LTIMindtree

Whitepaper

Reimagining Unstructured Document Processing for Global Banking

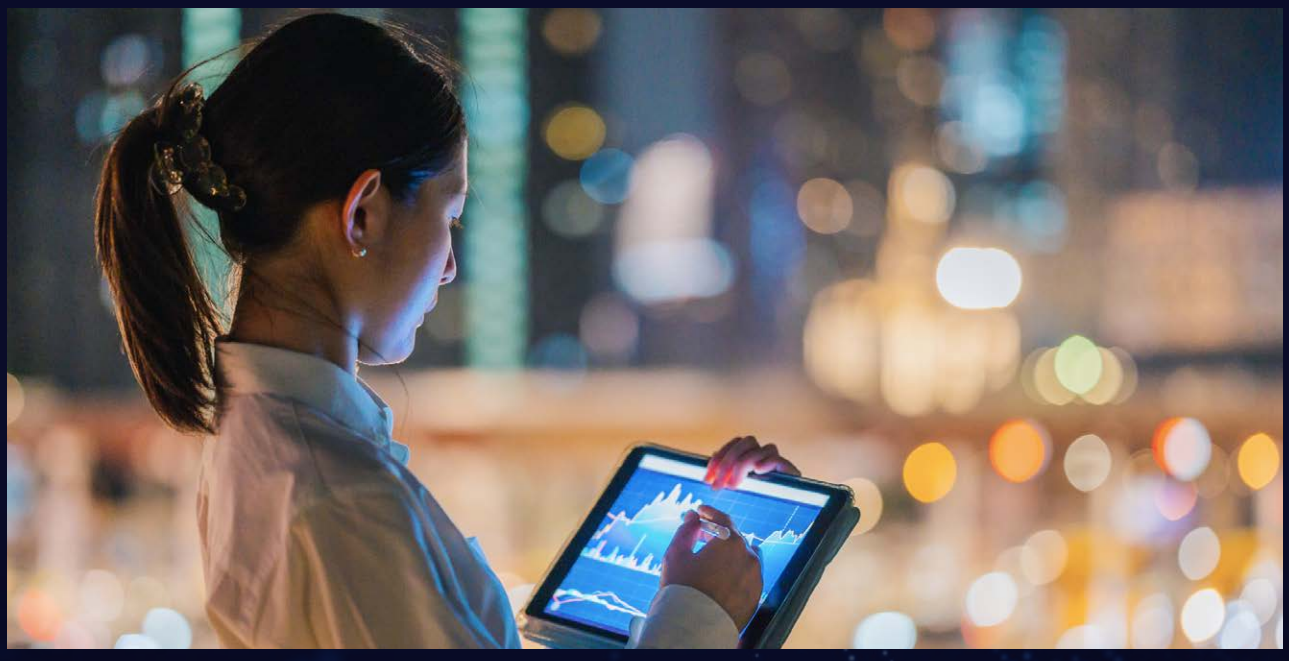
Table of contents

Executive summary

Global banks and multinational corporations face persistent challenges in managing unstructured documents across core functions like Trade Finance, Corporate Finance, KYC, Treasury, and Banking Operations. These high-volume, complex processes are hindered by fragmented systems, manual interventions, and inconsistent data formats, especially when handling multi-language, handwritten, and scanned documents. In today's fast-paced financial environment, delays, fraud risks, and compliance gaps arising from inefficient document handling directly impact operational performance and client satisfaction.

The white paper states a critical pain point in the Procure-to-Pay (P2P) cycle, where unstructured invoices in varied formats often lead to delayed payments, strained supplier relationships, missed validations, and weak audit trails.

To solve this, the proposed solution leverages Generative AI and Multimodal LLMs, specifically Google's Gemini, to power an "Intelligent Document Processing (IDP) as a Service" model. Using advanced prompt engineering, domain-tuned machine learning, and a configurable document intake pipeline with a maker-checker framework, the solution automates extraction, validation, and straight-through processing. Live implementations have shown significant cost reduction, faster turnaround times, enhanced compliance, and improved service delivery. This gen AI-powered IDP framework unlocks measurable value for global banks by transforming unstructured document processing into a scalable, intelligent, and efficient operation.



Introduction

In today's digital economy, invoices are everywhere—scanned PDFs, email attachments, multilingual formats from vendors across continents. For global banks, this isn't just paperwork—it's a tidal wave. With millions of invoices processed annually, the cost of manual handling and legacy OCR systems is staggering. According to the Institute of Finance & Management and other industry sources, manual invoice processing can cost up to $16 per invoice[1], not to mention the delays that ripple through payment cycles and vendor relationships.

Picture this: a bank in Singapore receives a handwritten invoice in Thai, while another in Frankfurt gets a scanned image in Cyrillic. Traditional OCR tools choke on format inconsistencies, and human processors struggle with language barriers. The result? Missed payments, compliance headaches, and frustrated vendors.

And with global regulations tightening around invoice data and audit trails, the stakes have never been higher.

That's why we're introducing a smarter way forward—an Intelligent Document Processing solution powered by Gemini LLM, a gen AI-led multimodal solution built for speed, scale, and precision. This whitepaper explores how an optimized variant of Gemini can transform invoice processing—cutting costs, boosting accuracy, and giving banks the agility they need to thrive in a data-heavy world.

Key challenges organizations face

Manual intervention

Despite the use of OCR technology, significant manual effort is still required in both the maker and checker stages to correct errors, validate entries, and ensure data accuracy. This not only slows down processing but also increases the risk of human error—especially critical in high-volume environments.

Template variability

Traditional solutions struggle with varying invoice templates, leading to inconsistent extraction of data.

Processing speed

Conventional OCR-based methods are slow and unable to meet the demands of high-volume processing. In today's fast-paced environment of global finance, slow processing translates directly to delayed payments, strained vendor relationships, and missed opportunities. This is particularly critical when processing millions of invoices annually.

Data accuracy

Handwritten or poorly scanned invoices present a significant challenge for traditional OCR solutions, resulting in frequent misinterpretations.

Multilingual support

Handling invoices in multiple languages often requires separate systems, leading to inefficiencies.

Figure 1: Key challenges with document processing operations

Proposed solution
Gemini LLM with prompt engineering

The Gemini LLM leverages advanced prompt engineering and multimodal capabilities to address the challenges outlined above. It integrates several specialized models that focus on specific tasks, such as OCR, language detection, and context understanding, into a unified framework. These models work in tandem to improve the overall data extraction process, ensuring a more accurate and efficient solution. By utilizing an optimized variant within the Gemini family, this Intelligent Document Processing solution is engineered for speed and efficiency, critical for handling the massive scale of invoice processing in large global banks, without incurring prohibitive computational costs.

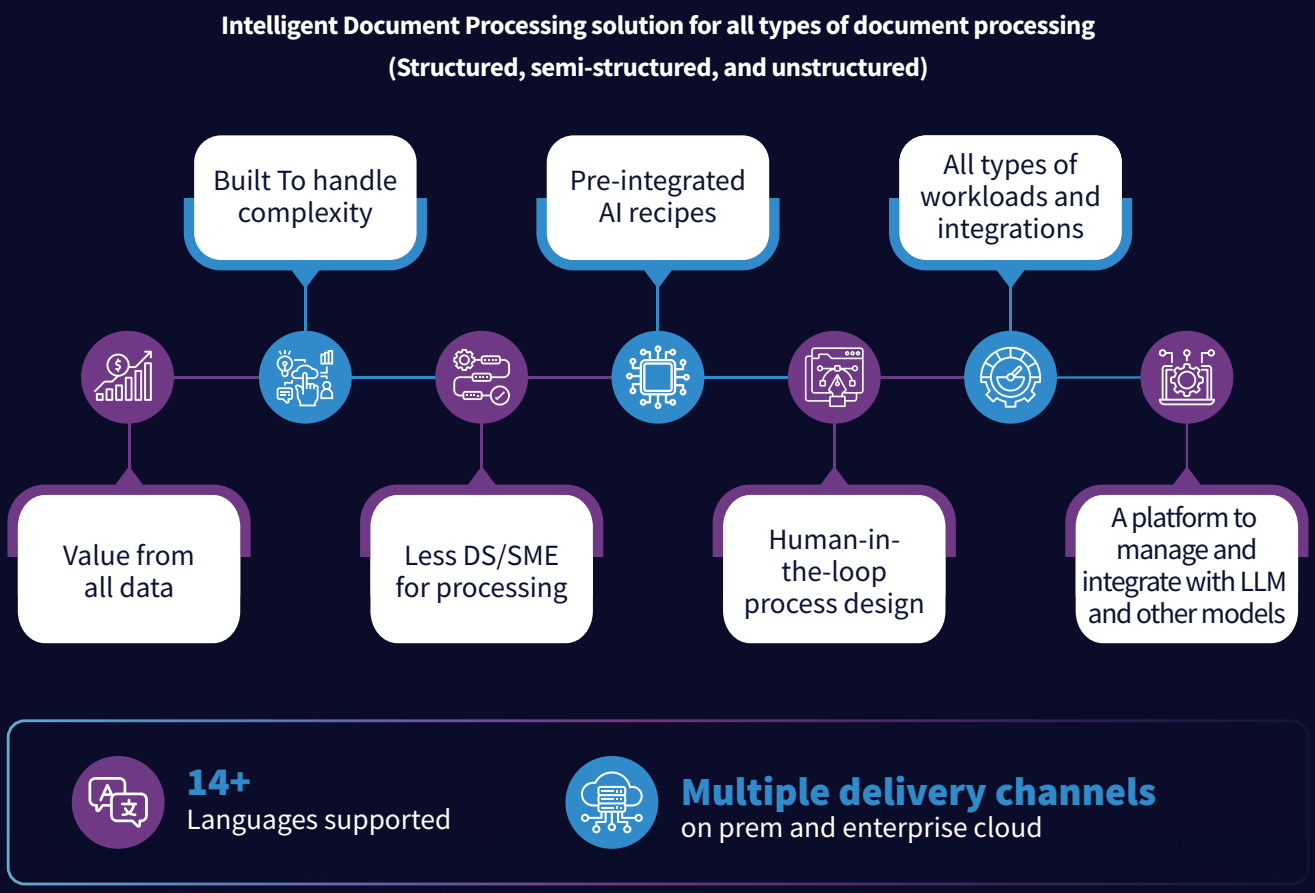


Figure 2: Document processing automation journey

Solution capabilities and benefits

The key features of the Intelligent Document Processing solution include:

High accuracy on handwritten text

Combines specialized models to process text, images, and layout — enabling precise extraction even from handwritten invoices.

Robustness to template changes

The model adjusts to varying invoice formats without needing frequent reconfiguration, reducing downtime and operational overhead.

Enhanced multilingual processing

Handles invoices in multiple languages with ease, supporting global operations and reducing the need for separate systems.

Increased processing speed

Optimized prompts enable faster and more accurate data extraction, improving throughput by up to 10x over traditional OCR.

Continuous improvement

Supports retraining and reinforcement learning from human feedback, allowing the system to evolve and improve over time.

Data completeness and integrity

Intelligent algorithms fill in missing or ambiguous fields using historical patterns and business rules, ensuring high data quality, and reducing manual checks.

Reduced manual effort

Automates validation workflows while preserving control, minimizing human intervention in both maker and checker stages.

Transparency, maintainability and usability

The solution is fully owned and maintained within the organization, ensuring transparency, ease of customization, and long-term sustainability. This enables teams to debug, enhance, and evolve the system without dependency on external vendors.

Gemini compared to traditional methods

We benchmarked Gemini LLM's performance against traditional OCR-based and manual invoice processing methods across key metrics. The table below summarizes the significant quantifiable gains achieved in accuracy, speed, manual intervention reduction, and multilingual efficiency. These improvements directly translate to substantial cost savings and operational enhancements.

Metric	Traditional Methods (Benchmark)	Gemini LLM Solution	% Improvement / Reduction	Business Impact
Data Extraction Accuracy	30%	95%	▲ +65%	Reduced errors, lower reconciliation costs, improved data quality
Handwritten Accuracy	7%	90%	▲ +83%	Enhanced accuracy for challenging documents, reduced manual correction
Processing Speed	100 pages/min	300 pages/min	▲ +30 times faster	Faster P2P cycles, improved vendor payments, working capital efficiency
Manual Intervention	Almost everything needed to be cross-checked	Minimal	-	Reduced labor costs, re-allocation of FTEs to strategic tasks
Multilingual Efficiency	15%	95%	▲ +80%	Streamlined global operations, unified workflows, faster processing
Overall Cost Reduction Potential (3-Year)	-	-	▲ 50%	Significant cost savings, improved ROI in P2P operations.

Table 1: Efficiency Gains from Gemini invoice document processing

Gemini model architecture

Gemini models are built on top of transformer decoders that are enhanced with improvements in architecture and model optimization to enable stable training at scale and optimized inference on Google's Tensor processing unit. These models are also trained to accommodate textual input interleaved with a wide variety of visual and audio inputs.

The 'pro' models are performance-optimized in terms of cost and latency and deliver significant performance across a wide range of tasks. These models exhibit strong reasoning performance and broad multimodal capabilities. They are pre-trained on a dataset that is both multimodal and multilingual, making use of their long context length effectively.

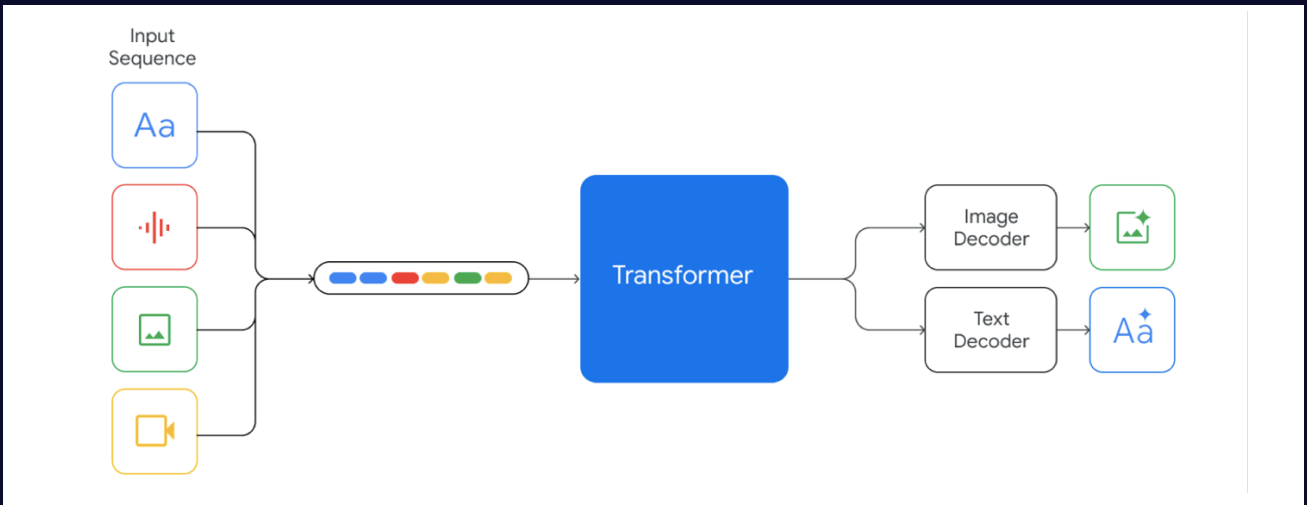


Figure 3: Gemini models support interleaved sequences of text, image, audio, and video as inputs.

	Gemini Ultra	Gemini Pro	GPT-4	GPT-3.5	PaLM 2-L	Claude 2	Inflection-2	Grok 1	LLAMA-2
MMLU Multiple-choice questions in 57 subjects (professional & academic) (Hendrycks et al., 2021a)	**90.04%** (CoT@32*) **83.7%** (5-shot)	**79.13%** (CoT@8*) **71.8%** (5-shot)	**87.29%** (CoT@32 via API**) **86.4%** (5-shot reported)	**70%** (5-shot)	**78.4%** (5-shot)	**78.5%** (5-shot CoT)	**79.6%** (5-shot)	**73.0%** (5-shot)	**68.0%***

Figure 4: Gemini performance on MMLU (Massive Multitask Language Understanding) benchmarks with external comparisons

How Gemini LLM integrates with our solution

Our solution's process is as follows:

Data collection

Unstructured invoice data, including text and images, is fed into the system. This could include handwritten or printed invoices in various formats and languages.

Preprocessing

The system applies OCR to extract textual content, while simultaneously detecting the language and layout of the invoice.

Multimodal processing

Various models within the Gemini LLM work together to interpret the data, applying context understanding to extract key information such as supplier details, invoice numbers, dates, and amounts.

Output generation

The extracted data is structured and outputted in a format that can be seamlessly integrated with other systems, such as ERP or accounting software.

Continuous improvement via RLHF

Human feedback is utilized to fine-tune the model, ensuring ongoing improvements in accuracy and handling new invoice formats.

Beyond immediate efficiency gains, the true transformative potential of Gemini in document processing lies in its ability to power the next generation of intelligent automation: Document Processing Agents.

Gemini-powered document processing agents

Gemini's evolution signals the rise of intelligent document processing agents in banking, transforming workflows beyond simple automation. Inspired by AI agents augmenting developer productivity, we envision Gemini-powered agents as co-pilots for finance professionals, autonomously streamlining document-centric tasks across P2P and wider operations.

Future capabilities that can be built

- **Proactive fraud prevention:** Real-time anomaly detection flags suspicious invoices for immediate review, shifting from reactive to preventative fraud measures.
- **Autonomous exception resolution:** Intelligent routing of exceptions, proactive data gathering, and draft communications for rapid resolution of non-standard invoices.
- **Personalized vendor management:** Agent-driven analysis of vendor patterns for tailored communication, proactive query anticipation, and dynamically adjusted interactions, ultimately enhancing vendor relationships.
- **Continuous process optimization:** Agents learn from every document and feedback loop, proactively identifying process bottlenecks, generating KPI dashboards, and driving data-led operational improvements.
- **Internal innovation platform:** Empowering bank developers to build custom agents for specialized workflows, fostering rapid in-house AI solution development leveraging Gemini's agent-building capabilities.

Gemini-powered agents go beyond just automating documents; they are a step towards true intelligence, autonomy and proactive digital assistants. The agentic era can lead to reshaping operations at the core.

Our approach to enterprise adoption and scale

For organizations interested in adopting unstructured data processing solutions will have differentiated benefits against products in the market. This is NIMBLE enough to Run "Platform as a Service", complement on existing client eco system to integrate and run. No Licensing Cost, Open Codebase for custom development, improvement and integration and scale. Furthermore, the efficiency of the underlying Gemini LLM model, designed for rapid processing and optimized resource utilization, directly contributes to the scalability and cost-effectiveness of this solution for large enterprise deployments.

- **Pilot program:** Conducting a pilot implementation to demonstrate the capabilities and benefits of the selected scope of work for MVP (minimum viable product) to run on client environment (on prem or cloud).
- **Integration:** Integrate the solution into existing systems, ensuring smooth workflows.
- **Training and optimization:** Continuously training the model and optimizing it based on business-specific requirements and feedback with minimal changes and configurable features with RBAC (role based access control).
- **Scalability and rollout:** Expanding the solution's implementation across various departments and regions to maximize its potential.

By following these steps, businesses can effectively implement unstructured document processing across multiple business function beyond invoice processing with flexibility extend on selected multimodal LLM and unlock significant efficiencies and straight through processing in operations.

Conclusion

The Gemini LLM represents a significant leap forward in the field of unstructured document processing. By integrating advanced multimodal capabilities, including prompt engineering, OCR, language processing, workflow and upstream downstream integration capability, the solution offers businesses a highly accurate, adaptable, and scalable method of processing invoices. The addition of reinforcement learning ensures continuous refinement, making it a future-proof solution capable of evolving alongside industry needs.

This innovative approach enables businesses to drastically reduce manual intervention, enhance multilingual capabilities, improve processing speed, and handle template variability with ease. By adopting Gemini LLM, organizations can streamline their invoice processing workflows, reduce operational costs, and position themselves for success in an increasingly data-driven world.

Succinctly put, for global banks ready to lead in the AI era, this is about turning document processing from a back-office necessity into a genuine strategic advantage.

Ready to transform invoice processing with GenAI?

LTIMindtree's AI-first approach delivered 50% cost savings and 95% accuracy at scale through intelligent automation, optimized maker-checker workflows, and multilingual support. Start your journey now! Write to us at **info@ltimindtree.com**

Reference

[1]*Invoice costs: What's the cost of processing an invoice?,* adobe, 2024: https://www.adobe.com/acrobat/business/hub/invoice-processing-costs.html

[2]*Introducing Gemini: our largest and most capable AI model,* Sundar Pichai, Demis Hassabis, Google, December 06, 2023: https://blog.google/technology/ai/google-gemini-ai/#sundar-note

About the author

Gunasiri Perera

Vice President, BFS AI & Data Analytics Leader Americas

Gunasiri Perera serves as Vice President and BFS AI & Data Analytics Leader – Americas at LTIMindtree, with over 27 years of global experience spanning BFSI, Telco, SCM, and the Airlines domain, Gunasiri Perera serves as Vice President and BFS AI & Data Analytics Leader – Americas at LTIMindtree. As an award-winning AI practitioner, enterprise AI strategist, and published inventor, Gunasiri enables global enterprises to drive business transformation

through AI, generative AI, advanced analytics, and scalable AI architectures.

He has successfully led strategic programs for leading institutions such as Citi, JPMC, HSBC, BNY, and NTB, holding key roles including Enterprise AI Architect, Senior Data Scientist, and Technology Advisor to CXOs. He has built and led high-performing teams across the U.S., Scandinavia, the UAE, Sri Lanka, and India. He has previously played a pivotal role at the C-level in two startups.

A Chartered Engineer and active member of the MIT alumni network, Gunasiri regularly contributes to academia and serves as a subject matter expert in the industry. He sholds a BSc (Hons) in Computer Science and engineering, is a certified Data Scientist from MIT, and has an MBA in Strategic Management.