentally avoided.
- Geometric Precision: Direct utilization of numerical coordinate data enables sub-millimeter–level accuracy in spatial relationship analysis.
- Structural Accessibility: CAD-native metadata, including layer hierarchies, block definitions, and line types, allows for explicit identification of semantic relationships among design elements.
- Computational Efficiency and Reliability: The approach maintains consistent processing performance independent of image resolution and guarantees reproducibility through deterministic, non-probabilistic outcomes.

Nevertheless, the DXF-based approach has its own limitations. When explicit boundary markers such as FROM/TO anchor texts are absent from the source drawing, the coordinate-based stream detection relies on spatial heuristics that can misidentify stream boundaries or flow directions. This was observed in A107 Process Streams, where the absence of boundary markers led to line number misreading, false extraction, and direction assignment errors, resulting in an F1 of 53.3%. These errors are inherent to text-based spatial matching rather than defects in the parser logic, and they define the current boundary condition of the framework.

Consequently, for design rule checking in plant engineering where high reliability and strict reproducibility are critical, the DXF-based structural approach is considered to provide a more stable and trustworthy data processing foundation than an LLM-only model. Table 8 presents this comparison: on the same drawing (A104), Approach B achieved 100% F1 across all categories, whereas Approach A reached 55.9%. When Approach B was extended to four drawings without parser modification, Overall F1 remained at 97.5%, with the reduction attributable solely to process stream detection in A107.

Table 8. Performance and iteration comparison table to match ground truth

Class	Approach A (LLM)	Approach B (Code)	
Drawing	A104	A104	A104-107
Input Data	PDF	DXF	
Method	Vision LLM	DXF Parsing	
Iteration	3	1	
Drawing info F1	100%	100%	100%
Equipment F1	100%	100%	100%
Line F1	48%	100%	100%
Process F1	71.4%	100%	86.3%
Overall F1	55.9%	100%	97.5%

Discussion

The experimental results demonstrate that the proposed framework achieves high accuracy in extracting Equipment and Line information from DXF-format P&ID drawings, successfully passing all 587 design rule checks across four test drawings containing 140 engineering objects. Notably, the parser and rule engine developed for A104 were applied to A105–A107 without modification, and Equipment and Lines maintained 100% F1 across all drawings. This provides evidence that the DXF parsing algorithm and FOL-based rule verification engine can generalize effectively across real-world plant engineering drawings of varying complexity. By structuring high-level and complex rules from the P&ID rule definition document into formal logic (FOL), the proposed approach mitigates the hallucination issues frequently observed in standalone LLM-based methods and enables deterministic and reproducible verification results.

When LLMs were applied independently, they exhibited strong reasoning capabilities but failed to provide reliable verification outcomes due to persistent hallucination problems. In the plant engineering domain, where design quality is directly linked to safety incidents, probabilistic models alone are insufficient to guarantee the level of accuracy and consistency required for design verification. Notably, AR and BR rules that could not be reliably validated by standalone LLMs were verified with high accuracy through the proposed framework, which combines LLM-based code generation with symbolic logic verification. This result highlights the practical effectiveness of the proposed approach.

Compared to traditional manual drawing review processes, the proposed automated data extraction and rule verification system significantly reduces review time and effectively minimizes the risk of human error. By integrating the flexibility of LLM-based code generation with the deterministic reliability of formal logic-based verification, the proposed approach overcomes key limitations of existing automated P&ID quality verification studies.

The proposed framework addresses LLM hallucination issues through a carefully designed separation of concerns. LLMs are restricted to two specific roles: (1) transforming natural language rule specifications into structured FOL representations, and (2) generating Python verification code from these formal logic structures. The actual verification process is performed entirely by deterministic symbolic reasoning engines, eliminating the probabilistic uncertainty inherent in direct LLM-based verification. This architectural design ensures that while the system benefits from LLM flexibility in code generation, the safety-critical verification outcomes remain reproducible and deterministic.

However, the framework's current boundary condition was revealed in A107 Process Streams, where F1 dropped to 53.3%. All errors occurred in streams lacking explicit FROM/TO boundary markers, indicating that the coordinate-based spatial matching is sensitive to the completeness of anchor text in the source DXF. This limitation does not affect Equipment or Line extraction, which rely on self-contained tag patterns, but it suggests that process stream detection requires either enhanced spatial heuristics or supplementary metadata to achieve robustness comparable to other categories.

Conclusions

This study proposes an automated data extraction and design rule compliance verification system for P&ID drawings in the plant engineering domain. The proposed system parses DXF-format P&ID drawings to extract engineering objects such as Equipment, Lines, and Instruments as structured data and automatically verify compliance with design standards using FOL-based rule engine. Experimental validation on four P&ID drawings containing 140 engineering objects demonstrates that the system achieves 100% F1 for Equipment and Line extraction and 97.5% Overall F1 and successfully performs exhaustive verification of 587 design rule checks across nine categories of AR and BR rules.

The main contributions of this research are as follows. First, a DXF-based P&ID parsing algorithm was developed to automatically extract engineering objects through layer structure analysis and text pattern recognition. Second, an FOL-based rule verification engine was implemented to formally represent plant design standards and perform systematic compliance verification. Third, an end-to-end automation pipeline was established that integrates drawing input, data extraction, rule verification, and result report generation, thereby substantially reducing the manual review burden on design engineers.

This study acknowledges several limitations that warrant further investigation. First, although validation was extended to four P&ID drawings comprising 140 engineering objects, all drawings originate from a single industrial project, which limits conclusions about generalization performance across diverse projects, drawing conventions, and industrial contexts. Second, the current implementation addresses nine categories of AR and BR rules, whereas real-world projects require more comprehensive coverage including equipment-specific requirements (e.g., manhole and tray installation specifications) and failure mode documentation standards. Future research will address these limitations through a phased approach:

- **Immediate Extension**: The scope of validation will be expanded to various industrial plant types, including petrochemical and power generation, and water treatment facilities. By securing real-world project drawings, the research will broaden the application of AR and BR rules to verify the system's generalization performance and derive industry-specific adaptation requirements.
- **Medium-term Goals**: Technical capabilities will be developed to support CRs, DRs and ERs. This involves advancing context-aware verification logic based on FOL expressions, integrating domain-specific calculation engines for pressure drop and heat balance verification, and establishing a cross-drawing consistency-checking mechanism.
- **Long-term Vision**: Achieving Digital Twin integration with real-time verification capabilities. This encompasses developing bidirectional synchronization between P&ID models and Digital Twin platforms, implementing continuous quality assurance during design iterations, and establishing automated compliance reporting aligned with international standards (ISO, ASME, API).

This study demonstrates the practical feasibility of automated P&ID quality verification in plant engineering contexts. The successful validation of AR and BR rules on industrial drawings, combined with the framework's user-centric design that enables non-programmers to define rules in natural language, suggests significant potential for industry adoption.

However, successful deployment in engineering organizations will require addressing several practical considerations: (1) integration with existing CAD workflows and document management systems, (2) establishment of verification protocols for LLM-generated code before production use, and (3) development of user training programs for effective rule specification and result interpretation.

Despite these implementation challenges, the adoption of automated verification systems is expected to deliver substantial benefits including reduced design review time (potentially 50-70% reduction based on pilot testing), improved consistency in rule application, and prevention of human errors in repetitive checking tasks. These improvements directly contribute to enhanced safety outcomes and economic efficiency in plant engineering projects, while freeing experienced engineers to focus on high-level design decisions and complex problem-solving activities.

In the future, this study lays the foundation for extending beyond single-project verification to automated quality assurance encompassing CRs, DRs and ERs. This extension will contribute to comprehensive quality assurance automation within digital twin environments, supporting broader digital transformation initiatives in plant engineering and operations.

Acknowledgments

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (No. RS-2024-00407028) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2025-25436231).

References

Balhorn, L. S., Seijsener, N., Dao, K., Kim, M., Goldstein, D. P., Driessen, G. H. M., & Schweidtmann, A. M. (2025). Rule-based autocorrection of Piping and Instrumentation Diagrams (P&IDs) on graphs (arXiv:2502.18493). arXiv. https://doi.org/10.48550/arXiv.2502.18493

Alimin, A. A., Goldstein, D. P., Balhorn, L. S., & Schweidtmann, A. M. (2025). Talking like Piping and Instrumentation Diagrams (P&IDs) (arXiv:2502.18928). arXiv. https://doi.org/10.48550/arXiv.2502.18928

Gómez-Vega, F. S., Acuña, O., Camargo, A. C., Jimenez, J. D., Galeano, S. M., Franco, I. E., Lozano, L. L., Vásquez, J., & Puertas, E. (2025). Enhancing consistency in piping and instrumentation diagrams using DistilBERT and smart PID systems. Systems and Soft Computing, 7, 200373. https://doi.org/10.1016/j.sasc.2025.200373

Pan, L., Albalak, A., Wang, X., & Wang, W. Y. (2023). Logic-LM: Empowering Large Language Models with Symbolic Solvers for Faithful Logical Reasoning (arXiv:2305.12295). arXiv. https://doi.org/10.48550/arXiv.2305.12295

Jianwu, L. W., Yew, L. S., On, L. K., Keong, T. C., Yuan Sheng, R. T., Sani, S. B., & Juan Agnes, T. H. (2024). Artificial intelligence-enabled evaluating for computer-aided drawings (AMCAD). International Journal of Mechanical Engineering Education, 52(1), 3–31. https://doi.org/10.1177/03064190231175231

Dzhusupova, R., Banotra, R., Bosch, J., & Olsson, H. H. (2023). Using artificial intelligence to find design errors in the engineering drawings. Journal of Software: Evolution and Process, 35(12), e2543. https://doi.org/10.1002/smr.2543

Kim, J.-B., Moon, Y., Han, S.-T., & Mun, D. (2025). Automated inspection of P&ID object recognition using deep learning. Scientific Reports, 15(1), 39031. https://doi.org/10.1038/s41598-025-25506-2