International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Bridging the Visual Gap: Integrating Vision Language Models (VLM) and Artificial Intelligence (AI) with Enterprise Resource Planning (ERP) Software System

T Bharath Chandra

Senior Product Owner, Product Managers, Aptean India Pvt Ltd, Bengaluru, India
E-mail: bharathchandra@gmail.com, linkedin.com/in/21bharathchandra

Abstract **- Businesses generate vast visual data (e.g., quality check photos, warehouse snapshots, invoices, customer images), but traditional Enterprise Resource Planning (ERP) systems, built for structured data, cannot process it. This study explores integrating Vision Language Models (VLMs), AI combining computer vision and language processing, with ERPs to automate tasks like quality control, inventory monitoring, and document processing. We assess integration feasibility with Microsoft Dynamics 365 Business Central, Salesforce, and SAP S/4HANA, proposing an API-driven system architecture. VLMs face precision challenges, and ERP readiness varies: Microsoft Dynamics needs custom development, Salesforce offers flexible APIs, and SAP S/4HANA is robust but complex. Strategic planning and leveraging VLM strengths enable AI-enhanced enterprise systems.**

Keywords: Vision Language Models, ERP Integration, Enterprise AI, Computer Vision, Automation, API Integration, Microsoft Dynamics 365, Salesforce, SAP S/4HANA, Multimodal AI.

I. INTRODUCTION

Modern businesses produce diverse visual data (e.g., production images, security videos, document scans, technician photos), rich with insights. Traditional ERPs, designed for structured data (e.g., inventory, financials), cannot leverage this, creating a "visual data gap." Vision Language Models (VLMs), like GPT-4V, Gemini, Flamingo, LLaVA, and Qwen-VL, combine AI-driven vision and language to automate tasks like quality management and document processing. This paper evaluates VLM integration with Microsoft Dynamics 365 Business Central, Salesforce, and SAP S/4HANA, proposing an API-based architecture and addressing VLM precision limitations.

1.1 Background

Visual Data Challenge in ERPs: ERPs prioritize structured data but lack AI to analyze images, leading to manual inspections or missed insights. AI integration can bridge this gap.

VLM Architecture: VLMs feature a vision encoder (e.g., Vision Transformers), language model, and alignment mechanism, enabling tasks like object recognition and text extraction. Models like CLIP and GPT-4V support applications in retail, healthcare, automotive, and manufacturing.

1.2 ERP Platform Capabilities for VLM/AI Integration

- Microsoft Dynamics 365 Business Central: Cloud-based with APIs (OData) and Microsoft ecosystem (Azure AI, Power Platform). Complex visual tasks require custom AL extensions or iPaaS.
- Salesforce: Flexible data model and APIs (REST, SOAP) with Einstein AI and AppExchange. Strong for customer tasks, weaker for manufacturing.
- SAP S/4HANA: Deep functionality and APIs (OData, BAPIs) for manufacturing but complex, needing specialized tools for VLM integration.

1.3 Objectives

1. Assess VLM integration feasibility with three ERPs.
2. Compare API capabilities, flexibility, and security.
3. Propose a VLM-ERP architecture (Section 3.3).
4. Identify high-value use cases (Section 3.4).
5. Estimate integration complexity (Section 3.1, Results 4.1).
6. Provide adoption recommendations (Conclusion).

II. LITERATURE REVIEW

AI in ERPs boosts efficiency (40-45% time savings), decision accuracy (70%), and automation. Most research focuses on structured data or RPA, not visual data with VLMs. VLMs enable defect detection, inventory tracking, and document processing, reducing costs. API-based integration (point-to-point, middleware like MuleSoft) varies in scalability and cost. Challenges include data quality, security,

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

and VLM precision limitations. This study evaluates ERP readiness for VLM integration.

III. METHODOLOGY

This qualitative, comparative analysis assesses ERP platforms for VLM integration, focusing on AI. It covers data types, preprocessing, architectures, interaction models, and metrics, applied via test scenarios.

3.1 Conceptual Data Sources

The "dataset" is dynamic visual data: product images (e.g., defect detection), warehouse visuals (e.g., pallets), scanned documents (e.g., invoices), and technician/customer photos. Varied resolution and lighting challenge AI consistency.

3.2 Data Preprocessing

- Resizing/Formatting: Adjust images (e.g., 224x224 pixels) and videos (frame extraction) to JPEG, PNG, MP4.
- Normalization: Pixel values to [0, 1].
- Quality Enhancement: Address lighting/blurriness (conceptual).
- ROI Extraction: Crop relevant areas.
- Augmentation Context: Considers real-world input variability. Managed by an Integration Layer (Section 3.3).

3.3 Architectures

1. VLM Architecture: Vision encoder, language model, alignment mechanism; pre-trained, black-box APIs.
2. VLM-ERP Architecture: Includes Visual Data Source (images/videos), ERP (three platforms), Integration Layer (API Gateway, preprocessing, data mapping), and VLM Service (cloud/on-premises REST API).

3.4 VLM Interaction Model

- API-Centric: RESTful API calls from ERP via Integration Layer.
- Input: Visual data and task prompts (e.g., classification).
- Output: JSON with AI insights (e.g., extracted text).
- Pre-trained: Minimal fine-tuning; focus on reliable API contracts.

3.5 Evaluation Metrics

Conceptual metrics include:

- Accuracy: Defect detection (>95%), inventory counting (>90%), document extraction (>98%), shipment verification (>95%).
- Efficiency: Processing time (<2-5s) reduced manual effort (>80%).
- Improvements: Error reduction (>40-90%).
- Feasibility: ERP API and flexibility (Section 3.1).

Applied in Section 3.6 scenarios; readiness discussed in Section 4.

3.6 Conceptual User Scenarios

Four AI-driven use cases—quality inspection, inventory monitoring, document processing, warehouse operations are analysed. Each details objectives, VLM-ERP workflows, and success metrics to evaluate feasibility.

i. VLM-Powered Quality Inspection with AI

- Objective: Automate quality checks for products like fruit baskets using AI to reduce manual effort and ensure consistent defect detection.
- Scenario: A camera captures an image or video of a "Tropical Fruit Basket (Organic)" containing assorted fruits on a production line. The ERP system logs a quality check to evaluate defects based on the criterion "BRUISES/MARKS VISIBLE on the fruit," with a threshold of 1–4 bruises per basket (sample size: 1).
- Workflow:
  - The ERP system device captures the fruit basket and sends it via a REST API to a VLM endpoint (e.g., Cloud-AI model like CLIP).
  - The VLM, as an AI system, analyses the image to detect and count bruises/marks, classifying the basket as "Pass" If this is less than 4 bruises, "Fail " more than 4 bruises found, or "Fail - Incorrect Fruit" (e.g., missing mangoes).
  - The VLM returns structured data (JSON format: {"status": "Pass", "bruises": 2}) to the ERP.
  - The ERP updates the Quality Order record for Item No. 0038, storing the VLM's AI-generated result in the report "Posted Quality Check List."
  - If the VLM detects more than 4 bruises, the ERP marks the "Blocked" field to halt processing and alerts the quality team.
- Success Metrics:
  - Accuracy of AI-driven bruise detection: > 95% compared to manual checks.
  - Processing time: > 12 seconds per image using the AI model.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

ii. VLM-Enhanced Inventory and Warehouse Management with AI

- Objective: Improve inventory accuracy and operational efficiency by monitoring stock levels and verifying item movements using AI.
- Scenario: A warehouse stores EURO pallets (Dimensions 1200x800 mm) and boxes labelled "Poly" on shelves (location A2-01). A camera captures a stack of 10 EURO pallets on the floor, a forklift moving 5 blue pallets to a shelf. But the ERP shows a negative inventory (-2 units) for Item No.0040.
- Workflow:
    - The warehouse camera captures images and sends them to the ERP, which forwards them via API to the VLM (AI service).
    - The VLM (AI) identifies and counts items (e.g., 10 EURO pallets, 12 Poly boxes), reads labels or barcodes to confirm details (e.g., "1200x800 mm" for pallets), and verifies that items moved by forklifts or loaded into containers matching ERP records.
    - The VLM returns structured AI-generated data (e.g., {"item": "EURO pallet", "count": 10, "location": "floor"}) to ERP software.
    - The ERP compares VLM counts to the Inventory tab (Item No. 0040) and flags discrepancies (e.g., 10 pallets detected by AI vs. -2 expected).
    - Discrepancies trigger cycle count adjustments as per the permission sets allowed or alerts for manual review.
- Success Metrics:
    - AI-assisted inventory count accuracy: >90% alignment with ERP records.
    - Verification time for shipments using AI: >5 seconds per image.
    - Reduction in cycle count errors through AI monitoring: >40%.

iii. VLM-Enhanced Document Processing with AI

Objective: Streamline data entry by automating the extraction of information from vendor invoices and warehouse documents using AI.

Scenario: A scanned Warehouse Receipt (dated 26-Dec-24) lists a shipment of 30 Quantity (Model ORB500M13-215) with serial numbers (e.g., OFB13N0N92211) and a barcode.

Workflow:

1. The VLM (AI system) receives the video or image document from Warehouse employee.
2. The VLM uses OCR and AI-powered layout analysis to extract fields like vendor name, invoice number, date, model number, quantity, and serial numbers, producing structured data (e.g., {"model": "ORB500M13-215", "quantity": 30, "date": "26-Dec-24", "serial": ["OFB13N0N92211", ...]}) ERP software will get notified via API.
3. The ERP Verifies fields in a new scanned Warehouse Receipt which matches the AI-extracted data to an existing Purchase Order, updating "Qty. on Purch. Order" for the corresponding item.
4. For warehouse documents, the VLM (AI) extracts data from receipts to verify and update inventory records in real-time during receiving operations.

- Success Metrics:
    - AI-based data extraction accuracy: >98% for key fields.
    - Processing time per document with AI: >3 seconds.
    - Reduction in manual data entry due to AI automation: >80%.

iv. VLM-Supported Warehouse Operations with AI

- Objective: Improve operational efficiency by verifying item movements and shipments in dynamic warehouse environments using AI visual analysis.
- Scenario: A warehouse handles EURO pallets and Poly boxes using equipment like forklifts and pallet jacks. A camera captures a worker loading 12 Poly boxes into a shipping container, a forklift moving 5 blue pallets to a shelf, and organized shelves with labeled items. The ERP tracks outgoing shipment and inventory movements for these items.
- Workflow:
    - The warehouse camera captures images or videos of operations and sends them to the ERP, which forwards them via API to the VLM (AI service).
    - The VLM (AI) identifies items and their quantities (e.g., 12 Poly boxes, 5 blue pallets), reads labels or barcodes, and verifies that items moved or loaded match ERP records.
    - The VLM returns structured AI-generated data (e.g., {"item": "Poly box", "count": 12, "action": "loading"} or {"item": "blue pallet", "count": 5, "action": "moving"}) to the ERP.
    - The ERP cross-checks this large-scale data with its records (e.g., confirming 12 Poly boxes match an outgoing sales order) and updates fields like "Qty. on Sales Order" or "Qty. on Component Lines."
    - For quality assurance, the ERP ensures items like the Tropical Fruit Basket (Item No. 0038) have passed

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

AI-assisted inspections (e.g., POC-CHECK003) before shipment.

- Success Metrics:
  - AI verification accuracy for shipments: >95% alignment with ERP orders.
  - AI processing time: >5 seconds per image or videoframe.
  - Reduction in shipment errors thanks to AI verification: >90%.

Summary (for 3.6 Conceptual User Scenarios):

AI-driven Vision Language Models (VLMs) enhance ERP systems by automating quality inspections, improving inventory count accuracy, accelerating document processing, and verifying warehouse operations. Future work should focus on real-world testing of these AI integrations to assess costs and benefits.

IV. RESULT

This section assesses the feasibility and anticipated outcomes of integrating Vision Language Models (VLMs) with Microsoft Dynamics 365 Business Central, Sales force, and SAP S/4HANA, based on the methodology, proposed architecture (Section 3.3), and conceptual test scenarios (Section3.6).

4.1 Platform Integration Readiness and Feasibility Assessment

Using the evaluation framework (Section 3.1), proposed architecture (Section 3.3), and test scenarios (Section 3.6), our analysis highlights varying readiness and complexity for integrating AI-powered VLMs with target ERPs. "Model Performance Metrics" refers to the integrated VLM-ERP solution’s ability to meet scenario-specific success metrics (Section 3.5, detailed in 3.6), not the intrinsic performance of pre-trained VLMs.

Table 1: Conceptual VLM/AI Integration Readiness: Microsoft Dynamics 365 Business Central

Feature	Assessment	Details
API Score (1-5)	3.7	Solid core APIs (OData, Custom)
Data Model Flexibility	Limited (Requires Ext.)	Customization needed via AL Extensions
AI Features	Basic (Azure Focus)	Relies heavily on external Azure AI services
Extension Mechanisms	Medium (AL, Power Plat.)	Good tools, but complex logic needs AL dev.
Est. Implementation Effort	Very High	For deep/complex visual AI integration
Key Strengths	Modern Architecture	Cloud-native, leverages Microsoft AI Ecosystem
Key Challenges	Customization Needs	Requires dev./middleware for advanced visual AI

Table 2: Conceptual VLM/AI Integration Readiness: Sales force Software

Feature	Assessment	Details
API Score (1-5)	4.8	Extensive, well-documented (REST, SOAP, Bulk etc.)
Data Model Flexibility	High	Easily handles custom objects/fields for AI data
AI Features	Extensive (Einstein AI)	Strong native AI capabilities
Extension Mechanisms	High (Flow, Apex)	Powerful low-code and pro-code options
Est. Implementation Effort	Moderate	Flexibility reduces integration complexity
Key Strengths	Flexibility, APIs, AI	Highly adaptable platform, strong ecosystem
Key Challenges	Core Process Depth	Less native depth in Mfg./SCM vs. SAP

Table 3: Conceptual VLM/AI Integration Readiness: SAP S/4HANA

Feature	Assessment	Details
API Score (1-5)	4.2	Robust standard APIs (OData, BAPIs) for core ops.
Data Model Flexibility	Medium	Can be extended, but often complex
AI Features	Emerging (SAP AI)	Growing native capabilities, BTP integration
Extension Mechanisms	Medium (ABAP, BTP)	Powerful but requires specialized skills
Est. Implementation Effort	High	Due to platform complexity and rigidity
Key Strengths	DeepProcess Integration.	Tightly couples AI with core SAP processes
Key Challenges	Complexity, Learning Curve	Steep learning curve, less agile integration

V. RESULTS AND DISCUSSION

Our comparative analysis, using the evaluation framework (Section 3.1), proposed architecture (Section 3.3), and user scenarios (Section 3.6), reveals varying readiness for integrating AI-powered Vision Language Models (VLMs) with Microsoft Dynamics 365 Business Central, Salesforce, and SAP S/4HANA. "Model Performance Metrics" refers to

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

the VLM-ERP solution's ability to meet scenario-specific success metrics (Section 3.5, detailed in 3.6), not VLM intrinsic performance.

Microsoft Dynamics 365 Business Central:

- **Strengths:**
  Cloud-native, scalable, with Microsoft's AI ecosystem (Azure AI, Power Platform). Power Apps/Automate enable low-code AI workflows.
- **Challenges:**
  Complex AI tasks (e.g., defect detection) need custom AL extensions or iPaaS, increasing cost and complexity. Effort: Very High.
- **Success Metrics:**
  Metrics (>95% defect detection, <2s processing) rely on custom development and Azure AI. Simpler tasks are moderately feasible.

Salesforce:

- **Strengths:**
  Flexible data model and APIs (REST, SOAP) with Apex, Flow, and Einstein AI support VLM outputs. Ideal for customer tasks.
- **Challenges:**
  Limited for manufacturing/inventory, needing extra logic or ERP integration. Effort: Moderate.
- **Success Metrics:**
  Strong for metrics (>98% document extraction) in customer or custom apps, leveraging automation.

SAP S/4HANA:

- **Strengths:**
- Deep manufacturing/logistics functionality with APIs (OData, BAPIs) and BTP. SAP AI enhances integration.
  Challenges:
- Complex models require ABAP/BTP expertise, with resource-intensive custom APIs. Effort: High.
  Success Metrics:
- High potential for core process metrics, but long development cycles raise costs.

VLM & AI Limitations: VLMs excel in document extraction but struggle with precision tasks. Solutions need human verification and error handling.

VI. CONCLUSION

VLM-ERP integration unlocks visual data insights, automating tasks and boosting efficiency. Integration is feasible but varies:

Microsoft Dynamics: Strong AI ecosystem, high custom effort.

Salesforce: Flexible for customer tasks.

SAP S/4HANA: Deep but complex integration.

Success requires robust architecture (Section 3.3), leveraging VLM strengths, and mitigating limitations via middleware.

Future Directions:

- Hybrid AI for precision.
- VLM fine-tuning for ERPs.
- Edge AI for real-time processing.
- Standardized ERP interfaces.
- Explainability (XAI) for trust.

Despite challenges, VLM-ERP integration promises smarter, automated systems, enhancing decision-making.

REFERENCES

[1] Daffodil Software. (2025, March 3). All You Need To Know About Vision Language Models. https://insights.daffodilsw.com/blog/all-you-need-to-know-about-vision-language-models
[2] IBM. (2025, May 6). What Are Vision Language Models (VLMs)? https://www.ibm.com/think/topics/vision-language-models
[3] Alayrac, J.B., et al. (2022). Flamingo: a Visual Language Model. arXiv:2204.14198. https://arxiv.org/abs/2204.14198
[4] Bai, J., et al. (2023). Qwen-VL: A Versatile Vision-Language Model. arXiv: teamwork2308.12966. https://arxiv.org/abs/2308.12966
[5] Liu, H., et al. (2023). Visual Instruction Tuning (LLaVA). arXiv:2304.08485. https://arxiv.org/abs/2304.08485
[6] Mhaskey, S.V. (2024, December). AI in ERP Systems: Opportunities and Challenges. Int. J. Comp. Eng. Res. Trends, 11(12), 1-9.https://www.ijcert.org/index.php/ijcert/article/view/1036

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 200-205, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202532
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Citation of this Article:

T Bharath Chandra. (2025). Bridging the Visual Gap: Integrating Vision Language Models (VLM) and Artificial Intelligence (AI) with Enterprise Resource Planning (ERP) Software System. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 200-205. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202532

****