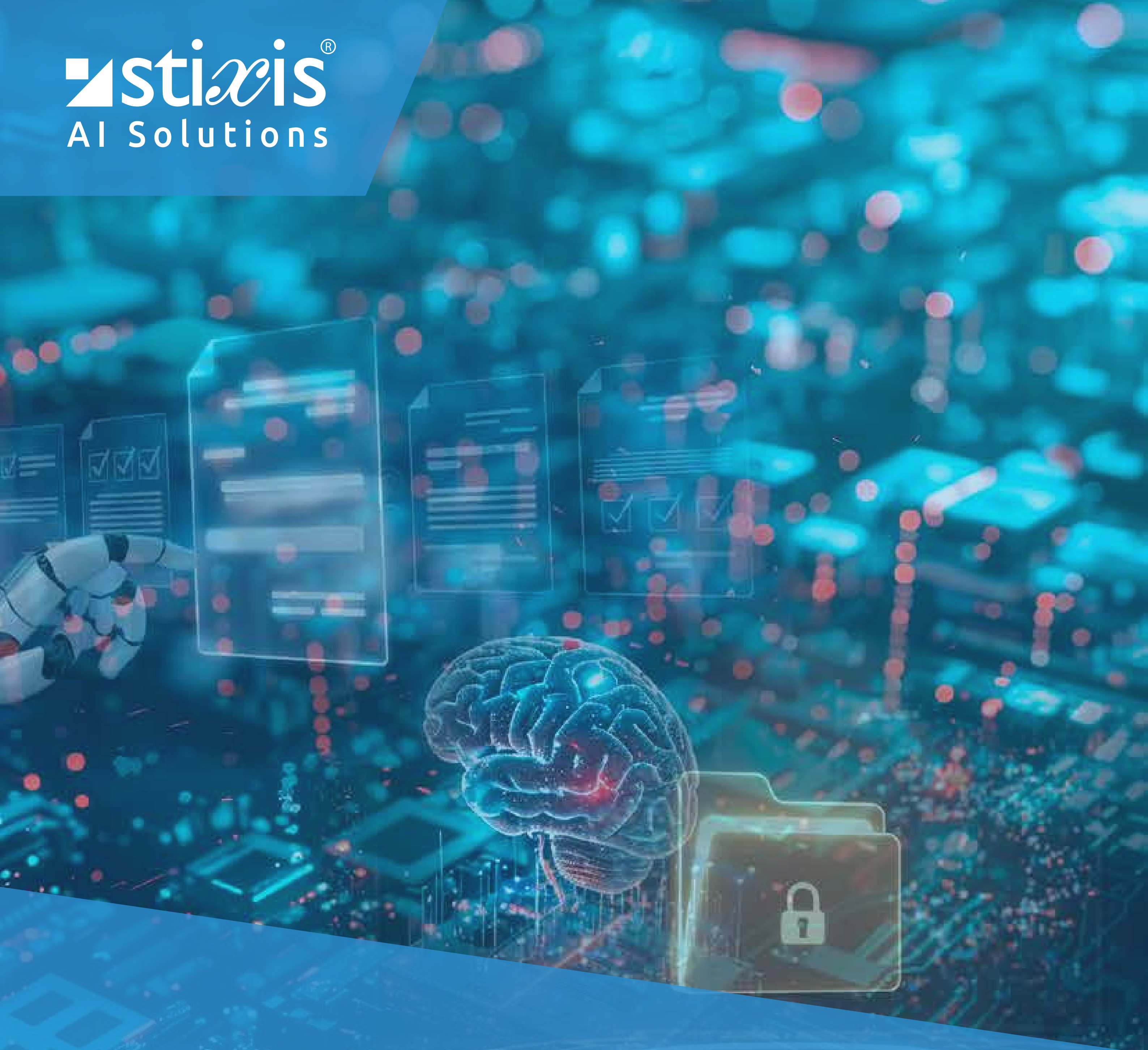
stixis®
AI Solutions



Transforming Document Processing with AI Agents

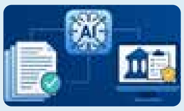
www.stixis.ai

Overview

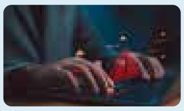
Revolutionising document processing workflows to unlock efficiency, reduce costs, and enable scalable growth.

Business Challenge

The client, a large enterprise managing millions of documents annually, faced severe inefficiencies in document processing workflows. Their operations relied heavily on manual data extraction and outdated tools, making the process



Time-consuming, with long turnaround times for document validation and extraction.



Error-prone, especially when dealing with unstructured or multi-format data (PDFs, scanned images, handwritten forms, etc.).

Expensive, requiring large teams to manage repetitive tasks.

Difficult to scale, as document volumes grew exponentially.

The organisation needed a modern, AI-driven solution to automate extraction, enhance data accuracy, and seamlessly integrate with existing systems all while maintaining cost efficiency and performance reliability.

Objective

Automate document classification and key data extraction from multi-format sources (PDF, Word).

Leverage AI and RAG architecture for domain-grounded document comprehension.

Ensure seamless integration with existing workflows and IT systems.

Improve accuracy, consistency, and processing speed at scale.

Minimise operational costs while maintaining compliance and security.

Build a future-ready foundation for intelligent, scalable document management.

Solution Approach

1. Discovery & Assessment

Conducted detailed workflow analysis to identify pain points and define measurable goals in accuracy, performance, and cost optimisation.

Mapped document formats, metadata structures, and user workflows to define data ingestion and output requirements.

2. AI Agent Design & Development

- Developed an AI Agent powered by a fine-tuned Large Language Model (LLM) to automate document classification and information extraction.
- Enabled intelligent recognition of key entities (names, dates, amounts, identifiers) from both structured and unstructured text.
- Automated metadata tagging and indexing to support quick retrieval and reporting.

3. Integration of RAG (Retrieval-Augmented Generation) Capability

To improve contextual accuracy and domain relevance, a RAG-based pipeline was integrated into the AI Agent.
This enhancement allowed the system to:

- Retrieve relevant contextual data from an enterprise knowledge base before generating outputs.
- Ensure extracted data aligns with company-specific policies, document templates, and industry standards.
- Enable real-time referencing of large document repositories (contracts, reports, case files) to verify extracted information.
- Support auditable, explainable outputs, as every extracted data point could be traced back to its original context.

Technical Highlights:

- Built a vector database using embeddings for document chunks, enabling semantic retrieval across millions of files.
- Combined the LLM's generative capabilities with retrieved source context to ensure accuracy and traceability.
- Deployed a hybrid search (keyword + semantic) mechanism for improved recall in complex document types.

4. Scalable Processing Infrastructure

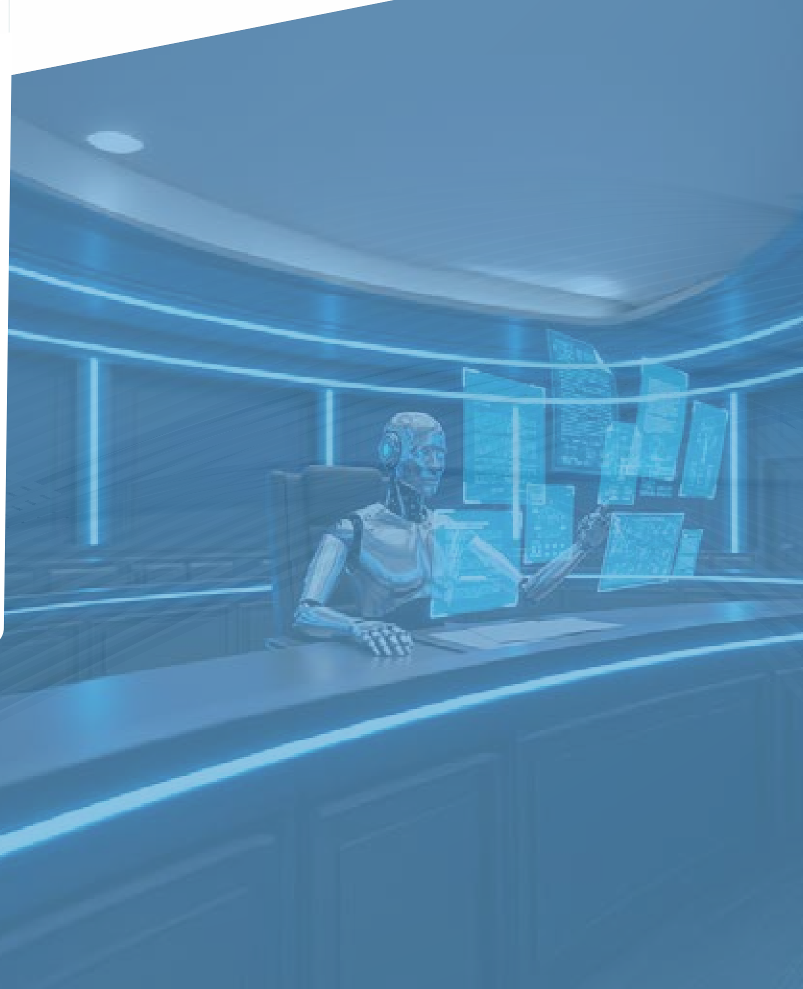
- Engineered a robust data pipeline capable of processing 11.5 million pages.
- Optimised cloud infrastructure for high throughput and cost efficiency.
- Deployed distributed workers to handle parallel processing of large files (>1,000 pages).

5. Performance Optimisation & Validation

- Created a custom evaluation suite to measure extraction accuracy, latency, and consistency across file formats.
- Reduced model token usage through efficient prompt structuring, achieving major cost savings per page.
- Conducted continuous model retraining with edge-case datasets to maintain precision

6. Seamless Integration

- Integrated processed outputs directly into existing ERP and workflow systems.
- Delivered real-time dashboards for monitoring extraction status, accuracy trends, and cost efficiency.
- Achieved deployment with zero downtime and minimal user retraining.



Results & Impact

90% reduction in manual document processing efforts.

80% faster document turnaround times.

11.5 million pages processed in 24 hours with 99% system uptime.

Enhanced accuracy and contextual reliability through RAG-based validation.

Significant cost savings via optimised LLM usage and automated scaling.

Full compliance with enterprise data governance and audit requirements

Key Takeaway

By combining LLM-based automation with RAG-enabled contextual retrieval, the client transformed its document processing workflows into a self-learning, scalable, and cost-efficient AI ecosystem.

This architecture not only accelerated operations but also ensured traceable, explainable, and highly accurate document intelligence at enterprise scale.

stixis®
AI Solutions

www.stixis.ai

Stixis AI Solutions Pvt. Ltd.

#477, Urban Vault, 18th Cross Road,
Parangipalya, Sector-2,
HSR Layout, Bangalore-560102.
Karnataka, India

Crestwood Center
1200 W Walnut Hill Ln,
Suite 2050, Irving TX 75038
U.S.A

Building No 8, Pinewood Square,
Pinewood Office Park, 33 Riley Road,
Woodmead, Johannesburg 2191
South Africa