УДК 004.896

ИНТЕГРАЦИЯ NLP-ПАЙПЛАЙНА НА ОСНОВЕ GPT В KINGDEE CLOUD GALAXY ДЛЯ СНИЖЕНИЯ ОШИБОК КОМПЛЕКТАЦИИ В РЕАЛЬНОМ ВРЕМЕНИ: ПОЛЕВОЙ ЭКСПЕРИМЕНТ НА СКЛАДЕ ВЫСОКОСКОРОСТНОЙ ОБРАБОТКИ НАПИТКОВ КОМПАНИИ GENKI FOREST (ТАЙЦАН, КИТАЙ)

Люй Синьтун, советник Школы искусственного интеллекта,
степень магистра гуманитарных наук.
Технологический институт Сучжоу Цзянь-шун,
город Тайцан, провинция Цзянсу, Китай
Научный руководитель — Буйная Елена Васильевна, доцент кафедры
прикладных информационных технологий, кандидат экономических наук
Кузбасский государственный технический университет
имени Т. Ф. Горбачёва,
Россия, Кемерово

Introduction. Widely adopted ERP systems such as Kingdee Cloud Galaxy have significantly improved warehouse management efficiency. However, as observed at the Genki Forest warehouse in Taicang, traditional menu-driven interfaces with strict input formats often cause operational delays and fail to prevent human errors during scanning and reporting.

The “last-meter” verification in picking is critical: operators may verbally confirm the wrong SKU. Existing solutions cannot semantically validate spoken SKU against expected SKU. This study offers an innovative fix by integrating a lightweight, fine-tuned large language model (LLM) into Kingdee Cloud Galaxy via Kingdee OpenAPI.

Objectives:

1. Design and deploy a voice-input, production-grade NLP pipeline for real-time SKU verification.
2. Evaluate its effectiveness via rigorous A/B testing in a live warehouse, focusing on error reduction and latency.
3. Demonstrate the architecture’s generality and replicability across ERP systems.

Materials and Methods.

Solution Architecture. Figure 1 illustrates the proposed NLP pipeline acting as an intelligent middleware without deep ERP modification. Core components include:
(a)Audio capture client.
(b)Automatic Speech Recognition (ASR) service.
(c)GPT-based semantic understanding model.
(d) API integration layer.

Workflow:

1. Operator speech is captured via headset.

2. Audio stream is sent to ASR.

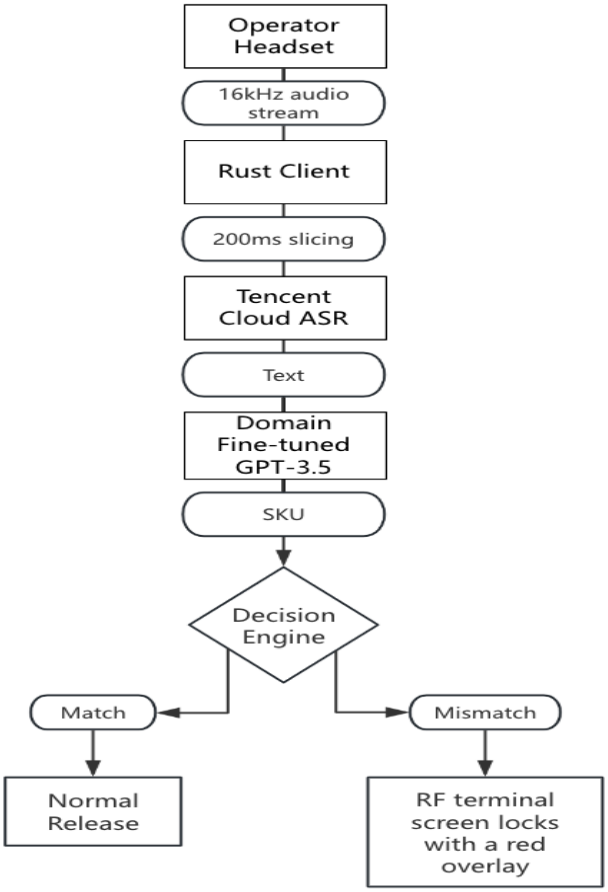

Figure 1 – Proposed Architecture of the NLP Pipeline

3. Transcript is parsed by a domain-fine-tuned GPT-3.5 model.
4. Predicted SKU is compared to expected SKU in the order.
5. Mismatch triggers an alert via Kingdee OpenAPI.
6. RF terminal shows red overlay and prompts re-scan.

Data Collection and Preprocessing. Operators wore Jabra Evolve 65 headsets (16 kHz). A local Rust service sliced audio into 200 ms chunks and forwarded them to Tencent Cloud ASR (beam size = 5). Over 10 shifts, 1,256 valid utterances were collected, containing 3,019 SKU entity mentions. Domain experts manually labeled SKU, quantity, and location. Data augmentation included synonym replacement and phonetic noise injection.

Domain Fine-Tuning and Baselines. We fine-tuned GPT-3.5-turbo using LoRA (rank = 8) for one epoch (lr = 2e-5). SKU-related tokens were weighted 3× in loss. Dataset split: 1,005 train, 251 validation. Baselines:
(a) Rule-based keyword matcher.
(b) Off-the-shelf BERT without fine-tuning.

Real-Time Decision Engine. If top-predicted SKU ≠ expected SKU, a JSON payload is sent to:
/k3cloud/api/v1/picking/verify
{"error": 1, "expected": "SKU10294", "detected": "SKU10301"}
RF terminal locks with red overlay until correct item is scanned.

Evaluation Protocol. Day 1: A/A test for baseline. Days 2–4: A/B test with 2-hour toggles to offset daily volume bias. Primary KPI: picking error rate per 100 orders. Secondary: verification latency.

Results.

Table 1: Summary of Error Rate and Latency Results

Metric	Baseline	NLP Pipeline	Δ
Error Rate (%)	3.40 ± 0.22	1.07 ± 0.15	−68%
Latency (s)	—	1.69 ± 0.14	—
SKU F1	—	0.92 ± 0.03	—

Statistical significance: $\chi^2 = 31.4$, $p < 0.001$, Cohen's $d = 2.7$ (large effect). 85% of requests processed within 1.5s; peak latency 2.2s did not disrupt throughput.

Residual Error Breakdown:

Table 2: Residual Error Analysis

Error Type	Count	Share
Phonetic similarity	12	28%
ASR misrecognition	18	42%
Model hallucination	8	19%
Other	4	11%

Discussion. The pipeline cut picking errors by two-thirds without reducing throughput (600 cartons/hour) or adding headcount. Most residual errors stem from ASR limitations and ambient noise, not semantic understanding. Future work should focus on noise-robust acoustic models and on-device ASR fine-tuning. The architecture is robust and corrects the vast majority of human slips.

Conclusion. We demonstrated that a lightweight, domain-fine-tuned GPT model, integrated via Kingdee OpenAPI, enables real-time voice-based SKU verification, reducing picking errors by over two-thirds. The hybrid LLM-ERP architecture offers significant potential for precision and efficiency gains in logistics and is readily deployable across any Kingdee Cloud Galaxy site.

List of literature:

1. M. Arslan, Z. Riaz, and C. Cruz, “Revolutionizing management information systems with natural language processing for digital transformation,” Procedia Computer Science, vol. 225, pp. 2835–2844, 2023.

2. M. Vega-Flores et al., “Natural language processing integrated into a CRM system for document control and sales logistics at a car dealership,” Journal of Information Systems Engineering and Management, vol. 10, no. 31s, 2025.

3. E. Niederwieser, D. Siegele, and D. T. Matt, “AI-driven ERP systems integrating large language models for enhanced customer interaction and operational efficiency,” KI – Künstliche Intelligenz, vol. 120, pp. 112–117, 2025.

4. S. Winasis and A. Terminanto, “Envisioning the future: the role of artificial intelligence in next generation ERP systems,” JMBI UNSRAT, vol. 12, no. 1, pp. 296–306, 2025.

5. D. Khan, “Unleashing business intelligence: text analytics and AI-driven ERP transformations towards an intelligent enterprise,” Social Sciences Spectrum, vol. 2, no. 1, pp. 103–110, 2023.

6. H. Aşan and V. Tecim, “Model design and development for a user support system using artificial intelligence techniques in enterprise resource planning software,” IJCEAS, vol. 13, no. 2, pp. 836–864, 2023.

7. A. Maged and G. Kassem, “Self-adaptive ERP: embedding NLP into Petri-Net creation and model matching,” in Proc. 3rd Int. Conf. on Design Science Research in Information Systems and Technology, 2024.

8. R. Hrischev and N. Shakev, “Artificial intelligence in ERP systems: SAP business technology platform case study,” Sciforce Journal, 2023.

9. M. E. Vallejo Venegas et al., “Modern ML methods for failure analysis automation on classical ERP systems,” Ciencia Latina Revista Científica Multidisciplinar, vol. 8, no. 5, 2024.

10. A. Faccia et al., “Blazing a new trail in ERP integration with NLP and generative AI through APIs: a fraud examination perspective,” in Proc. ICISE 2023, pp. 177–184, 2023.

11. V. Dachepalli, “A smarter ERP: how artificial intelligence is reshaping enterprise workflows,” Journal of Artificial Intelligence and Machine Learning, vol. 2, no. 2, pp. 1–11, 2024.

12. H. N. Patil, “ERP management system,” B.Eng. thesis, Parul University, Vadodara, India, 2023.

13. Kingdee Cloud Galaxy OpenAPI Specification v3.2, 2024.

14. Radford A. et al. Robust Speech Recognition via LargeScale Weak Supervision, 2023.