This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-032-30044-7_1. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

AI-powered Enterprise System for CRM and Inventory Management in Low-Digital MSMEs

Zachary Matthew Alabastro, Beatrice Anne Lui, Joseph Benjamin Ilagan, Jose Ramon Ilagan, and Gabrielle Anne Uy

Business Insights Laboratory for Development, Ateneo de Manila University, Quezon City, Philippines
{zachary.alabastro,beatrice.lui}@student.ateneo.edu
{jbilagan,jrilagan,gluy}@ateneo.edu ,
WWW home page: https://www.ateneo.edu/

Abstract. Many micro, small, and medium enterprises in developing economies continue to rely on handwritten records and informal chat-based transactions, limiting their ability to analyze data and make informed decisions. This paper presents the design and evaluation of an AI-powered enterprise system intended for MSMEs operating in low-digital contexts. The system integrates optical character recognition, large language models, entity resolution, and multilingual conversational interfaces to transform handwritten and chat-based business records into structured, analytics-ready data. Field deployments with Philippine MSMEs demonstrate how natural language querying and human-in-the-loop validation lower barriers to adoption while supporting trust, usability, and explainability. The study contributes to HCI research by showing how hybrid human–AI architectures can democratize data-driven decision-making for small businesses in emerging markets.

Keywords: ERP · MSME · Generative AI · GenAI · OCR · HITL · LLM

1 Introduction

Micro, small, and medium enterprises (MSMEs) in developing economies frequently rely on handwritten ledgers, informal spreadsheets, and chat-based transactions for daily operations. In the Philippines and similar contexts, these practices persist due to cost constraints, limited access to dedicated IT support, and the practical dominance of mobile messaging platforms for coordination. While enterprise resource planning (ERP) systems can improve operational control, inventory management, and reporting accuracy, conventional solutions often remain inaccessible to MSMEs due to licensing costs, implementation complexity, training requirements, and limited localization [8,9,15].

Recent advances in multimodal AI and large language models (LLMs) create opportunities for incremental digitization that does not require MSMEs to abandon existing workflows. Rather than assuming fully structured digital data entry,

AI-assisted pipelines can treat handwritten documents, receipts, and conversational orders as primary inputs. Optical character recognition (OCR) can extract text from images, while LLMs can enforce schemas, infer structure, and support multilingual interaction. However, deploying such systems raises important Human–Computer Interaction (HCI) questions concerning trust, usability, explainability, and accountability, particularly in low-digital-literacy environments.

Human-in-the-loop (HITL) frameworks offer one approach to balancing automation efficiency with user oversight. HITL systems emphasize verification, correction, and traceability to prevent the silent propagation of model errors [21]. In MSME contexts, this is especially relevant because inaccuracies in sales or inventory records can accumulate into operational and financial misjudgments. Hybrid architectures that combine automated extraction with user validation may therefore be more appropriate than fully autonomous systems.

This study presents the design and evaluation of an AI-powered enterprise system intended for MSMEs operating in low-digital contexts. The proposed architecture integrates OCR, LLM-based structured extraction and validation, entity resolution, multilingual conversational interfaces, and natural language querying into a unified pipeline. A text-to-SQL agent enables business owners to retrieve analytics using everyday language, reducing technical barriers to data access. Across modules, outputs adhere to a unified schema to support interoperability and traceability.

1.1 Objectives and Research Questions

The primary objectives of this study are to develop frameworks and techniques that simplify and augment manual business processes for data collection while retaining a human-in-the-loop approach. In parallel, the study seeks to extend existing, informal digital use cases by integrating AI in a more interpretable and accessible form, one that can coexist with, rather than disrupt, established MSME workflows. Collectively, these objectives motivate the following research questions:

RQ1) How can OCR–LLM pipelines transform handwritten and chat-based business records into structured, analytics-ready data for MSMEs?

RQ2) To what extent can multilingual conversational interfaces enhance MSME engagement with enterprise systems and encourage data-driven decision-making?

RQ3) What UX design principles support trust, usability, and sustained adoption of AI-driven enterprise tools in low-digital-literacy contexts?

1.2 Contribution

This study makes three primary contributions. First, it proposes a high-level architectural framework for an AI-augmented enterprise system designed to integrate OCR, LLMs, and HITL workflows for MSMEs. Second, it introduces empirical measures of accuracy and effectiveness to evaluate the performance of

the proposed OCR–LLM pipeline using representative sample tools and datasets. Third, it extends applied OCR research by demonstrating how handwritten and conversational business records can be operationalized for end-to-end enterprise integration, enabling AI-assisted digitization that remains compatible with predominantly manual MSME workflows [4].

2 Review of Related Work

2.1 ERP Adoption Barriers in MSMEs

ERP systems are widely recognized for improving coordination, operational visibility, and decision support in organizations. However, research consistently shows that ERP adoption is shaped not only by technical suitability but also by organizational readiness, resource constraints, and implementation complexity. For MSMEs, these constraints are amplified due to limited financial slack, limited IT expertise, and weaker process formalization [8].

Studies on ERP adoption in small businesses identify recurring barriers: high upfront and recurring costs, training demands, misalignment with informal workflows, and poor localization to language and context [9,15]. Even when lightweight ERP systems are available, adoption often fails when businesses cannot sustain data entry discipline, maintain data quality, or allocate personnel for ongoing administration. These issues suggest that MSME-oriented enterprise systems should prioritize low-friction data capture, incremental adoption, and support for semi-structured or unstructured inputs rather than assuming stable structured transaction encoding.

2.2 Digital Inclusion and Low-Digital Contexts

Digital inclusion research emphasizes that adoption of information systems in low-resource environments is mediated by access, skills, trust, and contextual fit. In many developing contexts, digital systems fail not because the benefits are unclear, but because they impose interaction patterns that are incompatible with existing work practices. MSMEs often operate with minimal formal processes and rely on social coordination rather than standardized digital workflows.

The Technology Acceptance Model (TAM) and its extensions provide a useful baseline for interpreting MSME adoption dynamics. Perceived usefulness and perceived ease of use remain strong predictors of adoption intention [17]. However, in low-digital contexts, ease of use must be interpreted not only as interface simplicity but also as the ability to align with users' mental models, literacy levels, and existing tools. This motivates design approaches that embed enterprise functionality into familiar interaction modes such as chat interfaces, image capture, and informal language input.

2.3 OCR for Handwritten and Semi-Structured Business Records

OCR remains a foundational technology for digitizing physical documents. While OCR for printed text is mature, handwritten OCR remains challenging due to

variability in handwriting, layout, and noise. Traditional OCR pipelines often assume relatively consistent document structure and high-quality scans, assumptions that do not hold in MSME settings where inputs may be captured using mobile cameras under poor lighting and informal formatting.

Recent applied research has explored combining OCR with downstream post-processing to improve robustness in real-world deployments, including the extraction of tables and receipts from semi-structured inputs [4]. In MSME contexts, a key requirement is not only transcription but also the transformation of extracted text into structured records that can be integrated into inventory and sales databases. This motivates architectures where OCR is paired with schema-aware extraction, entity linking, and validation.

2.4 Large Language Models for Information Extraction and Data Structuring

Large language models have demonstrated strong performance in information extraction, schema completion, and structured output generation. Unlike traditional rule-based or supervised extraction pipelines, LLMs can generalize across varied formats and can be guided using prompts and few-shot examples. This makes them attractive for low-resource enterprise settings where annotated datasets are unavailable and where document formats vary widely.

However, LLM-based extraction also introduces risks. Models may hallucinate values, over-infer missing data, or output structurally invalid results. These limitations make it difficult to deploy LLMs as fully autonomous extraction engines in enterprise workflows. Human-in-the-loop approaches are therefore commonly proposed to preserve accountability while benefiting from automation efficiency. HITL systems emphasize user verification, error correction, and traceability, ensuring that automation does not silently propagate incorrect records [21]. In enterprise contexts, this is especially important because small inaccuracies in sales or inventory data can accumulate into significant decision errors.

2.5 Entity Resolution and Data Quality in Small Business Systems

Entity resolution is a core challenge in operational systems, particularly when data originates from noisy or informal sources. In MSMEs, product names may be abbreviated, misspelled, or described using local terms, and the same item may appear under multiple names across records. OCR noise further increases ambiguity. Entity resolution techniques such as string normalization, approximate matching, and canonical identifier mapping are therefore critical for enabling coherent inventory reporting and preventing duplication.

In small-scale deployments, entity resolution must also remain interpretable to users. Fully automated resolution may reduce manual workload, but opaque matching decisions can reduce trust. Confidence scoring and user-correctable mappings are common strategies for improving usability and accountability. This aligns with broader HCI findings that user trust increases when systems provide transparent feedback and allow user control over uncertain outputs.

2.6 Conversational Interfaces, Multilinguality, and Informal Commerce

Conversational interfaces have become a practical entry point for enterprise interactions in contexts where messaging applications dominate. In many developing economies, informal commerce is coordinated through chat, including orders, confirmations, and payment coordination. This makes conversational systems a natural interaction model for MSMEs.

LLMs have improved multilingual and code-switched natural language understanding, enabling systems to parse mixed-language inputs such as Taglish. Recent work in conversational commerce highlights that conversational systems must support not only language coverage but also domain grounding and ambiguity management [11]. Users often express quantities, products, and preferences informally, and systems must either request clarification or provide uncertainty-aware outputs. This is particularly important in low-digital contexts where users may have limited tolerance for repeated failures and may quickly revert to manual processes.

2.7 Natural Language Querying and Text-to-SQL for Non-Technical Users

Business analytics traditionally requires technical skills such as spreadsheet modeling, database querying, or BI dashboard configuration. Natural language interfaces aim to lower these barriers by allowing users to query data using everyday language. Text-to-SQL systems translate user questions into executable SQL queries, enabling access to structured datasets without requiring SQL expertise.

Recent advances have improved Text-to-SQL accuracy, but practical deployment remains constrained by schema complexity, ambiguity in user intent, and the need for explainability. In MSME settings, the key value of Text-to-SQL is not high-volume analytics but accessible reporting for routine decisions (e.g., sales totals, best-selling products, inventory gaps). This suggests that lightweight schemas, constrained query templates, and clear result explanations may be more appropriate than fully open-ended natural language querying.

2.8 Trust, Explainability, and Human-Centered AI in Enterprise Tools

Trust is a central determinant of sustained adoption in AI-assisted systems. In enterprise contexts, trust is shaped by perceived accuracy, transparency, controllability, and the perceived consequences of failure. Users are more likely to adopt AI tools when systems provide feedback, support correction, and avoid presenting uncertain outputs as definitive. Explainability is therefore not only a technical property but also a UX requirement, especially when users rely on outputs for operational decisions.

HCI research on applied AI systems emphasizes the importance of designing for user mental models, supporting graceful failure, and making uncertainty

visible [18]. For MSMEs, these concerns are amplified because users may have limited time for training and may interpret AI errors as a sign that the system is unreliable. Hybrid human–AI workflows, where AI performs extraction and users validate or correct, can improve both trust and adoption while still reducing workload.

2.9 Summary and Research Gap

Prior work establishes that MSME adoption of enterprise systems is constrained by cost, complexity, and contextual misfit. OCR and conversational interfaces offer practical pathways for digitizing informal records, while LLMs introduce new capabilities for schema-aware extraction, multilingual interaction, and natural language analytics. However, there remains a gap in end-to-end enterprise system designs that integrate handwritten OCR, chat-based order capture, entity resolution, and natural language querying into a unified architecture optimized for low-digital MSMEs. In addition, empirical HCI-oriented evaluation remains limited for such hybrid pipelines, particularly regarding trust, usability, and sustained adoption in multilingual emerging market settings.

The present study addresses this gap by proposing and evaluating an AI-powered enterprise system that treats handwritten and chat-based records as primary inputs, supports multilingual interaction, and incorporates human-in-the-loop validation to maintain accountability and data quality.

3 Methodology

The proposed system architecture consists of four primary processing pipelines integrated into a cohesive data management framework (see Fig. 1). First, the OCR pipeline processes image inputs using Gemini 3 Flash as the primary extraction engine, followed by Anthropic Sonnet 4.5 for few-shot learning and validation; the results are structured into a JSON dictionary and converted into tabulated summaries [1,2]. To facilitate a comprehensive investigation, no confidence threshold is applied, with precision and recall serving as the primary statistical metrics for accuracy and utility [5].

Second, a custom conversational interface enables simulated order processing in English, Tagalog, and Taglish. This module extracts product data into JSON-formatted sets, designed for future integration into platforms like Messenger or Telegram to automate ordering and production workflows.

Third, an entity resolution pipeline processes receipt images to extract information and generate a consolidated set of orders or expenses, effectively minimizing data duplicates through automated cross-referencing. Fourth, a text-to-SQL agent provides a natural language interface for non-technical users to query underlying databases, utilizing a ReAct reasoning model to output digestible data results via a pre-built simulation UI [20,7]. All pipeline outputs adhere to JSON-valid formatting to ensure seamless display within the user interface. Finally, a

human-in-the-loop layer allows for manual intervention, enabling real-time editing and revision to ensure the integrity of the processed data [12]. These modules build upon established research regarding large language models for multilingual natural language understanding in conversational commerce [11].

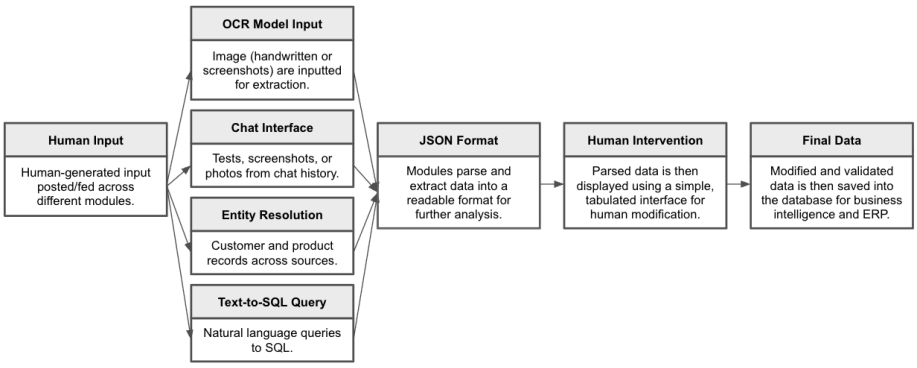


Fig. 1. Proposed methodological pipeline for modules.

3.1 Modular Development and Evaluation Methods

To operationalize the proposed AI-augmented ERP architecture, the study adopts a modular pipeline design in which each component is independently developed, validated, and evaluated while adhering to a shared data exchange schema. This design enables interoperability across modules while preserving traceability and HITL control. The following subsections describe the development and testing procedures for each module.

OCR–LLM Extraction Pipeline The OCR module is designed to convert handwritten sales records into structured, analytics-ready data while retaining human oversight. A predefined catalog of products and SKUs is supplied to the system as contextual grounding [19]. Handwritten inputs are processed using a multimodal large language model (Gemini) guided by structured extraction prompts [16]. The raw outputs are subsequently validated by a secondary language model (Anthropic), which performs cross-checking against known SKU–price pairs, enforces schema compliance, and converts the results into valid JSON [13]. The finalized outputs are then programmatically parsed and tabulated using Python.

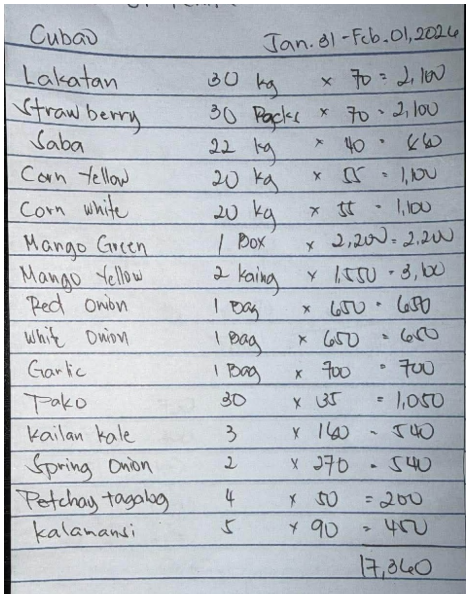


Fig. 2. OCR–LLM pipeline illustrating contextual grounding, multimodal extraction, validation, and structured data output.

Conversational Chat-Based Ordering The chat module enables MSME users to submit orders using informal, natural-language messages, reflecting SMS- or messaging-based interactions common in low-digital contexts [3]. Users provide free-text descriptions of intended purchases. The system parses these inputs to extract product entities, quantities, and pricing information, adhering to the same schema used by the OCR module. Validation steps ensure consistency across modalities and enable seamless downstream processing.

Table 1. Sample user messages used for chat-based order extraction.

User / Customer Message (Verbatim)
"boss pa order ulit ha for restock this week air filter 25pcs engine oil 4cs 1L 10w40 wiper blades 35 assorted spark plugs 80 engine coolant 2 case brake pads set 20sets motor batt 12pcs same address pa din. pa total nalang pag na compute mo"

Entity Resolution and Data Refinement Entity resolution addresses inconsistencies arising from OCR noise, informal naming conventions, or visual ambiguity [6,10]. Sample product images are provided as inputs to the resolution module. Python-based routines perform string normalization, SKU matching, and price correction by referencing the contextual product catalog. This process aligns extracted entities with canonical ERP identifiers and reduces duplication across records.

Table 2. Sample receipt line items with OCR noise and entity variations.

Row ID	Receipt ID	Quantity	Product Name	Subtotal
0	0001	1	PremiumUngr50ml	12.00
1	0002	2	TableSalt1kg	98.00
2	0002	1	PremiumVngr50ml	12.00
3	0003	1	PremiunVinegar50ml	12.00
4	0003	3	TblSalt1kg	147.00
5	0004	1	Premium Vinegar 50 ml	12.00
6	0004	2	Table Salt 1 kg	98.00
7	0005	1	PremVngr50ml	12.00
8	0006	1	TablSalt1kg	49.00
9	0006	2	PremiumVinegar50ML	24.00
10	0007	1	Premium Vinegr 50ml	12.00
11	0008	2	Table-Salt 1kg	98.00
12	0009	1	PremiumVinegar50mI	12.00

Text-to-SQL Query Generation The Text-to-SQL module enables analytics queries over structured ERP data using natural language [14]. Inputs are transmitted as multipart/form-data consisting of a user text query and zero, one, or multiple attachments (CSV or XLSX). Generated SQL queries operate over tabulated outputs from upstream modules, and results are returned using the same unified message schema as the chat interface. For the purpose of the discussion, an open-source dataset on product orders was used to capture and retrieve business intelligence results.

Table 3. Sample natural-language analytics queries for Text-to-SQL generation.

Natural-Language Query (Verbatim)
"How many ang benta or sales ko per category, and what is the distribution in percentages? Also who are the most frequent buyers? Generate visual charts of the data."

Summary By enforcing a unified schema and modular evaluation strategy, the system demonstrates how OCR-based digitization, conversational data entry, entity resolution, and analytics querying can coexist within a single AI-augmented ERP pipeline. This modular design supports extensibility while maintaining interpretability, human oversight, and compatibility with manual MSME workflows.

4 Results

This section reports the preliminary quantitative and qualitative results of the proposed AI-augmented ERP system, organized by module. Results are presented to reflect both system performance and user-perceived usefulness, consistent with the study's HITL design objectives.

4.1 OCR–LLM Extraction Performance

The OCR–LLM pipeline demonstrated practical accuracy for digitizing handwritten MSME sales logs. While precision varied across trials due to handwriting variability and layout inconsistencies, recall remained consistently high, indicating the system's ability to capture most relevant transactional entities. These results suggest that OCR augmented with LLM-based validation is suitable for semi-automated digitization workflows, where human verification remains available.

Table 4. Precision and recall across OCR extraction trials.

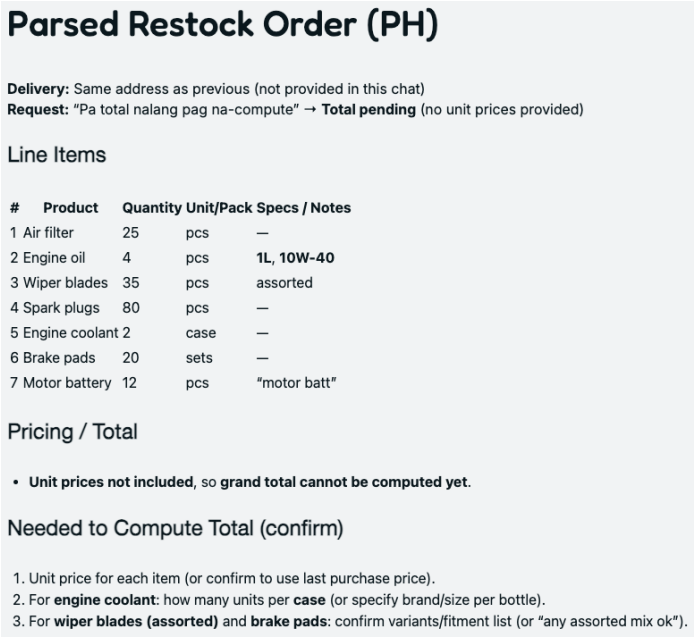
Trial	**Precision**	**Recall**
Trial 1	1.00	1.00
Trial 2	1.00	1.00
Trial 3	0.933	0.933
Average	0.971	0.971

Qualitative observations indicate that contextual grounding using predefined SKUs and price lists improved extraction stability, while post-processing by a secondary LLM reduced formatting and consistency errors. Users reported that minor inaccuracies were acceptable given the substantial reduction in manual encoding effort, particularly when paired with transparent validation and correction mechanisms.

4.2 Chat-Based Order Parsing

The conversational interface demonstrated reliable extraction of entities from bilingual (Taglish) messages, reflecting real-world MSME communication pat-

terns. The system successfully interpreted informal phrasing, abbreviated product names, and mixed-language inputs, enabling consistent downstream structuring of orders.



Parsed Restock Order (PH)

Delivery: Same address as previous (not provided in this chat)
Request: "Pa total nalang pag na-compute" → **Total pending** (no unit prices provided)

Line Items

#	Product	Quantity	Unit/Pack	Specs / Notes
1	Air filter	25	pcs	—
2	Engine oil	4	pcs	**1L, 10W-40**
3	Wiper blades	35	pcs	assorted
4	Spark plugs	80	pcs	—
5	Engine coolant	2	case	—
6	Brake pads	20	sets	—
7	Motor battery	12	pcs	"motor batt"

Pricing / Total

- **Unit prices not included**, so **grand total cannot be computed yet**.

Needed to Compute Total (confirm)

1. Unit price for each item (or confirm to use last purchase price).
2. For **engine coolant**: how many units per **case** (or specify brand/size per bottle).
3. For **wiper blades (assorted)** and **brake pads**: confirm variants/fitment list (or "any assorted mix ok").

Fig. 3. Chat-based ordering results extracted and parsed from natural language.

Evaluation focused on practical usability rather than purely technical accuracy. Users were able to submit orders conversationally and obtain structured outputs that matched their intended transactions. Qualitative feedback emphasized the importance of language familiarity and conversational tone. Participants reported higher confidence when the system acknowledged ambiguity and requested clarification instead of making silent assumptions. These findings suggest that conversational robustness and interaction transparency are critical to adoption in low-structure communication environments.

4.3 Entity Resolution and Data Refinement

Entity resolution reduced SKU duplication and naming inconsistencies introduced by OCR noise and informal user input. The module aligns observed product name variants with canonical ERP identifiers, improving the integrity of product dimensions and enabling consistent aggregation for inventory and sales reporting.

This process is consistent with prior work that models noisy product names as an entity resolution problem, where each observed string is treated as a reference to an underlying SKU entity. OCR-generated mutations such as substitution, insertion, and deletion errors prevent exact string matching and degrade the usability of product names as analytical dimensions. Clustering references based on string similarity enables multiple noisy observations to be consolidated into a single entity representation, thereby restoring dimensional consistency for business intelligence applications [**?**].

Consistent with findings in the literature, the resolution process was observed to be most effective in recovering coarse-grained product groupings, such as brand or generic product categories, which are sufficient for most operational analytics tasks. Participants reported that the consolidation of product variants improved inventory visibility and prevented fragmented stock records.

User feedback further indicated that confidence scores and transparent correction suggestions enhanced trust in automated consolidation. Rather than requiring fully autonomous correction, the human-in-the-loop design allowed users to validate uncertain matches, reducing manual data cleaning effort while maintaining perceived control over master data quality.

4.4 Natural Language Querying (Text-to-SQL)

The natural language query interface enabled users to retrieve analytics without prior technical or SQL knowledge. Participants issued queries related to sales summaries, customer behavior, product performance, and geographic trends using everyday conversational language.

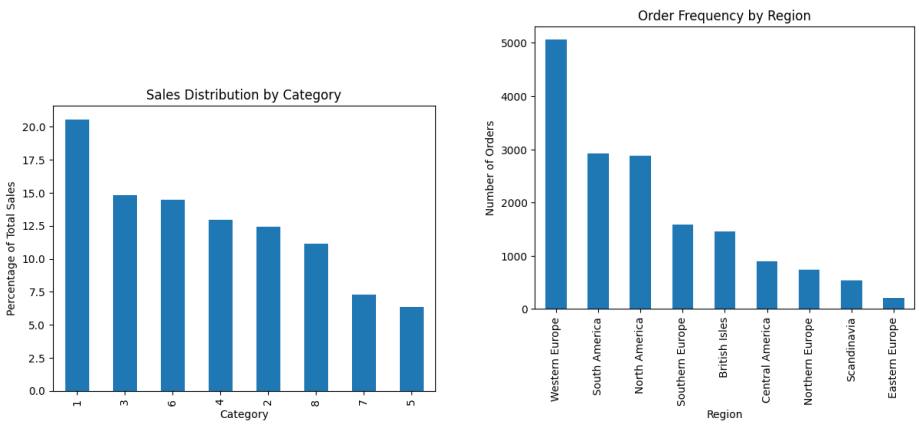


Fig. 4. Order frequency by region (left) and sales distribution by category (right).

Assessment focused on accessibility and decision usefulness rather than execution accuracy alone. Users were able to express analytical questions in natural language and receive structured summaries and visualizations without navigating predefined reports or technical interfaces. Participants reported that this interaction lowered the perceived barrier to data use and enabled more frequent exploratory queries.

Qualitative observations indicate that usefulness was driven by three factors: interpretability of outputs, alignment of results with business-relevant questions, and transparency in how answers were derived. When responses included labeled visualizations and clear aggregations, users expressed greater confidence in applying the results to operational decisions. Polite system feedback and clarification prompts in cases of ambiguity further contributed to trust, highlighting the importance of interaction design in hybrid human–AI analytics environments.

4.5 Summary of Findings

Across all modules, results indicate that accessibility, transparency, and user control are central to the successful integration of AI into MSME workflows. While quantitative performance metrics were available for the OCR module, evaluation of the remaining components emphasized practical usability, interpretability, and decision relevance.

Qualitative findings consistently suggest that contextual UX choices, including bilingual interaction, confidence signaling, and user-controlled correction, increase perceived usefulness and trust. These factors appear to play a critical role in enabling adoption in low-digital-literacy environments, where the value of AI systems is determined not only by technical performance but by their ability to reduce operational effort and support everyday business decision-making.

5 Discussion

This section interprets the findings in relation to the study's research questions and examines the implications for the design of AI-augmented enterprise systems for MSMEs operating in low-digital-literacy environments. The results suggest that the feasibility of such systems depends not only on model performance but on the interaction between technical capability, workflow integration, and user experience design.

RQ1) How can OCR–LLM pipelines transform handwritten and chat-based business records into structured, analytics-ready data for MSMEs?

Findings related to RQ1 indicate that OCR–LLM pipelines provide a practical foundation for converting low-structure business artifacts into structured data suitable for operational reporting and analysis. The combined use of OCR for visual extraction and LLM-based contextual validation enabled the system to recover key transactional entities despite handwriting variability and layout inconsistencies. Rather than requiring fully autonomous accuracy, the system

was designed to support human-in-the-loop verification, allowing users to review and correct extracted outputs before finalization.

This hybrid configuration shifts the role of AI from replacement to augmentation. The primary value observed was the reduction of manual encoding effort and the acceleration of record digitization, even when minor errors remained. Such findings support a progressive automation strategy, in which AI-assisted digitization serves as an incremental pathway toward structured data availability and business intelligence capability. As advances in OCR and multimodal language models continue, the reliability of automated extraction is likely to improve, further reducing the need for manual intervention while preserving user oversight.

Beyond extraction, the integration of entity resolution was critical for ensuring downstream analytical usability. Without consolidation of noisy product names, dimensional inconsistencies would limit aggregation and reporting. The results therefore highlight that data refinement processes, including clustering and canonical mapping, are not ancillary components but essential elements in transforming raw AI outputs into analytics-ready enterprise data.

RQ2) To what extent can multilingual conversational interfaces enhance MSME engagement with enterprise systems and encourage data-driven decision-making?

With respect to RQ2, the findings suggest that multilingual conversational interfaces substantially lower the interaction barriers associated with traditional enterprise systems. Allowing users to submit orders, retrieve reports, and request summaries using conversational language aligned system interaction with existing communication practices. The ability to use mixed-language inputs (e.g., Taglish) was particularly important in reducing formality requirements and increasing user comfort.

Importantly, the value of conversational interaction extended beyond input flexibility. Natural language querying enabled users to access analytics without navigating predefined dashboards or understanding database structures, effectively democratizing access to business insights. Participants were able to pose operational questions and immediately receive interpretable summaries and visualizations, supporting exploratory data use and more frequent decision-oriented queries.

However, the results also reveal boundary conditions for conversational robustness. Dialectal variation, shorthand expressions, and highly informal phrasing introduce ambiguity that may not be fully resolved through general-purpose language understanding. Performance was most stable when conversational inputs were grounded by domain schemas, few-shot examples, or constrained response formats. These observations suggest that conversational AI for enterprise contexts should be designed as a controlled natural language interface rather than an unconstrained dialogue system, with domain grounding playing a central role in maintaining reliability.

RQ3) What UX design principles support trust, usability, and sustained adoption of AI-driven enterprise tools in low-digital-literacy contexts?

The findings related to RQ3 emphasize that user experience design is a primary determinant of trust and sustained use, often outweighing marginal differences in technical accuracy. Participants consistently preferred interaction patterns that resembled familiar mobile messaging environments, indicating that familiarity and perceived simplicity reduce cognitive load and encourage engagement.

Several design principles emerged as critical for adoption. First, transparency mechanisms such as confidence indicators, clarification prompts, and editable outputs increased user confidence in automated results. Rather than expecting error-free performance, users valued visibility into system reasoning and the ability to intervene when necessary. Second, minimizing cognitive and technical demands, by avoiding code-based interactions and complex navigation, enabled effective use among individuals with limited digital experience. Third, reliance on existing devices, particularly mobile phones for image capture and messaging, reduced infrastructural barriers and supported inclusivity.

Taken together, these findings suggest that trust in hybrid human–AI enterprise systems is constructed through interaction design that supports user control, feedback clarity, and alignment with everyday technological practices. In low-digital-literacy contexts, perceived agency and familiarity appear to be as important as algorithmic performance in determining system acceptance.

6 Conclusion

This study presented the design and preliminary evaluation of an AI-augmented enterprise system for MSMEs operating in low-digital contexts. The proposed architecture integrates OCR, large language models, entity resolution, and multilingual conversational interfaces within a human-in-the-loop framework to transform handwritten and informal transactional data into structured, analytics-ready information.

The findings contribute to HCI research by demonstrating how hybrid human–AI architectures can be designed to align with the operational realities of small enterprises in emerging markets. Rather than assuming fully automated decision-making, the system emphasizes assisted intelligence, where automation accelerates routine tasks while users retain oversight through validation, correction, and interpretability mechanisms. This balance between efficiency and control was observed to support user confidence, reduce perceived risk, and encourage engagement with digital tools.

Three key implications emerge from the study. First, empathy-driven design, including language localization, conversational interaction, and alignment with familiar communication patterns, plays a critical role in building trust in AI-assisted workflows. Second, human validation mechanisms preserve accountability while enabling meaningful efficiency gains, suggesting that transparency and user control are central to adoption in low-digital-literacy environments. Third, the integration of extraction, refinement, conversational interaction, and natural language analytics demonstrates how multiple AI capabilities can be or-

chestrated to extend inclusivity in enterprise system design, enabling data-driven decision-making without requiring technical expertise or significant infrastructure investment [4][15][17][18].

More broadly, the study contributes to ongoing discussions in HCI and information systems regarding the role of explainability and user agency in AI deployment. The results suggest that in resource-constrained and low-literacy contexts, perceived control and interaction clarity may be more influential for trust than marginal improvements in algorithmic accuracy. This highlights an important design trade-off: systems that prioritize interpretability, feedback, and recoverability may achieve higher real-world adoption even when automation performance remains imperfect.

6.1 Implications

The findings offer several implications for research and practice. For HCI research, the study reinforces the importance of contextualized AI design, where interaction models reflect users' existing technological habits rather than requiring behavioral change. Conversational interfaces, bilingual support, and mobile-first workflows demonstrate how familiar interaction paradigms can reduce cognitive barriers and increase system accessibility.

For enterprise system design, the results suggest that digital transformation strategies for MSMEs should prioritize workflow compatibility over feature complexity. Systems that integrate with existing practices, such as handwritten logs and messaging-based communication, are more likely to achieve sustained use than traditional ERP platforms that require formal data entry and structured navigation.

For AI deployment in emerging markets, the study highlights the value of incremental automation. Instead of positioning AI as a replacement for human work, the proposed architecture illustrates how layered assistance—extraction, consolidation, and interpretation—can gradually build data maturity. This approach lowers adoption risk while enabling organizations to transition from informal record-keeping toward analytics-driven decision-making.

Finally, the results suggest that data accessibility, rather than model sophistication alone, is a critical determinant of impact. By enabling users to ask questions in natural language and immediately receive interpretable outputs, the system supports everyday operational decisions, thereby increasing the practical value of enterprise data.

6.2 Future Work

Several directions for future research emerge from this work. First, longitudinal studies are needed to examine sustained adoption, behavioral change, and business outcomes over extended deployment periods. Understanding how usage patterns evolve and whether access to analytics leads to measurable operational improvements remains an important area for investigation.

Second, future work should expand linguistic and domain coverage. While the system demonstrated robustness for bilingual conversational input, regional dialects, industry-specific terminology, and highly informal language structures present ongoing challenges. Adaptive learning mechanisms that incorporate user corrections may improve performance over time while maintaining system reliability.

Third, improvements in entity resolution and data refinement should address the trade-off between coarse-grained consolidation and fine-grained product differentiation. Techniques that incorporate contextual signals such as pricing patterns, transaction frequency, or user feedback may enhance precision for inventory-sensitive applications.

Fourth, scalability and deployment considerations, including offline functionality, low-bandwidth optimization, and integration with financial and supply chain services, should be explored to support broader adoption across diverse MSME environments.

Finally, future research should examine governance and ethical considerations associated with AI-assisted enterprise systems, including data ownership, privacy, and the design of safeguards that maintain user autonomy while enabling automated insight generation.

Acknowledgments. We thank the John Gokongwei School of Management, Ateneo Business Insights Laboratory for Development (BUILD), and the Ateneo University Research Council for supporting our study and its future presentation.

References

1. Gemini 3 Developer Guide | Gemini API, https://ai.google.dev/gemini-api/docs/gemini-3
2. Introducing Claude Sonnet 4.5, https://www.anthropic.com/news/claude-sonnet-4-5
3. Ahmed, R., Gopal, A.B.: Consumer Perception of Conversational Commerce from the Perspective of Bangladesh. Dhaka University Journal of Business Studies pp. 53–76 (2025). https://doi.org/10.3329/dujbst.v44i1.84710, https://banglajol.info/index.php/DUJBSt/article/view/84710
4. Alabastro, Z.M., Ilagan, J.B., To, L.A., Ilagan, J.R.: Applied Optical Character Recognition and Large Language Models in Augmenting Manual Business Processes for Data Analytics in Traditional Small Businesses with Minimal Digital Adoption. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI, vol. 15822, pp. 267–276. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-93429-2_18, https://link.springer.com/10.1007/978-3-031-93429-2_18, series Title: Lecture Notes in Computer Science
5. Buckland, M., Gey, F.: The relationship between Recall and Precision. Journal of the American Society for Information Science **45**(1), 12–19 (1994). https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L, https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199401%2945%3A1%3C12%3A%3AAID-ASI2%3E3.0.CO%3B2-L, _eprint:

https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-4571%28199401%2945%3A1%3C12%3A%3AAID-ASI2%3E3.0.CO%3B2-L

6. Christophides, V., Efthymiou, V., Palpanas, T., Papadakis, G., Stefanidis, K.: An Overview of End-to-End Entity Resolution for Big Data. ACM Comput. Surv. **53**(6), 127:1–127:42 (Dec 2020). https://doi.org/10.1145/3418896, https://dl.acm.org/doi/10.1145/3418896
7. Gao, D., Wang, H., Li, Y., Sun, X., Qian, Y., Ding, B., Zhou, J.: Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation (Nov 2023). https://doi.org/10.48550/arXiv.2308.15363, http://arxiv.org/abs/2308.15363, arXiv:2308.15363 [cs]
8. Haddara, M., Elragal, A.: ERP adoption cost factors identification and classification: a study in SMEs. International Journal of Information Systems and Project Management **1**(2), 5–21 (2013), https://revistas.uminho.pt/index.php/ijispm/article/view/3951
9. Ikhtiari, K., Muslim, M., Nurfadila, N.: Improving MSME Accounting Financial Recording Skills Based on Android Applications. Advances in Community Services Research **2**(2), 62–73 (2024), http://advancesinresearch.id/index.php/ACSR/article/view/137
10. Ilagan, J.R., Ilagan, J.B.: Graph-partitioning entity resolution for resolving noisy product names in OCR scans of retail receipts. Procedia Computer Science **239**, 338–345 (2024), https://www.sciencedirect.com/science/article/pii/S1877050924014236
11. Ilagan, J.B., Ilagan, J.R., Zulueta, P.Y., Rodrigo, M.M.: Optimizing Conversational Commerce Involving Multilingual Consumers Through Large Language Models' Natural Language Understanding Abilities. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI, vol. 14736, pp. 47–59. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-60615-1_4, https://link.springer.com/10.1007/978-3-031-60615-1_4, series Title: Lecture Notes in Computer Science
12. Kumar, S., Datta, S., Singh, V., Datta, D., Kumar Singh, S., Sharma, R.: Applications, Challenges, and Future Directions of Human-in-the-Loop Learning. IEEE Access **12**, 75735–75760 (2024). https://doi.org/10.1109/ACCESS.2024.3401547, https://ieeexplore.ieee.org/abstract/document/10530996
13. Li, X., You, J., Li, H., Peng, J., Chen, X., Guo, Z.: CM-SQL: A cross-model consistency framework for text-to-SQL. Neurocomputing **658**, 131708 (Dec 2025). https://doi.org/10.1016/j.neucom.2025.131708, https://www.sciencedirect.com/science/article/pii/S092523122502380X
14. Rane, N., Paramesha, M., Choudhary, S., Rane, J.: Business Intelligence through Artificial Intelligence: A Review (May 2024). https://doi.org/10.2139/ssrn.4831916, https://papers.ssrn.com/abstract=4831916
15. Sagala, G.H., Őri, D.: Toward SMEs digital transformation success: a systematic literature review. Information Systems and e-Business Management **22**(4), 667–719 (Dec 2024). https://doi.org/10.1007/s10257-024-00682-2, https://link.springer.com/10.1007/s10257-024-00682-2
16. Vafaie, M., Hertling, S., Banse-Strobel, I., Dubout, K., Sack, H.: End-to-end Information Extraction from Archival Records with Multimodal Large Language Models. In: Proceedings of the 34th ACM International Conference on Information and Knowledge Management. pp. 6075–6083. CIKM '25, Association for Computing Machinery, New York, NY, USA (Nov 2025). https://doi.org/10.1145/3746252.3761503, https://dl.acm.org/doi/10.1145/3746252.3761503

17. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science **46**(2), 186–204 (Feb 2000). https://doi.org/10.1287/mnsc.46.2.186.11926, https://pubsonline.informs.org/doi/10.1287/mnsc.46.2.186.11926
18. Wang, D., Yang, Q., Abdul, A., Lim, B.Y.: Designing Theory-Driven User-Centric Explainable AI. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–15. ACM, Glasgow Scotland Uk (May 2019). https://doi.org/10.1145/3290605.3300831, https://dl.acm.org/doi/10.1145/3290605.3300831
19. Xu, D., Xie, T., Xia, B., Li, H., Bai, Y., Sun, Y., Wang, W.: Does Few-Shot Learning Help LLM Performance in Code Synthesis? (Dec 2024). https://doi.org/10.48550/arXiv.2412.02906, http://arxiv.org/abs/2412.02906, arXiv:2412.02906 [cs]
20. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K.R., Cao, Y.: ReAct: Synergizing Reasoning and Acting in Language Models (Sep 2022), https://openreview.net/forum?id=WE_vluYUL-X
21. Zanzotto, F.M.: Human-in-the-loop artificial intelligence. Journal of Artificial Intelligence Research **64**, 243–252 (2019), http://www.jair.org/index.php/jair/article/view/11345