Revolutionizing Pathology Data Entry with Agentic AI and Google's Handwriting LLM

EXECUTIVE SUMMARY

In pathology services, precision and speed are essential. A mid-sized company with five regional centers faced a bottleneck due to manual data entry from handwritten forms, achieving only 60% accuracy despite two years of digitization efforts. When an IT firm proposed automating their Accounts Payable systems with Agentic AI, the pathology company shifted focus to automating the transcription of handwritten forms. The IT firm accepted the challenge, utilizing Google's Large Language Model for handwriting recognition and Agentic AI for intelligent document processing. The results were transformative:

- Handwritten document readability improved to ~90%
- IDP bot costs reduced by 90%
- Processing became instant, eliminating delays
- Infrastructure and licensing costs significantly decreased
- Data extraction accuracy reached nearly 100%

This case study examines the transition from manual inefficiency to AI-driven precision, highlighting technical challenges, strategic choices, and business impact.

CLIENT BACKGROUND

The pathology company operates five diagnostic centers, each staffed with two data entry personnel. Their primary task is to transcribe patient and test information from pre-printed pathology forms into a proprietary application. One staff member enters the data, while the other verifies it for accuracy.

This process, though reliable, was labor-intensive and slow. The company had attempted to digitize the workflow twice in the past two years using traditional OCR and form recognition tools. However, both attempts failed to meet the required accuracy threshold, resulting in continued reliance on manual entry.

THE CHALLENGE

The pathology forms were handwritten by various medical professionals, each with unique writing styles. Two major challenges emerged:

Tick Marks Spanning Multiple Boxes

Forms often included checkboxes for selecting test types or patient details. A common issue was tick marks that extended across multiple boxes, making it difficult for traditional OCR systems to determine the intended selection. This ambiguity led to frequent misclassifications.

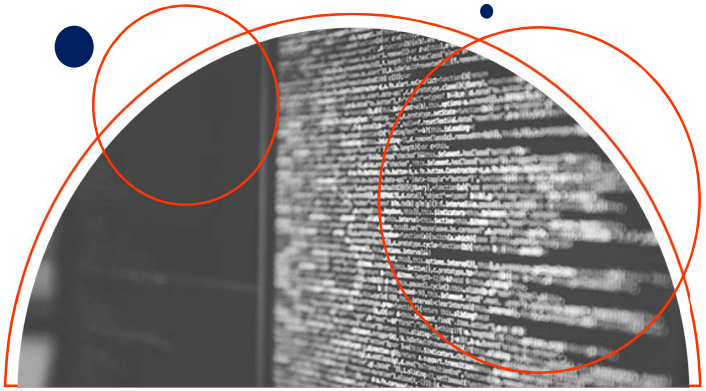
Background Printing Interference

Many fields, such as date entries, had pre-printed text (e.g., “DDMMYYYY”) within the input boxes. When users wrote over these fields, the scanned image included both the handwriting and the background print. This confused earlier models, which couldn’t distinguish between the foreground (user input) and background (template text).

ENGAGEMENT OVERVIEW

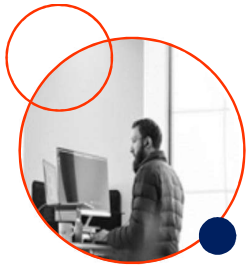
The IT services firm initially engaged with the pathology company to automate their Accounts Payable (AP) systems using Agentic AI. During discussions, the client expressed a more urgent need: digitizing handwritten pathology forms.

Given the firm's ongoing work with LLMs and document intelligence, the team decided to take on the challenge. The client set a clear benchmark—accuracy must exceed 60%, or the project would be shelved.



TECHNICAL APPROACH

Evaluating LLMs for Handwriting Recognition



The team evaluated several LLMs and handwriting recognition models. Google's handwriting LLM stood out due to its:

- High accuracy in recognizing diverse handwriting styles
- Robust training on medical and structured forms
- Ability to contextualize ambiguous inputs using surrounding text

Agentic AI for Intelligent Document Processing

Agentic AI was used to orchestrate the document processing pipeline. Unlike traditional automation, Agentic AI adapts to context, learns from feedback, and makes decisions dynamically. It was instrumental in:

- Resolving checkbox ambiguities using spatial and semantic cues
- Filtering out background noise from pre-printed templates
- Validating extracted data against known patterns (e.g., date formats)

System Architecture

The solution architecture included:

- Document ingestion module: Scans and preprocesses forms
- LLM-based handwriting recognition engine: Extracts text and selections
- Agentic AI orchestrator: Validates, corrects, and routes data
- Integration layer: Pushes data into the pathology application.

EVALUATING LLMS FOR HANDWRITING RECOGNITION

Handwriting recognition has traditionally been the domain of OCR (Optical Character Recognition) systems. However, the rise of Multimodal Large Language Models (MLLMs) has significantly advanced the field, enabling more nuanced and context-aware interpretation of handwritten text. These models combine visual understanding with language modeling, allowing them to process handwriting with greater flexibility and accuracy.

Key LLMs and Models Evaluated

During the evaluation phase, the IT services team considered several leading models:

AMAZON TEXTRACT

*- **Strengths:** Strong layout analysis, good integration with AWS services.*
*- **Limitations:** Moderate performance with highly variable handwriting.*
*- **Use Case Fit:** Considered but not selected due to lower accuracy in benchmark tests.*

GPT-4O (OPENAI)

*- **Strengths:** Multimodal capabilities, strong contextual reasoning, excellent zero-shot performance.*
*- **Limitations:** Requires careful prompt engineering; not optimized for structured form layouts.*
*- **Benchmark Performance:** Among the top performers in handwriting benchmarks, especially in free-form text.*

CLAUDE 3.5 SONNET (ANTHROPIC)

*- **Strengths:** Exceptional zero-shot performance, strong in modern handwriting recognition.*
*- **Limitations:** Less effective in structured form parsing and layout-sensitive tasks.*
*- **Benchmark Insight:** Outperformed open-source models in zero-shot settings, particularly for English handwriting.*

AMAZON TEXTRACT

*- **Strengths:** Strong layout analysis, good integration with AWS services.*
*- **Limitations:** Moderate performance with highly variable handwriting.*
*- **Use Case Fit:** Considered but not selected due to lower accuracy in benchmark tests.*

MICROSOFT AZURE COMPUTER VISION API

*- **Strengths:** Good general OCR capabilities, scalable.*
*- **Limitations:** Lower performance in handwriting benchmarks compared to Google and GPT-4o.*

TRANSKRIBUS (OPEN SOURCE)

*- **Strengths:** Specialized in historical handwriting, customizable models.*
*- **Limitations:** Requires extensive training data and manual annotation.*
*- **Benchmark Insight:** Comparable to LLMs in some cases, but lacked the flexibility and scalability needed for this project.*

BENCHMARKING INSIGHT

A 2025 benchmark study compared these models using 50 handwritten samples from five different writers. The findings revealed:

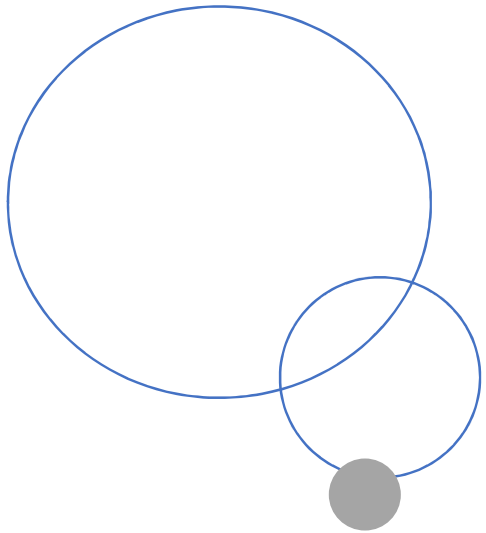
- GOOGLE CLOUD VISION API, GPT-4O, AND CLAUDE 3.5 SONNET WERE THE TOP PERFORMERS.
- AVERAGE RECOGNITION ACCURACY ACROSS MODELS RANGED FROM 64% TO 90%, DEPENDING ON HANDWRITING COMPLEXITY AND LAYOUT.
- LLMS DEMONSTRATED SUPERIOR ADAPTABILITY TO HANDWRITING VARIABILITY, ESPECIALLY IN ZERO-SHOT

WHY GOOGLE'S LLM WAS CHOOSEN

The decision to use Google's handwriting LLM was based on several factors:

- HIGH ACCURACY IN STRUCTURED FORM RECOGNITION
- EFFECTIVE HANDLING OF BACKGROUND INTERFERENCE (E.G., PREPRINTED DATE FORMATS).
- STRONG SUPPORT FOR LAYOUT-SENSITIVE PARSING, CRUCIAL FOR CHECKBOX INTERPRETATION.
- SCALABILITY AND INTEGRATION WITH EXISTING CLOUD INFRASTRUCTURE.

This model's ability to distinguish between foreground handwriting and background template text was particularly valuable in overcoming one of the project's core challenges.



IMPLEMENTATION

Training and Fine-Tuning

The team fine-tuned the Google LLM using a curated dataset of scanned pathology forms. Special attention was given to:

- Variability in handwriting styles
- Common tick mark patterns
- Background print interference

Handling Tick Mark Ambiguity

Agentic AI used spatial analysis and historical data to resolve ambiguous tick marks. For example, if a tick spanned two boxes, the system considered:

- The center of the tick
- Historical frequency of selections
- Semantic relevance of adjacent fields

Background Print Filtering

To address background interference, the team implemented a layered image processing technique:

- Layer 1: Isolated handwriting using contrast and stroke detection
- Layer 2: Suppressed background print using template matching
- Layer 3: Applied LLM interpretation only on the filtered layer

RESULT

Deployment Results: AI Impact Summary

Metric	Before AI	After AI
Accuracy of data extraction	<60%	~100%
Readability of handwritten forms	~50%	~90%
Processing time	Scheduled (batch)	Instant
IDP bot cost	High	Reduced by 90%
Infra & licensing cost	High	Significantly reduced
Staff redeployment	Not possible	Fully redeployed

The system was rolled out across all five locations. Staff previously engaged in manual entry were redeployed to higher-value tasks such as quality control and patient engagement.

IMPACT

Operational Efficiency

Instant processing replaced batch jobs, enabling real-time data availability and improving turnaround times for test results and workflows.

Cost Savings

Agentic AI and LLMs eliminated costly OCR licenses, reducing IDP bot costs by 90% and enhancing scalability.

Workforce Optimization

Automation of data entry allowed staff to focus on roles that improved service quality and customer satisfaction, boosting morale and agility.

Strategic Value

The project's success positioned the pathology company as a pioneer in AI in healthcare diagnostics, enabling further automation initiatives.

LESSON

Importance of Model Selection

Choosing the right LLM was critical. Google's model, with its robust handwriting capabilities, outperformed others in handling medical forms.

Adaptive Intelligence Matters

Agentic AI's ability to adapt to context and learn from feedback was key to resolving ambiguities and improving accuracy.

Pre-processing Is Crucial

Effective preprocessing—especially background filtering—was essential for accurate data extraction. This step often determines the success of the entire pipeline.

Collaboration Drives Innovation

Close collaboration between the IT services team and the pathology company ensured that domain-specific challenges were addressed effectively.

CONCLUSION

This case study demonstrates how Agentic AI and LLMs can revolutionize document processing in healthcare. By overcoming the challenges of handwriting variability and background interference, the IT services firm delivered a solution that exceeded expectations. The pathology company now enjoys:

- Near-perfect data accuracy
- Real-time processing
- Significant cost savings
- Redeployment of staff to strategic roles

This success story underscores the transformative potential of intelligent automation and sets a benchmark for future AI-driven initiatives in healthcare and beyond.