29th International Congress on Project Management and Engineering
Ferrol, 16th-17th July 2025

# 01-048 – Analysis of electrical projects from single-line diagrams using generative AI – Análisis de proyectos eléctricos a partir de diagramas unifilares mediante IA generativa

*Onabanjo, Olumayowa*[1]; *Ortega Fernandez, Francisco*[1]; *Díaz Piloñeta, Marina*[1]; *Martínez Huerta, Gemma*[1]; *Moreno-García, Carlos*[2]
(1) Universidad de Oviedo, (2) Robert Gordon University

📄Spanish 🎤Spanish

This communication presents a methodology for the automatic generation of electrical projects using generative artificial intelligence, based on single-line diagrams. The proposed approach aims to optimize the process of developing technical documentation in electrical projects, significantly reducing the time and resources required, while ensuring compliance with current regulations. The developed system employs advanced image processing and deep learning techniques to interpret single-line diagrams, extracting crucial information about the topology and components of the electrical installation. In this first stage, the topology is extracted and interpreted, associating characteristics such as power, etc., to each element or line. This allows for calculations to be performed and, ultimately, through natural language models specifically trained in electrical regulations and project standards, to automatically generate the justifying calculations. Results from the system's implementation in real case studies are presented, demonstrating its effectiveness in terms of accuracy, coherence, and compliance with legislation. Furthermore, the implications of this technology for engineering firms are discussed, highlighting its potential to improve efficiency and quality in project development.

**Keywords:** *Generative artificial intelligence; Electrical projects; Single-line diagrams; Automation; Electrical regulations*

Esta comunicación presenta una metodología para la generación automática de proyectos eléctricos utilizando inteligencia artificial generativa, partiendo de diagramas unifilares. El enfoque propuesto busca optimizar el proceso de elaboración de documentación técnica en proyectos eléctricos, reduciendo significativamente el tiempo y los recursos necesarios, a la vez que se garantiza el cumplimiento de la normativa vigente. El sistema desarrollado emplea técnicas avanzadas de procesamiento de imágenes y aprendizaje profundo para interpretar los diagramas unifilares, extrayendo información crucial sobre la topología y componentes de la instalación eléctrica. En esta primera etapa, se extrae e interpreta la topología, asociado a cada elemento o línea sus características como potencia, etc. De modo que se puedan realizar sus cálculos y, en última instancia, mediante modelos de lenguaje natural entrenados específicamente en normativas eléctricas y estándares de proyectos, generar automáticamente los cálculos justificativos. Se presentan los resultados de la implementación del sistema en casos de estudio reales, demostrando su eficacia en términos de precisión, coherencia y conformidad con la legislación. Además, se discuten las implicaciones de esta tecnología para las ingenierías, destacando su potencial para mejorar la eficiencia y calidad en el desarrollo de proyectos.

**Palabras claves:** *Inteligencia artificial generativa; Proyectos eléctricos; Diagramas unifilares; Automatización; Normativa eléctrica*

## Acknowledgments:

Grants for Research Groups of Public R&D&I Organizations in the Principality of Asturias, Call 2024: 'Project Engineering and Sustainable Engineering' IDE/2024/000758"

## 1. Introduction and Objectives

Generative Artificial Intelligence (GAI) is a subset of artificial intelligence (AI) and a transformative technology that utilizes insights from vast amounts of data to deliver new content based on user instructions (Byrne et al., 2025). Among its most prominent developments are generative Large Language Models (LLMs), whose intuitive conversational capabilities have rapidly gained attention across disciplines, including engineering design research (Doris et al., 2024; Chiarello et al., 2024; Göpfert et al., 2024).

In the context of increasing demands for energy efficiency, agile and accurate engineering design has become critical, particularly in low-voltage electrical projects. However, many engineering firms still depend on manual documentation processes, which limit productivity and introduce frequent errors. Moreover, engineering documentation often includes multimodal formats such as diagrams, tables, and graphs that go beyond what traditional text-based LLMs can interpret effectively. The emergence of Multimodal Large Language Models (MLLMs) addresses this limitation, offering promising capabilities for automating complex documentation tasks (Doris et al., 2024).

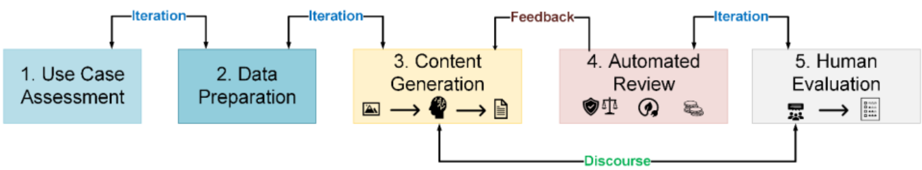
Despite these advances, electrical engineering diagrams can be complex, and MLLM spatial understanding is still evolving (Chiarello et al., 2024; Göpfert et al., 2024). These limitations signal the need for domain-specific benchmarks (B. Song et al., 2024; Doris et al., 2024; Geipel, 2024; Göpfert et al., 2024), as well as more human-centric evaluations (Göpfert et al., 2024). Systematic human evaluations, particularly those conducted by researchers, identify GAI shortcomings and drive subsequent improvements. For example, newer models promise better understanding of complex diagrams and charts (Poccia, 2024). Additionally, earlier models are constantly refined for reasoning and calculations (OpenAI, 2024).

To ensure these improvements translate to increased industrial application, research on MLLM capabilities for design-specific issues is essential. However, there are currently few studies, with limited examples on electrical diagrams such as by Geipel (2024). To address this gap, a scalable methodology is proposed for electrical project documentation generation using low-voltage single-line diagrams as a case study. This methodology is based on best practice from literature and is validated on anonymized real-world diagrams on critical design tasks (multimodal design reasoning and calculations), using state-of-the-art models. The work also provides practical implementation guidelines to support adoption by industry professionals.

## 2. Methodology Overview

This section presents a structured methodology to automate electrical project documentation using Generative Artificial Intelligence (GAI), following a five-stage pipeline: (1) Use Case Assessment, (2) Diagram Digitization, (3) Content Generation, (4) Automated Review, and (5) Human Evaluation. These stages are derived from best practices in AI implementation (Dzhusupova et al., 2024), design discourse between users and GAI (Göpfert et al., 2024) and aligned with the CRISP-DM framework. Figure 1 summarizes the full workflow. Each stage is detailed below, including challenges, tools, and integration points.

**Figure 1: Electrical Project Documentation with Generative Artificial Intelligence (Methodology).**



## 2.1 Use Case Assessment

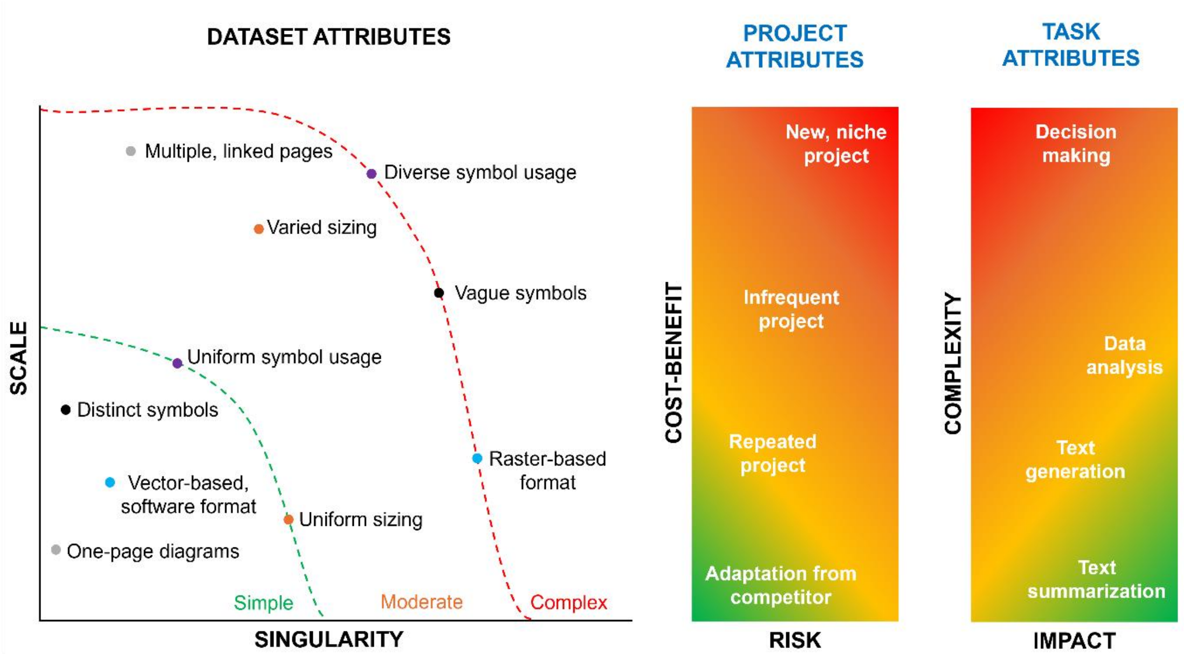
The methodology begins with an assessment of the use case feasibility, based on diagram, project, and task attributes. Electrical single-line diagrams (SLDs) are utilized in private, corporate, and public projects to effectively communicate complex system configurations. SLDs are exchanged among designers, engineers, and clients in various formats. Proprietary software and vector-based formats are easier for AI to process but are not widely transferred between organizations, while raster-based formats are more common but challenging to digitize. However, raster digitization has advanced due to novel deep learning-based methods (Bhanbhro et al., 2023c).

Digitization alone does not ensure viable MLLM integration. Firms must contextualize their case to identify specific challenges that GAI can address. Figure 2 presents a visual representation of the criteria used to support this assessment, expanded from a GAI map (Project Management Institute (PMI), 2023). Diagram attributes are Scale (i.e., for data processing) and Singularity (i.e., format difficulty for GAI). Project attributes are Risk and Cost-Benefit (of GAI usage). Task attributes are Complexity and Impact.

Diagram feasibility determines if a diagram is simple (green), moderate (orange) or complex (red). Project determines if GAI usage on a project is safe (green), intermediate (orange) or risky (red). While tasks are classed as minor (green), major (orange) or critical (red). Color coding simplifies assessment outcomes for communication with decision-makers.

For simple cases (i.e., simple datasets, safe projects for GAI, minor GAI tasks), conceptual design software providers already incorporate GAI models for product design generation (Byrne et al., 2025). Moderate cases require more custom solutions such as Drawer AI that generates documentation for electrical vectorized diagrams (i.e., moderate datasets, intermediate projects, major tasks). However, complex cases remain underserved (i.e., complex datasets, intermediate projects, critical tasks), making them the focus of this study, i.e., investigating optimal GAI implementation conditions for these cases. The first step in exploring this opportunity is Data Preparation, which is discussed in the next section.

**Figure 2: Use Case Assessment.**



### 2.2 Data Preparation (Digitization)

The main elements of SLDs are symbols, text, and lines. SLD digitization studies (Table 1) are analyzed and insights (Figure 3) show that symbols are the most common target component (50% of publications), followed by text 27% and lines 23%. Research focus is mainly driven by institutional demand for digitizing legacy documents (L. Yang et al., 2024b) or paper copies (Cao et al., 2025) for high-voltage power distribution facilities.

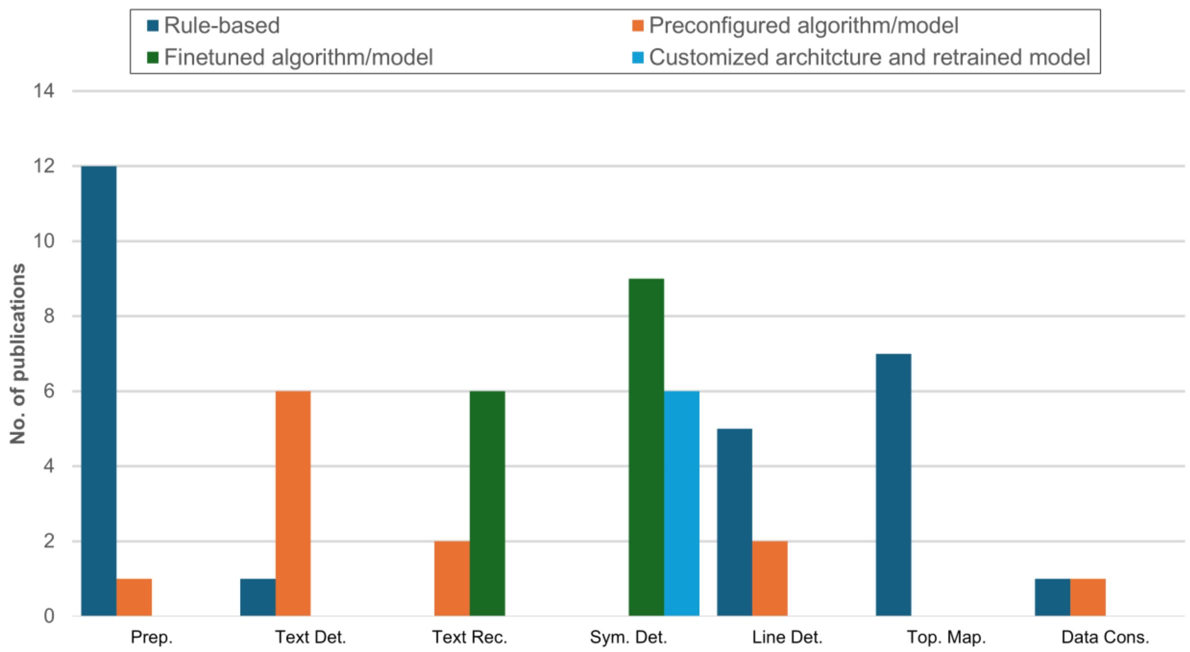
State-of-the-art SLD digitization methods are reviewed as a guide for firms seeking to digitize archived diagrams for training GAI models.

- Preprocessing: Techniques are applied according to dataset characteristics (i.e., data quality) and target component (symbols, lines or text). Selected techniques are mostly rule-based algorithms, usually grayscale conversion (for optimized symbol downstream processing), resizing (for symbol detection) and noise removal (for line detection).
- Text Detection: Identifying and isolating text regions in complex diagrams is a crucial decluttering step to enhance the subsequent detection of symbols and lines (Cao et al., 2025; Mao et al., 2023; C. Yang et al., 2023).
- Text Recognition: Information retrieval through Optical Character Recognition (OCR) for detected text regions is crucial for SLD understanding. Although many out-of-the-box options exist, researchers often customize existing solutions to improve accuracy due to SLD layout dissimilarity from OCR training data, viz., through fine-tuning or training (Chen et al., 2021; Mao et al., 2023; Shen et al., 2022; A. Song et al., 2021; C. Yang et al., 2023; L. Yang et al., 2024a) or augmenting with error checking algorithms (Cao et al., 2025).
- Symbol Detection: Transfer learning is used to adapt pretrained object detection algorithms to SLD symbols. You-Only-Look-Once (YOLO) is preferred for speed (used in 67% of the analyzed publications), followed by a Region-based Convolutional Neural Network i.e., Faster R-CNN (20%). Previous challenges such as insufficient data, poor image quality, small symbol-to-image size ratio, text-symbol overlap, class imbalance, etc., have largely been resolved using dataset augmentation with synthetic images (Bhanbhro et al., 2023a; L. Yang et al., 2024a), symbol size tolerant detection (Cao et al., 2025; L. Yang et al., 2024a), low-resolution image enhancement (L. Yang et al., 2024b), and text-symbol-line separation (L. Yang et al., 2024a). However, 100% detection (measured by accuracy or F1 Score) for all symbol classes is yet to be reported (see Table 1), hence, AI or human verification augments detection pipelines (Klinsrisuk & Witayangkurn, 2024).
- Line Detection: Straight line patterns algorithms are used to detect lines in SLDs. Masking is typically applied to reduce interference from graphic or text elements (Mao et al., 2023; L. Yang et al., 2024a).
- Topography Mapping: Symbol and line coordinates obtained from previous steps are combined to trace electricity flow through the SLD. Due to varying diagram styles, rule-based algorithms are more suitable (Cao et al., 2025; Li et al., 2021; A. Song et al., 2021; C. Yang et al., 2023; L. Yang et al., 2024a).
- Data Consolidation: A final step is to compile extracted information into a structured data format for language models. The translation of extracted information into structured data is less frequently discussed in literature. Custom algorithms (Mao et al., 2023) and open-source tools (Klinsrisuk & Witayangkurn, 2024) are used.

**Table 1: Digitization Single-line Diagram.**

| Authors | Dataset | Access Options | Symbol Classes | Detection Accuracy % | F1 Score % |
|---|---|---|---|---|---|
| (Dongxu et al., 2020) | Proprietary | NA | NA | NA | NA |
| (Chen et al., 2021) | Proprietary | NA | 11 | 90 | NA |
| (Rezaeva & Semendyaev, 2021) | Synthetic | NA | 28 | 92 | NA |
| (A. Song et al., 2021) | Synthetic | NA | 11 | 94 | 83 |
| (Shen et al., 2022) | Proprietary | NA | NA | 98 | NA |
| (Bhanbhro et al., 2023a) | Augmented | On request | 16 | 95 | 99 |
| (Bhanbhro et al., 2023b) | Proprietary | Open | 18 | 95 | 87 |
| (Bhanbhro et al., 2023c) | Augmented | NA | 22 | 96 | 91 |
| (C. Yang et al., 2023) | Synthetic | NA | 7 | 95 | NA |
| (Mao et al., 2023) | Proprietary | NA | NA | 95 | NA |
| (Klinsrisuk & Witayangkurn, 2024) | Augmented | NA | 37 | NA | 93 |
| (L. Yang et al., 2024a) | Augmented | Code | 8 | NA | 99 |
| (L. Yang et al., 2024b) | Augmented | Code | 8 | NA | 99 |
| (Cao et al., 2025) | Augmented | NA | NA | 99 | 99 |

**Figure 3: Comparison of Single Line Digitization Methods According to Approach Applied.**




Where Prep. is Preprocessing, Text Det. is Text Detection, Text Rec. is Text Recognition, Sym. Det. is Symbol Detection, Line Det. is Line Detection, Top. Map. is Topography Mapping, and Data Cons. is Data Consolidation.

### 2.3 Content Generation

From the use case assessment, project reports emerged as a significant case due to the potential benefits of GAI automation. These documents are redacted repeatedly, making them a priority. Reports are also based on diagrams and integrate various MLLM capabilities, including information retrieval, text generation, image comprehension, reasoning, and calculations. Prior studies evaluated MLLMs on image comprehension and information retrieval through Visual Question Answering (Abdul Razak et al., 2024; Doris et al., 2024; Geipel, 2024), as well as text generation and reasoning (Kunze & Fay, 2024). To the best of our knowledge, calculation performance is still unexplored in the engineering design context.

While the use case is challenging, Byrne et al. (2025) highlights the importance of these engineering cases to advance GAI research, noting incremental performance improvement over trials. Additional recommended techniques to enhance MLLM content generation accuracy include domain-specific training or finetuning (Abdul Razak et al., 2024; Byrne et al., 2025; Doris et al., 2024) and Retrieval Augmented Generation (RAG) (Doris et al., 2024).

Doris et al. (2024) also highlight limited resources for MLLM dataset preparation and model evaluation. Additionally, the context windows of MLLMs are small when compared to the length of project documentation (Kunze & Fay, 2024). Hence, exploratory human-centric evaluations can be used instead of metrics, combining techniques including few-shot-prompting (Kunze & Fay, 2024) and input data variation (Abdul Razak et al., 2024; Byrne et al., 2025; Kunze & Fay, 2024). Göpfert et al. (2024) also suggests using specialized software to augment models.

### 2.4 Automated Review

To enhance output validity and build user trust, content review can be automated by multi-agent systems (Arkoudas & Health, 2023) based on guidance documents (i.e., technical, safety, ethical, and sustainability). Researchers in engineering design also advocate this approach (Göpfert et al., 2024; B. Song et al., 2024).

### 2.5 Human Evaluation

Despite being resource-intensive, human verification remains crucial, especially for technical or sensitive domains (Arkoudas & Health, 2023). In a survey by PMI (2024), GAI response verification was identified as a crucial skill. Verification enables practitioners to improve their attention to detail, feedback mechanisms, and subject mastery. Discourse is also beneficial for models, particularly when conversational data is repurposed to enhance future responses.

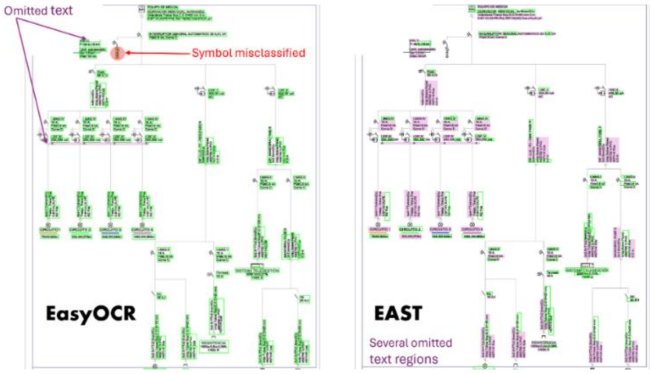
## 3. Case Study

This case study applies the proposed GAI-based methodology to a real-world dataset of low-voltage SLDs. The goal is to evaluate the effectiveness of current digitization techniques and GAI models in automating project documentation tasks. The dataset includes 40 complex SLDs sourced from public repositories, annotated and processed using open-source tools and state-of-the-art models.

### 3.1 Evaluation of Digitization Methods

<u>Text detection</u>: Out-of-the-box text detection tools (i.e., EasyOCR and EAST) are compared for complex images. Examples of detected text bounding boxes (in green) are shown in Figure 4. EasyOCR (with support for Spanish) outperforms on SLDs across sample images with fewer instances of omitted text and misclassification (highlighted for clarity).

**Figure 4: Text Detection Performance (Preconfigured Tools).**




Symbol detection: To evaluate the symbol detection performance, pretrained YOLO models are finetuned for a binary classification task. 40 real-world SLDs are collected from a public database, 477 symbols manually annotated using the LabelStudio annotation tool (available at https://labelstud.io) and exported to a data format suited to the YOLO models. Input image sizes averaged 1303 pixels by 1470 pixels with a maximum of 2841 pixels by 2692 pixels. 75% of the images were used to train the models with the remaining used to evaluate performance.

Model training setup: Training was done in the Google Colab environment, with a Tesla T4, 15102MiB Graphics Processing Unit, software versions were Python (3.10.12), Ultralytics (8.3.14), Torch (2.4.1) + cu121 CUDA:0. Table 2 shows metrics and Figure 5 shows results on a sample SLD.

**Table 2: Symbol Detection Results.**

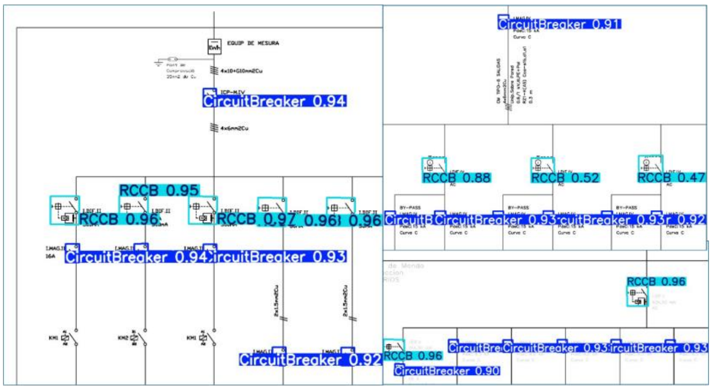
| Model | Circuit Breaker (CB) Class 0 | | | Residual Current CB (RCCB) Class 1 | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | F1-score | Precision | Recall | F1-score |
| yolo_v11n_detection (baseline) | **1** | 0.958 | 0.979 | 0.916 | 1 | 0.956 |
| yolo_v8n_obb | 0.946 | 0.999 | 0.972 | 0.950 | 1 | 0.974 |
| yolo_v11n_obb | 0.926 | 0.938 | 0.932 | **0.996** | 1 | **0.998** |
| yolo_v8s_obb | 0.979 | **1** | 0.989 | 0.981 | 1 | 0.990 |
| yolo_v11s_obb | 0.961 | **1** | 0.980 | 0.993 | 1 | 0.996 |
| yolo_v8l_obb | 0.999 | **1** | **0.999** | 0.991 | 1 | 0.995 |
| yolo_v11l_obb | **1** | 0.996 | 0.998 | 0.993 | 1 | 0.996 |

Note: n stands for *nano*, s for *small*, and l for *large,* while obb stands for *Oriented Bounding Boxes.*

Models performed according to symbol characteristics, for instance, average metrics for Class 0 (Circuit Breaker) are slightly lower (i.e., Precision: -0.13%, Recall: -1.58%, and F1-score: -0.82%) because of its smaller size and resemblance to other symbols. There are also more pronounced variations in its representation, making detection challenging on large complex images. Overall performance for both classes is acceptable, as symbols are amply represented. For rare classes, novel methods i.e., (Jamieson et al., 2024) are recommended.

Due to diverse symbol orientations, Oriented Bounding Boxes (OBB) demonstrated superior detection compared to the baseline. Lighter models show good performance, e.g., the nano size of v11 achieves the best performance for Class 1 while v8 small performs comparably with larger models across classes. Also, the newer YOLO model (i.e., v11) did not outperform its predecessor for Class 0, while it consistently improved performance on Class 1.

**Figure 5: YOLO Symbol Detection Performance (Binary Classification).**



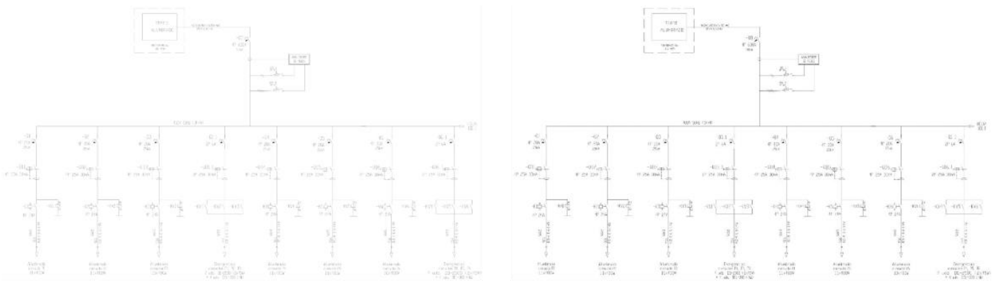

## 4.2 Evaluation of Content Generation

After data preparation (via digitization), state-of-the-art web chat-based MLLMs are evaluated using a Multimodal Information Retrieval and Generation task. A benchmark proposed by Geipel (2024) for industrial applications is adapted to define the task as shown in Table 3.

Input data used for evaluation (sample in Figure 6) was selected because of its complex attributes. I.e., the original PDF format does not contain extractable text, and when converted to an image, it requires preprocessing to enhance its readability for OCR. Despite image enhancement, proximity between text and symbols is challenging for preconfigured OCR tools. Some symbol class images were not found online, challenging MLLMs trained on web data.

Experimental setup is designed to check spatial understanding on real-world complex diagrams before comparing performance when varying levels of supporting data are provided.

**Figure 6: SLD Sample Image for MLLM Assessment.**



Original image Enhanced image

<u>Task description</u>: The given task is to calculate percentage voltage drop (for lighting cables). Voltage drop calculations across circuits are common to electrical projects and their compliance is verifiable via publicly available electrical regulations "*Reglamento electrotécnico para baja tensión*" (i.e., "Electrotechnical regulations for low voltage" in English).

Voltage drop is calculated using Equation 1 and 2 for single-phase and three-phase cables respectively. Where ΔV is the voltage drop in Volts, I is the current in the cable in Amperes, L is the cable length, R is the resistance per unit length in Ohms, X is the reactance per unit length in Ohms, while φ is the phase angle.

$$\Delta V = 2 \cdot I \cdot L \cdot (R \cdot cos(\varphi) + X \cdot sin(\varphi)) \quad (1)$$

$$\Delta V = \sqrt{3} \cdot I \cdot L \cdot (R \cdot cos(\varphi) + X \cdot sin(\varphi)) \quad (2)$$

MLLMs are tested under three conditions. For the Constrained case, MLLMs are given a complex diagram and a short prompt (see Table 3). MLLM evaluations by (Abdul Razak et al., 2024; Doris et al., 2024) on engineering diagrams showed that MLLMs struggle with fine details such as dimensioning. To account for this limitation, the diagram was partially simplified, i.e., extracted text using a commercial OCR service (Google Document AI) is included in a structured prompt (Table 3), this is termed the Midway case.

Fully digitized data was simulated by providing a hierarchy table in English showing component relationships and ratings (Unconstrained case). Hierarchy tables were quickly generated from diagrams using ChatGPT o3-mini model. ChatGPT generated a hierarchy table from the diagram image, OCR text from the Midway case, and a hierarchy table from another project.

An engineer with over five years of industry experience reviewed and corrected model output, identifying omitted components, misidentified equipment, incorrectly assigned specifications, and erroneous component relationships. The corrected table was then included in the same prompt as the Midway case with additional design assumptions provided (e.g., cable length).

Task evaluation criteria: MLLMs are compared qualitatively, a similar approach is utilized by Kunze & Fay (2024). Performance is measured by full, partial or unsatisfactory fulfilment of five criteria, i.e., prompt interpretation, design considerations, formula selection, arithmetic, calculation steps. For example, full prompt interpretation requires that all provided ratings are used for the correct cables. For the design considerations criteria, MLLM responses are checked for application of domain knowledge based on all the information given in the prompt. For formula selection, formulas in Equation 1 and 2 are applied correctly. For example, Equation 1 is used for single-phase cables only and not three-phase cables and vice versa. Similarly, the selected formulas are based on the given regulations, not general knowledge.

The emphasis of this evaluation is on MLLM reasoning processes rather than the final answer, scoring exact answers is more appropriate for extensive quantitative evaluations. Table 4 shows results of a comparison of MLLM calculation steps with detailed steps by an engineer from existing project documentation.

**Table 3: Task Description.**

| Aspect evaluated | Description |
|---|---|
| Task type | Multimodal reasoning and information extraction from SLD image |
| Input Data | Full-page electrical diagram (PDF converted to image) |
| Domain Knowledge Required | High – design knowledge, electrical regulations, and project interpretation |
| Reasoning steps | Multiple (identify components, extract specs, apply standards, calculate) |
| Prompt types | *Short prompt* (95 characters): What is the percentage voltage drop across the lighting cables in the given diagram? |
| | *Midway* (~2100 characters): With extracted OCR text. |
| | *Unconstrained* (~4200 characters): With hierarchy table and assumptions |
| Models evaluated | ChatGPT (o3 mini), Gemini 2.0, Claude Sonnet, Copilot |

The structured prompt used in the unconstrained case is given below for reference. This prompt simulates a realistic scenario where an engineer receives both diagram and extracted information (e.g., hierarchy tables, OCR text, and assumptions), and performs calculations following Spanish low-voltage electrical regulations. The goal was to assess the model's ability to extract relevant data, apply regulatory logic, and sequence calculations correctly.

> You're an electrical engineer and specialize in low-voltage electrical projects in Spain. You interpret electrical diagrams and calculate design parameters based on regulations provided at https://www.boe.es/eli/es/rd/2002/08/02/842/con
>
> What is the percentage voltage drop across the lighting cables in the given diagram?
>
> In your response, only state each cable and its corresponding voltage drop.
>
> It is recommended that you:
>
> 1. Determine the specifications of the lighting cables.
>
> 2. Consider the length, cross-sectional area, cable material, cable maximum temperature, phase, current, and power factor of the cables in your calculations.
>
> [Additional information] is provided as an aid. Account for [Additional information] errors and adjust where necessary. [Design assumptions given for Unconstrained case].
>
> Let me know if you need additional information.
>
> [Additional information]: OCR text or Hierarchy Table...

<u>MLLM strengths</u>: In the Midway & Unconstrained case, models sourced information that was not directly provided. E.g., models identified material types from specifications, even when not explicitly stated, and used this information to estimate resistivity values for calculations. Correct design considerations included recognizing subtle differences between equipment (e.g., single-phase vs. three-phase cables). Not all the models applied the information in their formula selection and calculations, leading to incorrect results.

Gemini selected the correct formulas to match its interpretation of the prompt (Unconstrained). Claude selects formulas only when design assumptions are correct. While Copilot, ChatGPT, and Claude selected generic formulas unsuited to the context. Only ChatGPT's o3-mini model requests more information when it is not provided (Midway case). MLLMs performed arithmetic, including Bing Copilot, mainly designed for browsing support.

<u>MLLM limitations</u>: Prompt interpretation is the most challenging task for MLLMs. None of the models comprehended complex diagrams unaided (i.e., Constrained). Models refrain from answering except Copilot which gives incorrect answers based on fabricated component specifications. Performance improved when given partial information (i.e., Midway) but with inconsistent performance due to incorrect assumptions. However, adding more information does not always give commensurate performance improvement (Unconstrained). Models missed details in longer prompts. For example, Claude uses an assumed voltage rating based on outdated regulations for one component while making the correct selection for another. This suggests that models may be overwhelmed by lengthy tables presented as unstructured text.

Copilot was more likely to be incorrect across trials, likely struggling to discern the veracity of web sources. Copilot is selected as a baseline to represent out-of-the-box models that do not have a reasoning mode. Overall, the evaluation showed promising results. However, current MLLMs are yet to achieve reliable results on design calculations and require finetuning or specialized software support.

**Table 4: MLLM Comparison (Unconstrained): Full (F), partial (P) or unsatisfactory (U) fulfilment.**

| MLLM Chat Tool (Mode or Model) | Prompt Interpretation | Design Considerations | Formula Selection | Arithmetic | Calculation Steps |
|---|---|---|---|---|---|
| Copilot | P | P | P | F | P |
| ChatGPT (Reason / o3 mini) | P | F | P | F | F |
| Gemini (2.0 Flash) | P | F | F | F | F |
| Claude (3.7 Sonnet) | P | F | P | F | F |

## 4 Discussion

This assessment is not considered a ranking because model performance varies between trials as observed by Abdul Razak et al. (2024). Evaluations show current limitations but are static while models constantly evolve. For example, experiments by Byrne et al. (2025) indicated that Gemini fails to match ChatGPT in arithmetic performance which later assessments do not support. And future assessments will likely contradict our study.

Model response also depends on testing conditions. E.g., performance improves in the Midway & Unconstrained case compared to the Constrained case, showing the benefits of tailored prompts and MLLM augmentation. Discourse also has a positive effect as responses improve when models are prompted for explanations or the reasoning behind design assumptions.

This work tested models using a challenging, domain-specific, multilingual task with varying levels of supporting information. Translating diagrams directly into text is not always effective and can confuse models. Information representation techniques such as Knowledge Graphs (KG) can be explored for representing diagram configurations for MLLMs in future work.

### 4.1 Practical Implementation Recommendations

To translate the findings of the case study into actionable strategies for industry adoption, Table 5 presents a set of practical recommendations based on six typical classes of electrical engineering documentation projects. These classes are derived from real-world use cases and classified along three dimensions. Recommendations in this work focus on technical implementation options for GAI in the context of engineering diagrams, with insights drawn from work by Dzhusupova et al. (2024) on AI implementation options for engineering firms, PMI (2023) and PMI (2024) for GAI implementation for general project management. The cases are determined by outcomes from the Use Case Assessment in Figure 2, i.e.,

- Dataset Attributes (Simple, Moderate, or Complex Dataset)
- Project Attributes (Safe, Intermediate, or Risky Project)
- Task Attributes (Minor, Major, or Critical Tasks).

Each class considers these factors and is associated with a set of suitable implementation approaches, such as Custom Platform (CP), Out-of-the-box models (MLLM), Finetuned or Trained model (MLLM+), Retrieval Augmented Generation (RAG), Specialized Software (SOFT), and Digitization (DX). As well as levels of human oversight, such as Low Human Involvement (LH) and High Upfront Human Involvement (HH). Options are now discussed.

<u>Case 1 (Simple Dataset, Safe Project for GAI, Minor GAI Task)</u>: Out-of-the-box MLLMs are sufficient for straightforward, low-risk use cases, depending on data privacy policy and

deployment cost. An example case is automated metadata generation for a large collection of digitized diagrams with clear text descriptions. Dzhusupova et al. also suggest off-the-shelf AI for cases that do not enhance a firm's competitive advantage.

Case 2 (Simple Dataset, Intermediate Project, Major Task): Case 2 requires RAG with increased human oversight. Examples are Visual Question Answering for compliance checking by Doris et al. or innovation options exploration. However, intellectual property (IP) rights on the ideas generated by human-GAI discourse should be clarified beforehand.

Case 3 (Moderate Dataset, Intermediate Project, Major Task): A firm can build its own MLLM-based solution (MLLM+ in Table 5) if its project type is business-critical, not covered by current providers or extra options such as multilingual support are required. Digitization tools are optional but useful for integrating legacy formats.

Case 4 (Complex Dataset, Intermediate Project, Critical Task): Mid-complexity tasks on image-based diagrams benefit from digitization pipelines. While mature, these tools can be costly, so collaborations with academia or open-source approaches are recommended. Critical documentation generation from images demands the full stack: MLLM+, RAG for information retrieval, SOFT for calculation logic, and DX for input processing. Initially, human involvement will be substantial until reliability is established.

Case 5 (Complex Dataset, Risky Project, Critical Task): Successful GAI automation on the previous cases frees up engineer's time for this case. Thus, current workflows that rely on human input, supported by software and diagram digitization can be maintained. Although MLLMs are expected to support risky projects in the future, successful implementation at this scale is feasible primarily through coordinated, cross-sector partnerships.

**Table 5: GAI Recommendations for Electrical Project Diagrams.**

| Project Type | | | | GAI implementation Options | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Case** | **Dataset Attributes** | **Project Attributes** | **Task Attributes** | **CP** | **MLLM** | **MLLM+** | **RAG** | **SOFT** | **DX** | **LH** | **HH** |
| 1 | Simple | Safe | Minor | | ✓ | | | | | ✓ | |
| 2 | Simple | Intermediate | Major | | ✓ | | ✓ | | | | ✓ |
| 3 | Moderate | Intermediate | Major | ✓ | | ✓ | ✓ | | | | ✓ |
| 4 | Complex | Intermediate | Critical | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ |
| 5 | Complex | Risky | Critical | | | | | ✓ | ✓ | | ✓ |

Note: CP stands for *Custom Platform*, MLLM for *Out-of-the-box Multimodal Large Language Model (MLLM)*, MLLM+ for *Finetuned or Trained MLLM,* RAG for *Retrieval Augmented Generation,* SOFT for *Specialized Software,* DX for *Digitization,* LH for *Low Human Involvement,* and HH for *High Upfront Human Involvement.*

## 5. Conclusion

Documentation automation using generative artificial intelligence (GAI) is a key research area in engineering design. This study presents a scalable methodology for generating electrical project documents using MLLMs, with low-voltage single-line diagrams as a case study. It provides practitioners with a first-hand demonstration of MLLM capabilities on a real project, fostering interest in practical industrial applications and paving the way for more agile, secure, and automated engineering design.

Cutting-edge digitization methods are validated on a real-world dataset, showing strong text and symbol detection performance, despite minor limitations due to inherent diagram complexity. Symbol detection results highlight the importance of resource-efficient models tailored to dataset requirements.

Digitization results were then applied to MLLM evaluations on a calculation task for a real-world electrical project, requiring complex reasoning and adherence to current legislation. MLLMs, particularly models enhanced for reasoning, demonstrated promising performance in electrical design knowledge and clear calculation steps. The Gemini 2.0 Flash model achieved the best results across the evaluated criteria.

Diagram interpretation was suboptimal without digitization. Similarly, models tend to miss details when presented with unstructured tables, indicating the need for information retrieval support and continuous refinement on domain-specific problems to enhance generalization.

Finally, insights from the model evaluations are distilled into practical recommendations for GAI implementation to align with common industry use cases. Future research should focus on simplifying complex textual information from digitization via Knowledge Graphs with Retrieval-Augmented Generation (RAG), combined with fine-tuning or prompt engineering, to improve performance on complex engineering diagrams. Similarly, data-driven evaluations of MLLM performance, supported by metrics, are essential to fully convince practitioners of their reliability. This study establishes a foundation for further progress in automating documentation generation throughout diverse domains of engineering design.

## 6. References

Abdul Razak, A. N., Lim, M., Tomeo-Reyes, I., & Li, D. D. (2024). Exploring the Capabilities and Limitations of Generative AI in Providing Feedback on Engineering Drawings: A Case Study. *2024 World Engineering Education Forum - Global Engineering Deans Council, WEEF-GEDC 2024*. https://doi.org/10.1109/WEEF-GEDC63419.2024.10854935

Arkoudas, K., & Health, D. (2023). *GPT-4 Can't Reason*. https://arxiv.org/abs/2308.03762v2

Bhanbhro, H., Hooi, Y. K., Kusakunniran, W., & Amur, Z. H. (2023a). Symbol Detection in a Multi-class Dataset Based on Single Line Diagrams using Deep Learning Models. *International Journal of Advanced Computer Science and Applications*, *14*(8), 43–56. https://doi.org/10.14569/IJACSA.2023.0140806

Bhanbhro, H., Hooi, Y. K., Zakaria, M. N. Bin, Hassan, Z., & Pitafi, S. (2023b). Single Line Electrical Drawings (SLED): A Multiclass Dataset Benchmarked by Deep Neural Networks. *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET)*, 66–71. https://doi.org/10.1109/ICSET59111.2023.10295140

Bhanbhro, H., Kwang Hooi, Y., Kusakunniran, W., & Amur, Z. H. (2023c). A Symbol Recognition System for Single-Line Diagrams Developed Using a Deep-Learning Approach. *Applied Sciences 2023, Vol. 13, Page 8816*, *13*(15), 8816. https://doi.org/10.3390/APP13158816

Byrne, D., Hargaden, V., & Papakostas, N. (2025). Application of generative AI technologies to engineering design. Procedia CIRP, 132, 147–152. https://doi.org/10.1016/J.PROCIR.2025.01.025

Cao, W., Chen, Z., Wu, C., & Li, T. (2025). A Layered Framework for Universal Extraction and Recognition of Electrical Diagrams. *Electronics*, *14*(5), 833. https://doi.org/10.3390/electronics14050833

Chen, Y., Jiang, W., Wang, Y., Hu, J., Guan, L., & Zhu, Z. (2021). Research on Deep Learning-based AI Information Extraction Methods of Substation Engineering Design. *2021 4th*

*International Conference on Energy, Electrical and Power Engineering (CEEPE)*, 971–976. https://doi.org/10.1109/CEEPE51765.2021.9475686

Chiarello, F., Barandoni, S., Škec, M. M., & Fantoni, G. (2024). Generative large language models in engineering design: opportunities and challenges. *Proceedings of the Design Society*, *4*, 1959–1968. https://doi.org/10.1017/PDS.2024.198

Dongxu, Z., Baohong, G., Rui, B., & Yonggang, F. (2020). Research on the Analysis and Check of Electrical Secondary PDF Drawings Based on Deep Learning. *2020 5th International Conference on Power and Renewable Energy (ICPRE)*, 507–511. https://doi.org/10.1109/ICPRE51194.2020.9233302

Doris, A. C., Grandi, D., Tomich, R., Alam, M. F., Cheong, H., & Ahmed, F. (2024). DesignQA: A Multimodal Benchmark for Evaluating Large Language Models' Understanding of Engineering Documentation. Journal of Computing and Information Science in Engineering, 25(2). https://doi.org/10.1115/1.4067333/1210215

Dzhusupova, R., Bosch, J., & Olsson, H. H. (2024). Choosing the right path for AI integration in engineering companies: A strategic guide. *Journal of Systems and Software*, *210*, 111945. https://doi.org/10.1016/J.JSS.2023.111945

Geipel, M. M. (2024). Towards a Benchmark of Multimodal Large Language Models for Industrial Engineering. *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–4. https://doi.org/10.1109/ETFA61755.2024.10711022

Göpfert, J., Weinand, J. M., Kuckertz, P., & Stolten, D. (2024). Opportunities for large language models and discourse in engineering design. *Energy and AI*, *17*, 100383. https://doi.org/10.1016/j.egyai.2024.100383

Jamieson, L., Elyan, E., & Moreno-García, C. F. (2024). Few-Shot Symbol Detection in Engineering Drawings. *Applied Artificial Intelligence*, *38*(1), 1–19. https://doi.org/10.1080/08839514.2024.2406712

Klinsrisuk, C., & Witayangkurn, A. (2024). ASDR: Automatic System to Diagnose and Recognize Electrical Drawings. *Proceedings - 21st International Joint Conference on Computer Science and Software Engineering, JCSSE 2024*, 118–125. https://doi.org/10.1109/JCSSE61278.2024.10613721

Kunze, F. C., & Fay, A. (2024). Automated Generation of AML Models for Industrial Plants Using LLM Chat Applications. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*. https://doi.org/10.1109/ETFA61755.2024.10710654

Li, H., Guan, T., Wang, S., Shi, W., Liu, Z., & Liu, X. (2021). Topological relation detection technology of substation wiring diagram in electric power system. *Beijing Hangkong Hangtian Daxue Xuebao/Journal of Beijing University of Aeronautics and Astronautics*, *47*(3), 531–538. https://doi.org/10.13700/J.BH.1001-5965.2020.0476

Mao, M., Zhao, K., Li, M., Xu, H., Yue, W., & Huang, K. (2023). Research on Substation Diagram Automatic Generation Based on Artificial Intelligence Algorithms. *2023 4th International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2023*, 219–225. https://doi.org/10.1109/ICHCI58871.2023.10277966

OpenAI. (2024). *Learning to reason with LLMs | OpenAI*. https://openai.com/index/learning-to-reason-with-llms/

Poccia, D. (2024). *Introducing Amazon Nova foundation models: Frontier intelligence and industry leading price performance | AWS News Blog*. Amazon Web Services.

https://aws.amazon.com/blogs/aws/introducing-amazon-nova-frontier-intelligence-and-industry-leading-price-performance/

Project Management Institute. (2023). Shaping the Future of Project Management With AI | PMI. https://www.pmi.org/learning/thought-leadership/ai-impact/shaping-the-future-of-project-management-with-ai

Project Management Institute. (2024). Benefits of Adopting Generative AI for Project Management | PMI. In *Project Management Institute*. https://www.pmi.org/learning/thought-leadership/benefits-of-ai-for-project-management

Rezaeva, M. A., & Semendyaev, R. Y. (2021). Development and Application of Convolutional Neural Network for the Recognition of Objects in the Scheme of Electric Grid. *Journal of Physics: Conference Series*, *2096*(1), 012020. https://doi.org/10.1088/1742-6596/2096/1/012020

Shen, C., Lv, P., Mao, M., Li, W., Zhao, K., & Yan, Z. (2022). Substation One-Line Diagram Automatic Generation Based On Image Recongnition. *2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT)*, 247–251. https://doi.org/10.1109/GCRAIT55928.2022.00059

Song, A., Kun, H., Peng, B., Chen, R., Zhao, K., Qiu, J., & Wang, K. (2021). EDRS: an Automatic System to Recognize Electrical Drawings. *2021 China Automation Congress (CAC)*, 5438–5443. https://doi.org/10.1109/CAC53003.2021.9728054

Song, B., Zhou, R., & Ahmed, F. (2024). Multi-Modal Machine Learning in Engineering Design: A Review and Future Directions. *Journal of Computing and Information Science in Engineering*, *24*(1). https://doi.org/10.1115/1.4063954

Yang, C., Wang, J., Yang, L., Shi, D., & Duan, X. (2023). Intelligent Digitization of Substation One-Line Diagrams Based on Computer Vision. *IEEE Transactions on Power Delivery*, *38*(6), 3912–3923. https://doi.org/10.1109/TPWRD.2023.3290945

Yang, L., Wang, J., Zhang, J., Li, H., Wang, K., Yang, C., & Shi, D. (2024a). Practical single-line diagram recognition based on digital image processing and deep vision models. *Expert Systems with Applications*, *238*, 122389. https://doi.org/10.1016/j.eswa.2023.122389

Yang, L., Zhang, J., Li, H., Ren, L., Yang, C., Wang, J., & Shi, D. (2024b). A comprehensive end-to-end computer vision framework for restoration and recognition of low-quality engineering drawings. *Engineering Applications of Artificial Intelligence*, *133*, 108524. https://doi.org/10.1016/j.engappai.2024.108524

## Use of Generative Artificial Intelligence

Generative Artificial Intelligence is used for hierarchy table preparation (with human oversight) and diagram comprehension experiments (i.e., ChatGPT, Copilot, Claude and Gemini).

## Communication aligned with the Sustainable Development Goals