



POSITION PAPER



Use of AI Techniques in Intelligent Document Processing

Table of Contents

The Role of AI in Modern Intelligent Document Processing	03
The LLM Question	04
The question is not only whether LLMs are secure, it is also whether they are necessary.	05
Using the Right AI for the Job	06
Building an IDP Demonstration is easy. Building a proven AP Platform is not.	07
Purpose-built Document Intelligence	08
Purpose-Built AI That Learns	09
Layered Intelligence	10
Accuracy requires more than an AI answer	11
Questions Every Organization Should Ask	12
More Than a Decade of Development. More Than 2,000 Customers.	13
About ancora Software	14



The Role of AI in Modern Intelligent Document Processing



Modern Intelligent Document Processing (IDP) is far more complex than simply extracting information from a page. Today's enterprise environments require flexible, multi-featured systems capable of handling unpredictable inputs, multiple document types, diverse workflows, and highly customized outputs.

Documents may arrive in paper form although increasingly rarely or electronically through email, web portals, or other channels. Formats can range from TIFF, JPEG, and PNG to PDF (both image based and vector), XLS/XLSX, and DOC/DOCX. Documents may also be embedded directly in the body of an email.

In many cases, multiple documents arrive together as a single file. In others, each document arrives as an individual file. Documents may contain attachments that must be identified and processed appropriately, while in other circumstances attachments must be excluded. The system may need to identify and classify documents, separate them from one another, ignore attachments, or process only a specific subset of pages.

The workflow itself varies according to the business process. Purchase-order-based invoices, service invoices, statements, purchase orders, and business forms can each require different processing logic. In many cases, the IDP system must communicate with downstream applications typically an ERP or financial system for validation, enrichment, matching, and field population.

Enterprise deployments also require multi-tenancy, security, user management, scalability, auditability, and the ability to support multiple organizations and workflows simultaneously. Consequently, capturing the content of a document page represents only one relatively small component of the overall production workflow.

ancora has developed and provides a complete ecosystem for processing documents that includes data capture and versatile workflows and supports both on-premises and cloud installations with secure customer data handling.

The LLM Question



The rapid emergence of publicly available Large Language Models (LLMs) has created an understandable perception that a single technology can address virtually every aspect of document processing.

LLMs are extraordinarily capable technologies. They have opened important new possibilities across enterprise applications, including document understanding, summarizing, classification, conversational interfaces, and knowledge management.

However, technological capability does not necessarily mean technological necessity.

This raises an important question for finance and technology leaders.

“LLMs are impressive, our position is not that LLMs are inherently insecure. The question is much simpler: Why does my sensitive invoice data need to go through a third-party LLM in the first place?”

Noel Flynn
CEO at ancora Software, Inc.

If highly accurate invoice extraction can be achieved without sending sensitive financial information through a third-party generative LLM, why introduce that additional processing layer?

ancora believes that LLMs have an important role to play in enterprise technology. But the newest technology is not necessarily the right technology for every business process, particularly when processing invoices containing exact and often highly confidential information such as negotiated pricing, discounts, supplier relationships, banking information, payment terms, and contractual terms.

The question is not only whether LLMs are secure, it is also whether they are necessary.



Leading enterprise LLM providers have implemented security, privacy, and tenant-isolation controls. When appropriately configured, enterprise services can provide some protections against one customer's information being exposed to another and may prevent customer information from being used to train shared foundation models.

Those safeguards are important.

But they do not answer the more fundamental question for Accounts Payable leaders: once invoice data enters that third-party layer, where does it go, who can see it, and is the layer necessary in the first place? (See "Questions Every Organization Should Ask" below for the full list of items worth raising with any vendor.)

"LLMs are impressive technology, and our position is not that LLMs are inherently insecure. The question is much simpler: Why does my sensitive invoice data need to go through a third-party LLM in the first place?", said Noel Flynn, CEO of ancora Software.

Using the Right AI for the Job

The strength of the ancora approach is its use of complementary AI technologies rather than dependence on a single technique.

Different AI approaches have different strengths. Combining purpose-built machine learning, automated learning, document intelligence, and language technologies can produce a system better suited to the specific requirements of transactional document processing than relying on one general-purpose model for every task.

The objective is not to reject AI.

It is to use the right AI for the job.

Building an IDP Demonstration is easy.

Building a proven AP Platform is not.

The rapid advancement of generative AI and LLMs has significantly lowered the technical barrier to demonstrating document extraction capabilities. It has also contributed to an increasingly crowded IDP market, with more than 300 providers now tracked across the broader global IDP landscape according to industry analyst coverage.

An LLM can make it possible to demonstrate document extraction in minutes or hours rather than the weeks or months traditionally required to build an initial prototype.

But demonstrating that a system can extract information from an invoice is fundamentally different from building an enterprise platform capable of processing financial documents reliably, securely, and at scale every day.

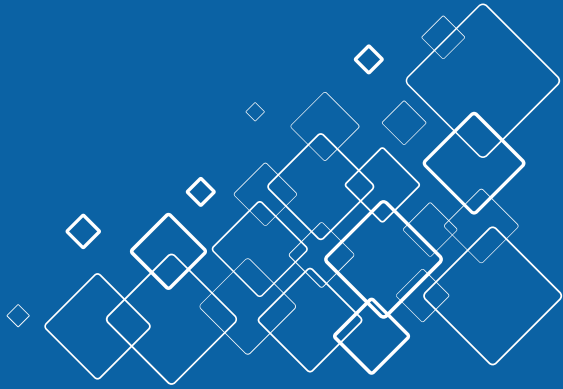
Production Accounts Payable automation requires much more than identifying fields on an invoice.

It requires:

- High and consistent accuracy.
- Continuous learning.
- Validation and business-rule enforcement.
- Complex line-item extraction.
- Vendor-specific intelligence.
- Document classification and separation.
- Exception management.
- Security and tenant isolation.
- Auditability and traceability.
- Scalability and reliability.
- Workflow orchestration.
- Integration with ERP and financial systems.
- Support for real-world variations in document formats and business processes.



Purpose-built Document Intelligence



ancora's purpose-built document intelligence reflects more than ten years of focused engineering on exactly these challenges, including patented unassisted and assisted machine learning technologies built specifically for document processing rather than adapted from a general-purpose generative LLM.

"LLMs have made it easier than ever to demonstrate document extraction. That's not the same thing as building an enterprise AP automation platform.", Flynn adds.

"We've solved the difficult parts: accuracy, learning, validation, line-item extraction, vendor-specific intelligence, security, scalability, workflow and integration into real-world financial processes."



Purpose-Built AI That Learns



Rather than requiring a general-purpose generative LLM for core invoice extraction, ancora uses patented unassisted and assisted machine learning technologies and purpose-built document intelligence to identify, extract, learn from, and validate information contained in invoices and other business documents.

An important element of this approach is continuous, customer-specific learning.

When a user corrects an extracted value, the system can remember that correction and apply what it has learned when another invoice from the same vendor is received.

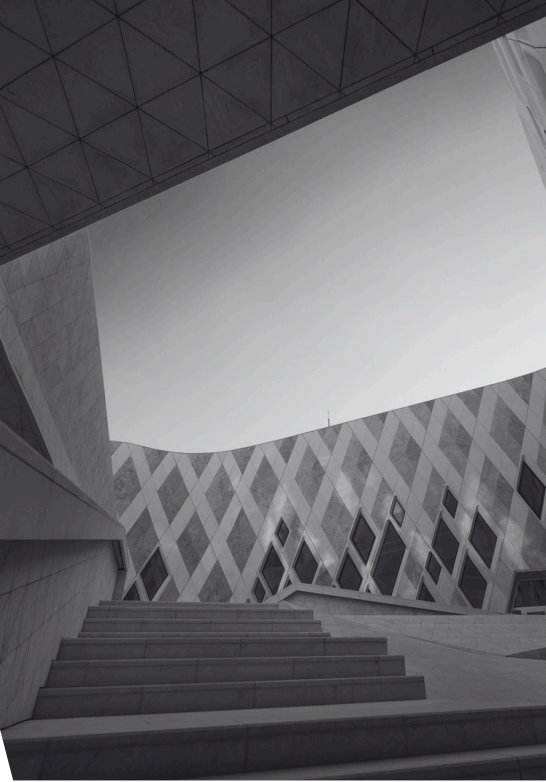
The correction therefore does more than fix a single invoice.

It becomes part of the customer-specific and vendor-specific intelligence that can improve future processing and reduce repetitive human intervention.

Correct it once, if necessary, and apply what was learned the next time.

Layered Intelligence

ancoraFusion further extends this approach through a layered architecture that combines universal intelligence, customer specific intelligence, and vendor specific intelligence with a small language model (SLM) designed specifically for document understanding.



This architecture reflects ancora's fundamental position:

AI versus no AI is the wrong question.
The right question is which AI is
appropriate for the job.

By using technologies specifically designed for transactional document processing, ancora provides organizations with an alternative to architectures that require invoice content to be sent to third-party generative LLMs for core data extraction.

The result is a self-contained, continuously improving AI system designed specifically for document processing.



Accuracy requires more than an AI answer

LLMs are probabilistic by nature. That characteristic contributes to their extraordinary power in conversational, generative, and reasoning applications.

Financial document processing presents a different challenge.

An invoice total cannot be approximately correct.

A purchase order number cannot be approximately correct.

Supplier banking information cannot be approximately correct.

Accounts Payable automation requires information to be accurate, validated, consistent, traceable, and auditable.

The objective is therefore not simply to produce an answer that appears plausible. The objective is to produce reliable financial data that can confidently move through downstream business processes.

"Our objective isn't simply to produce an answer," says Flynn. "It is to produce accurate financial data that can be validated, traced, and confidently passed into downstream AP and ERP processes. And when human intervention is necessary, the system should learn from it so the user isn't correcting the same issue over and over again."

Questions Every Organization Should Ask

Organizations evaluating IDP and Accounts Payable automation platforms should consider asking every prospective vendor, including ancora, the following questions:

- Does the platform send our invoices or extracted information to a third-party generative LLM?
- What specifically does the model do, and can core invoice extraction operate without it?
- Where is our invoice information processed, stored, and retained?
- Is our information ever used to train, fine-tune, optimize, or improve models or services?
- How is our information isolated from other customers?
- When our users correct an extraction, does the system remember that correction and apply what it learned to future invoices from that vendor?
- How is AI-generated output validated before information reaches our ERP or payment process?
- Can every extracted value be traced and audited?
- What independent security and compliance credentials support the platform?
- What happens to document processing if a third-party AI service becomes unavailable?
- What execution time to be expected for each image processed?
- What costs is the organization to incur per image processed by the LLM?

"We encourage prospects to ask these questions of every vendor, including ancora," Flynn adds. "The issue isn't whether a company can put an LLM into invoice processing. The question is whether doing so provides enough additional business value to justify the additional dependency, processing layer, and cost."



More Than a Decade of Development.

More Than 2,000 Customers.

The difference between a promising technology and a proven enterprise platform is ultimately demonstrated in production, not in a demo. Here is what that looks like at scale:

More than 2,000 customers globally use ancora technology today, and the company's technology processes more than \$50 billion in annualized invoice transactions across its customer base, based on ancora's internal transaction data.

ancora combines patented unassisted and assisted machine learning, automated learning, customer-specific and vendor-specific intelligence, validation, and workflow to automate invoice processing without requiring third-party generative LLMs for core invoice extraction.

With ancora's Intelligent Document Processing technology and ancoraFlow, organizations can automate the invoice lifecycle from ingestion and extraction through validation, matching, exception handling, approvals, GL coding, and ERP integration.

ancora's capture platform can also be deployed independently. This allows organizations and software partners with established AP workflow platforms to use ancora's Intelligent Document Processing technology without being required to purchase a bundled workflow solution.

Contact Us



1.844.6ANCORA
+44 01752 875655



sales@ancorasoftware.com



About ancora Software

ancora Software specializes in intelligent document classification and data capture solutions. Founded in 2013, we have been at the forefront of developing advanced technologies to automate and optimize document-driven processes for businesses.



With a mission to simplify information capture and management, we have continuously evolved our solutions to meet the growing needs of organizations across various industries. We have built a strong reputation for our expertise in document classification, data extraction, and workflow automation.

Over the years, ancora Software has established itself as a trusted provider of intelligent data capture solutions that leverage cutting-edge technologies such as machine learning, natural language processing, and optical character recognition (OCR). By harnessing the power of these technologies, we enable businesses to extract valuable information from documents with speed, accuracy, and efficiency.

Our solutions have found applications in diverse sectors, including finance, healthcare, manufacturing, and more. By automating manual document processing tasks, we help businesses streamline operations, reduce errors, enhance compliance, and ultimately improve overall productivity and customer satisfaction.

Through continuous innovation and a customer-centric approach, we have grown our customer base and established strong partnerships with organizations worldwide. We remain committed to delivering industry-leading solutions that empower businesses to unlock the potential of their data, optimize processes, and drive digital transformation.