Docxtract

A DOCXTRACT WHITE PAPER

# Closing the Last-Mile Gap

Why document intelligence is the missing layer in enterprise AI orchestration.


PREPARED BY
Docxtract

PUBLISHED
July 2026

AUDIENCE
CIOs, CTOs & Automation Leaders

EXECUTIVE SUMMARY

# Orchestration is scaling faster than the data feeding it.

Enterprises are pouring budget into AI orchestration. The market for AI-driven automation is on pace to reach $169 billion in 2026 alone, and by year's end an estimated 40% of enterprise applications will carry task-specific AI agents, up from under 5% in 2025. Boards are approving orchestration platforms, agent frameworks, and workflow engines at a pace that would have seemed reckless three years ago.

Yet a growing body of research says the orchestration layer is not where enterprise automation is stalling. The constraint sits one level down, in what those orchestration engines are being asked to run on: roughly 80% of enterprise data is unstructured — invoices, contracts, claims forms, purchase orders, PDFs and scans that no workflow engine can act on without help getting turned into data first. Stonebranch's 2026 survey of 402 automation leaders found that the share of organizations with streamlined, AI-integrated workflows across business functions actually fell from 30% to 16% year over year, even as automation investment rose. Orchestration platforms are proliferating faster than the data feeding them is getting clean.

This paper argues that document intelligence — the ability to reliably turn unstructured documents into structured, governed data — is the missing layer between an enterprise's AI ambitions and its orchestration spend. Without it, orchestration platforms coordinate empty pipelines. With it, every downstream agent, workflow, and ERP integration gets trustworthy input to act on. We lay out why this gap exists, what it costs, and a practical reference model for closing it.

**80%**
of enterprise data is unstructured

**16%**
of orgs have AI-integrated workflows enterprise-wide in 2026, down from 30% in 2025

**60%**
of AI failures trace to governance gaps, not model performance

CONTENTS


01 The Orchestration Boom and Its Blind Spot 04

02 The 80% Problem: Unstructured Documents Are the Bottleneck 05

03 Why Orchestration Platforms Stall at the Document Layer 06

04 Document Intelligence as the Missing Layer 07

05 A Reference Model: From Document to Orchestrated Action 08

06 The Business Case 09

07 Conclusion & Recommendations 10

01 — THE BLIND SPOT

# The Orchestration Boom and Its Blind Spot

2026 is the year orchestration went from IT hygiene to board-level strategy. Stonebranch's annual survey of automation leaders shows why: 88% of organizations now run hybrid IT environments, 64% are investing in cloud automation, and 50% are investing in workload automation and service orchestration platforms (WLA/SOAP). Sixty-nine percent of organizations planning a platform change now cite functionality and modernization, not cost, as the driver, up 21 points since 2025.[3] Orchestration has become the control plane enterprises are betting on to make AI operational.

**But the same data shows the strategy is running ahead of execution. Eighty-nine percent of enterprises already run multiple WLA/SOAP platforms (three is the most common count), and 78% plan to add or replace one in the near term.[3] That is not consolidation; it is proliferation. And despite the investment, only 21% of organizations have reached enterprise-wide production for AI workflows.[3] More platforms are being bought to solve a problem that more platforms cannot solve on their own.**

The reason is structural. Orchestration engines, workflow automation tools, and agent frameworks are built to move and coordinate data that is already structured and addressable: API payloads, database records, job triggers, message queue events. They are exceptional at deciding what happens next once they have a clean fact to act on. They are not built to read a scanned invoice, a signed contract, or a handwritten claims form. That gap doesn't show up in an orchestration platform's demo. It shows up the moment real enterprise paperwork hits the pipeline.

02 — THE BOTTLENECK

# The 80% Problem: Unstructured Documents Are the Bottleneck

Most of the friction in enterprise AI is not a model problem. It is a document problem. Gartner's widely cited estimate that up to 80% of enterprise data is unstructured has held up for years, but the scale is compounding: Komprise's 2026 survey found 74% of IT and data leaders now manage at least 5 petabytes of unstructured content, a 57% increase over 2024.[4] Separately, 45% of companies cite fragmented, unstructured data as a top AI challenge, while overall business data volume grows roughly 63% a month, up to 90% of it unstructured.

**The consequence shows up in project outcomes, not just infrastructure metrics. Gartner research indicates that roughly 30% of AI and generative AI projects will be abandoned or fail to scale, primarily due to poor data quality, governance gaps, or unclear business value.[5] Separately, 84% of organizations say they are betting big on AI, but only 17% report it has actually succeeded. Most recent analysis puts an even sharper number on the governance piece: by 2026, 60% of AI failures will trace back to governance gaps, not model performance.[6] Enterprises are not failing because the AI can't reason. They are failing because the AI is being asked to reason over data nobody has validated, structured, or traced.**

Invoices, purchase orders, contracts, claims, onboarding forms, remittance advices: these are exactly the categories that carry the highest volume and the highest variance in format, and they sit at the center of the highest-value automation targets — accounts payable, procurement, claims processing, customer onboarding. They are also precisely the content orchestration platforms were never designed to parse.

03 — THE STALL POINT

# Why Orchestration Platforms Stall at the Document Layer

Orchestration and workflow engines assume a clean, structured, addressable input: a field, a record, an API response. When a document enters a process, it typically forces one of three costly detours: a person keys the data in manually, a brittle point-to-point OCR script tries to catch it and breaks the next time a vendor changes their invoice template, or the item sits in a queue until someone with the right context resolves it by hand. None of these are orchestration failures. They are the absence of a layer that should have handled the document before orchestration ever saw it.

## Where the gap actually bites

Multi-platform complexity compounds the problem: 89% of enterprises already juggle multiple automation platforms, and each one typically expects clean inputs formatted its own way.

Even fully structured file movement struggles with integration depth: 90% of organizations connect managed file transfer tools to their automation platforms, but only 49% report those connections are fully integrated.[7] Unstructured content is a harder problem than file transfer.

Governance is the layer most exposed: without extraction that carries lineage, confidence scores, and an audit trail, a document-derived data point entering an orchestrated workflow is a compliance liability, not an asset. This is consistent with governance, not model quality, now driving the majority of AI failures.

None of this means orchestration platforms are the wrong investment. It means they are being asked to do a job — reading and trusting a document — that sits outside their design, and the mismatch is exactly where automation programs quietly stall.

04 — THE MISSING LAYER

# Document Intelligence as the Missing Layer

Document intelligence — what the market broadly calls intelligent document processing, or IDP — is AI purpose-built to read, classify, extract, validate, and structure content from documents, then hand that content to orchestration and ERP systems as clean, governed records. Where orchestration coordinates the middle mile of an enterprise workflow, system to system, step to step, document intelligence covers the last mile: paper and PDF to structured, trustworthy data.

**This is not a niche capability. The IDP market is projected to grow at roughly a 26–29% compound annual rate through the early 2030s, and finance and accounting already account for the largest functional share of that market, an estimated 45.6% in 2026,[7] because accounts payable, procurement, and invoicing are simultaneously the highest-volume document workloads and the most ERP-adjacent. It is the natural first place enterprises close the gap between paperwork and orchestrated action, and the pattern generalizes to claims, contracts, and onboarding wherever documents feed a governed workflow.**

THE DISTINCTION THAT MATTERS

Document intelligence is not a substitute for orchestration, and it is not a better OCR tool. It is the layer that makes every dollar already committed to orchestration and agent platforms actually usable, because those platforms can only be as reliable as the data arriving at their front door.

05 — THE REFERENCE MODEL

# A Reference Model: From Document to Orchestrated Action

Enterprises that get this right treat document intelligence as infrastructure that sits ahead of orchestration, not a point solution bolted onto one process. A simple four-stage model:

| Stage | What Happens | Risk It Removes |
|---|---|---|
| **1. Capture & Classify** | Documents ingested from email, scan, portal, or upload; automatically identified by type (invoice, contract, claim, etc.). | Eliminates manual sorting and routing delays. |
| **2. Extract & Validate** | AI reads fields and line items across varying layouts and vendors; results are checked against business rules. | Removes keying errors and format-driven breakage. |
| **3. Structure & Govern** | Validated data is normalized into structured records with confidence scores, lineage, and an audit trail. | Closes the governance gap behind most AI failures. |
| **4. Route & Orchestrate** | Clean records are pushed via API or connector (e.g., SAP, Oracle, Zoho, UiPath, Power Automate) into ERP and orchestration platforms to trigger the next step. | Gives orchestration and agents a trustworthy input to act on. |

This is where Docxtract sits by design: an AI-powered document intelligence layer that reads and structures data from any business document, then delivers it into ERP and orchestration systems through pre-built connectors or a standard API, so the platforms an enterprise has already invested in receive governed data instead of a scanned PDF.

06 — THE BUSINESS CASE

# The Business Case

For a C-suite evaluating where to spend the next automation dollar, the case for treating document intelligence as foundational infrastructure comes down to three arguments.

### It protects the orchestration investment already made

Every WLA/SOAP platform, agent framework, or ERP workflow an enterprise has bought is only as productive as the data reaching it. Fixing the document layer is the fastest way to raise the return on automation spend that has already been approved, without adding another orchestration platform to the 89% that already run more than one.

### It targets the specific failure mode enterprises are experiencing

With governance gaps, not model performance, now cited as the source of roughly 60% of AI failures, and generative AI project abandonment running near 30%, the highest-leverage fix is upstream: structured, validated, auditable data before it ever reaches an agent or workflow. Document intelligence is where that lineage and validation is created in the first place.

### It has a proven, ERP-adjacent starting point

Finance and accounting's roughly 46% share of the IDP market is not an accident. Accounts payable, procurement, and invoicing are high-volume, document-heavy, and directly tied to ERP systems the business already depends on. It is the lowest-risk, fastest-payback place to prove the model before extending it to claims, contracts, or customer onboarding.

07 — CONCLUSION & RECOMMENDATIONS

# Conclusion & Recommendations

Enterprise AI orchestration is not overhyped. The coordination layer enterprises are building is genuinely necessary. But the data in this paper points to a consistent conclusion: orchestration investment is outrunning the ability to feed it trustworthy data, and that gap, not model capability, is the reason so many automation programs stall short of enterprise-wide scale.

Four recommendations for leadership teams evaluating their AI and automation roadmap:

1. **Audit unstructured document volume before the next orchestration purchase.** Know how much of the target process runs through documents before buying another coordination layer to sit on top of it.

2. **Treat document intelligence as infrastructure, not a point tool.** It should sit ahead of every orchestration and ERP integration, not be bolted onto one workflow at a time.

3. **Start where document volume and ERP dependency are both highest.** Accounts payable, procurement, and claims are the proven entry points.

4. **Demand lineage and governance, not just extraction accuracy.** Given that governance gaps now drive the majority of AI failures, an extraction layer without an audit trail recreates the problem it was meant to solve.

Orchestration decides what happens next. Document intelligence decides whether what happens next can be trusted. Enterprises that fund both, in the right order, are the ones that will move from AI pilots to AI at scale.

ABOUT DOCXTRACT

Docxtract is an AI-powered intelligent document processing platform that extracts structured data from business documents and integrates it directly into ERP and orchestration systems, closing the gap between unstructured paperwork and automated enterprise workflows.

**Book a demo →**

**docxtract.io | connect@docxtract.io**



Docxtract