



# Build versus Buy

Why buying a best of breed IDP platform is the [best] investment

# Table of Contents

P. 04

1\ TCO Executive Summary

P. 07

2\ Dark Green Costs vs Light Green Benefits

P. 10

3\ The Bigger Picture of IDP Success

P. 13

4\ Determining the Technical Infrastructure

P. 15

5\ Understanding Business Factors Affecting the Total Cost of Ownership (TCO) for AI

P. 17

6\ Unique Agentic Approach to Extraction

P. 19

7\ Conclusion

P. 20

8\ Appendix



# Introduction

// Organizations today are inundated with choices when it comes to automating document-heavy processes—ranging from decades-old OCR solutions to an ever-expanding menu of hyperscaler services and AI-powered tools. Against this backdrop, boards, CEOs, and government leaders alike are pressing for clarity: What is our AI strategy? How do we deploy AI to disrupt rather than be disrupted? How fast can these investments deliver measurable ROI?

For most enterprises, the fastest path to meaningful AI outcomes begins not with bold experiments but with their back-office and the documents at the center of their most critical operations. By approaching Intelligent Document Processing (IDP) as a strategic priority, organizations can lay the groundwork for AI success—streamlining document-heavy workflows, cutting manual effort, improving the speed and quality of customer interactions, and delivering the structured data foundation that GenAI depends on.

## Who is Hyperscience

*Hyperscience* offers an Intelligent Document Processing (IDP) platform that manages the entire document lifecycle from ingestion and pre-processing to accurate extraction, human validation, and seamless downstream orchestration. Reducing document processing cycle times with AI brings benefits beyond improved efficiency: organizations also see increased employee satisfaction, lower attrition, improved customer experiences, revenue growth, and better business outcomes.

## Purpose of this Paper

This white paper is designed to help you evaluate two distinct approaches for addressing your IDP requirements:

## BUILD



1 / Build a platform leveraging tools and services from a hyperscaler.

## BUY



2 / Buy a customizable ML-native document processing platform, such as *Hyperscience*.

## TCO Executive Summary

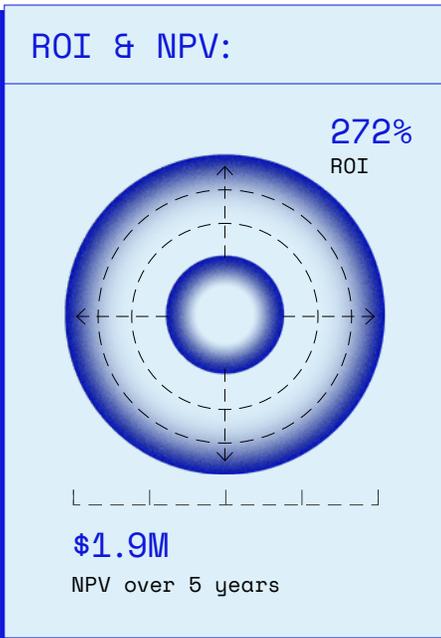
To measure the comparative costs and benefits of the hyperscaler-built solutions versus *Hyperscience*, we developed a 5-year Total Cost of Ownership (TCO) model.

### Key Conclusion

*Hyperscience* delivers a **272% ROI** with a Net Present Value of **\$1.9M** in benefits over a five year period, when compared to building an IDP platform from scratch with a hyperscaler.

### Solutions Evaluated

This model contrasts the value of the unified, out-of-the-box platform from *Hyperscience* against the build-it-yourself approach required by 3 hyperscaler offerings:



	Textract, Bedrock, A2I, Lambda and additional services
	Document AI, Vertex AI, Cloud Functions, Gemini, Custom Data Extractor, and other add-ons
	Azure AI Document Intelligence, Azure Machine Learning, Microsoft Purview, Microsoft Power Automate
<b>(h[s])</b>	Hyperscience Hypercell

<b>1M</b> pages
<b>5</b> Years
<b>TCO 10%</b> Discount Rate
<b>MEDIUM</b> complexity

### Key Assumptions

For simplicity, we analyzed the published model architectures of each of these three hyperscalers, as they map their architecture for IDP to the stages of Document Ingestion and Text Extraction, Document Classification, Entity Extraction and Document Enrichment, and Results Validation.

We studied the hyperscaler offerings, paying special attention to:

- The components used at each stage of the document processing pipeline
- The cost per page data of using each service component
- A representative blend of average cost per page for an end-to-end IDP solution built on a hyperscaler

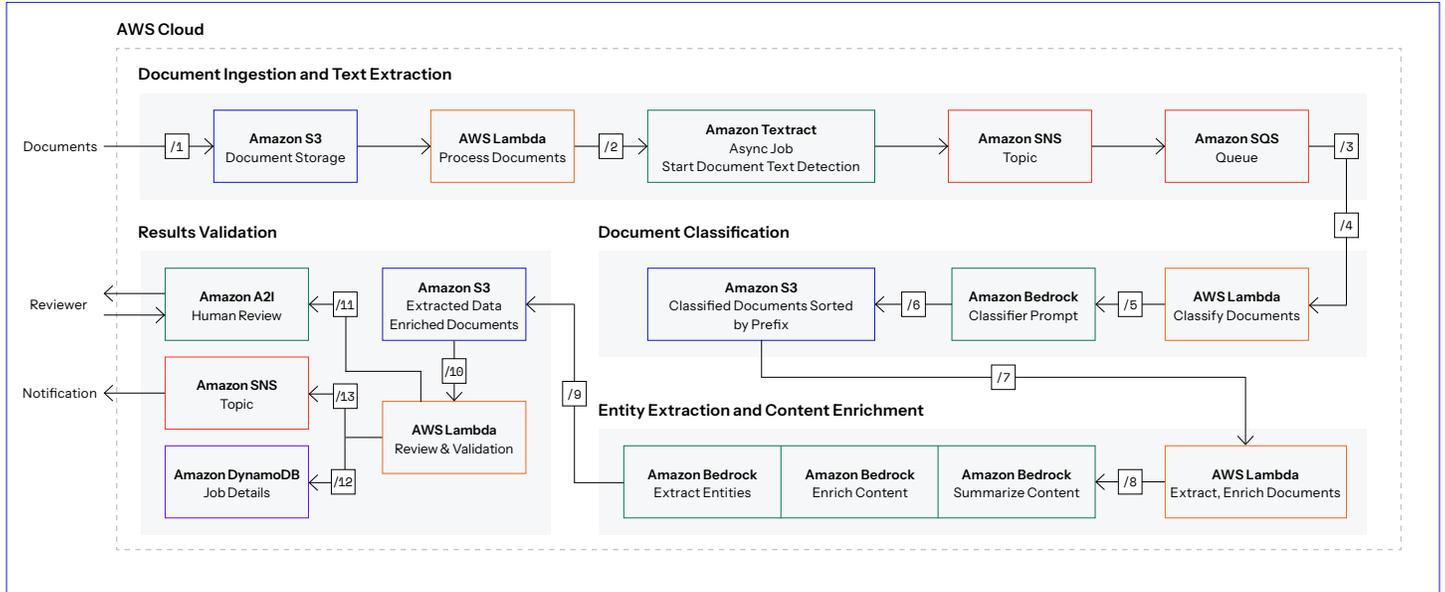
### Document Volume

The TCO model assumes that the IDP system is managing a volume of 1M document pages per year.

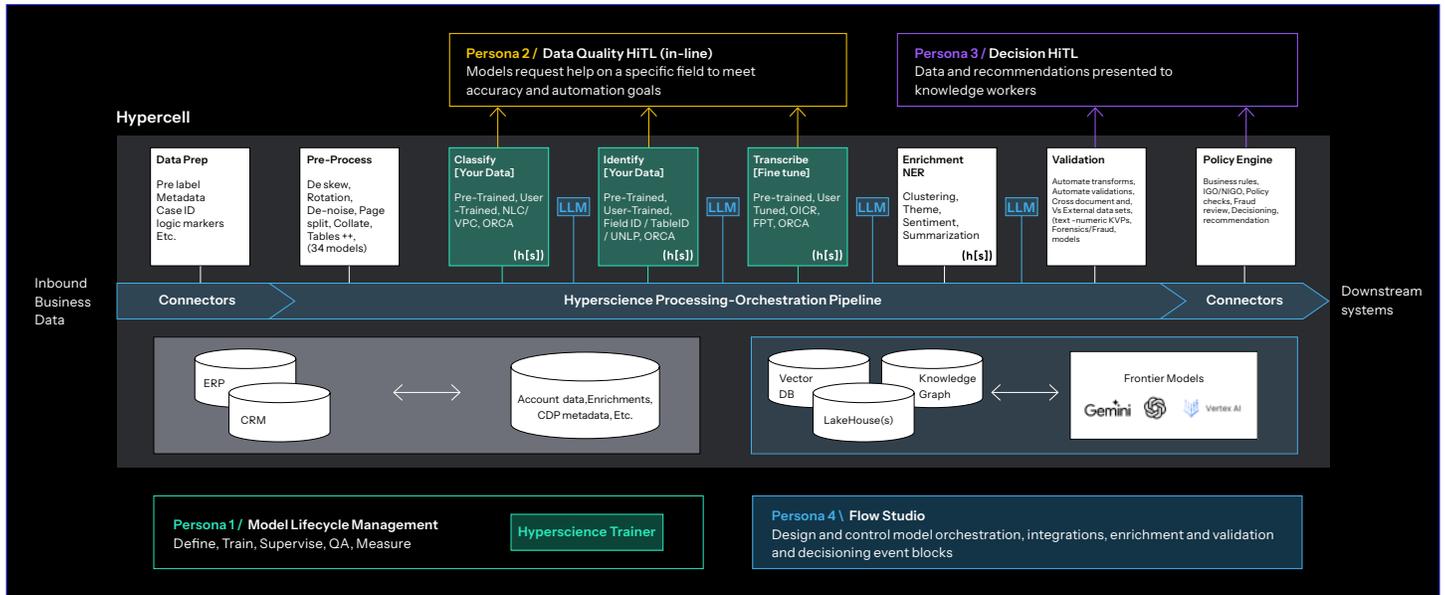
## Deployment

The *Hyperscience* Hypercell is available for deployment as on-premises, hybrid, SaaS, and FedRAMP High. For the purposes of this study, we are exclusively comparing *Hyperscience's* proprietary SaaS offering deployed on hyperscaler infrastructure to a solution built using hyperscaler document processing APIs and IDP services.

## AWS Architecture



## Hyperscience Architecture



## Use Case Complexity

The model assumes a medium complexity use case for Document Ingestion, Classification, Entity Extraction and Document Enrichment, and Results Validation. Medium complexity is a construct used to compare IDP tasks across different platforms, by comparing the number of times a model needs to be prompted or invoked, to execute a specific task.

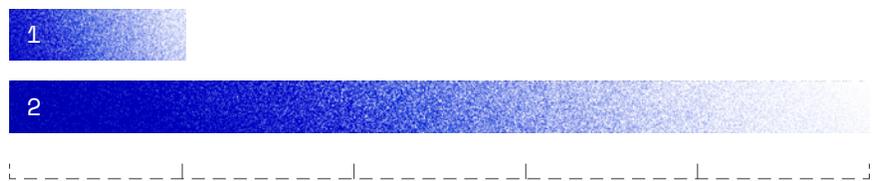
This allows us to compare cost and operational execution between *Hyperscience's* collection of ML models and approaches (ranging from small models to LLMs or VLMs [vision language models]) to those AI components used by hyperscalers for IDP. It is important to note that this use case represents a full end-to-end document processing pipeline, not simply text extraction.

### Technical Labor Assumptions

The model assumes specific technical labor costs for key staff used for project deployment and maintenance, including infrastructure engineers, machine learning model engineers, security engineers, and project managers. These costs may vary based on the specifics of each labor market where the services can be procured, and are detailed in the *Analysis of Benefits* section. For the purposes of this study, all labor cost assumptions are in-house staffing, not offshoring or nearshoring.

### Timeframe

The model distinguishes between two cost timelines:



**1 / Initial Setup** - costs involved in getting the system implemented & functional.

**2 / Ongoing Maintenance** - collective costs for years 1-5 post go-live to maintain performance, compliance, and ongoing evolution of the IDP project.

Some notable exemptions of IDP platform capabilities we have not expressly scoped for building with a Hyperscaler include:



**Full Human-in-the-Loop (HITL)** capabilities for document review and iterative model refinement.



**Full Model Lifecycle Management** capabilities to train, monitor, QA, and improve trained model performance over time and handle Document Drift Management.



**Accuracy Harnessing** to threshold the required accuracy level and flag HITL when each model is uncertain of the result provided in a specific identification, classification, or extraction step.

*Hyperscience Hypercell* includes these capabilities out of the box, but a full analysis of building comparable functionality with hyperscalers is out of the scope of this study. Therefore our hyperscaler infrastructure and technical labor requirements are a conservative estimate. Actual costs may be higher if these more advanced capabilities are built with a hyperscaler. With those limitations in mind, the following assumptions were applied to establish a consistent baseline for cost modeling across both approaches.

### Discount Rate

A yearly Discount Rate of 10% is assumed for this analysis, in order to calculate the Net Present Value of Benefits, Costs, and ROI.

## Dark Green Costs vs Light Green Benefits

### Dark Green Costs of Implementing a Hyperscaler

**Risk-adjusted PV quantified cost of building in-house using hyperscaler services include:**

Initial technical labor cost of \$587,500

Initial setup and configuration in the first year of the project requires the following employee investment to implement a hyperscaler offering:

- 2 full-time Infrastructure Engineers
- 1 full-time Model Engineer
- 1/4 of a Security Engineer
- 1/4 of a Project Manager

Technical labor cost of \$337,500 for years 1–5

Deploying an IDP solution on a Hyperscaler requires ongoing maintenance:

- 1 full-time Infrastructure Engineer
- 1/2 of a Model Engineer
- 1/4 of a Security Engineer
- 1/4 of a Project manager

Technical infrastructure cost of \$107,775 for years 1–5

Deploying an IDP solution on a Hyperscaler requires:

- AI API service: \$101,500 (1M API calls x \$.1015/document)
- Infrastructure Monitoring: \$5,075 (5% x 101,500)
- Deployment pipelines that must be installed, configured, and maintained throughout the life of the project: \$1,200 (\$99/month)

**PV total cost of building in-house IDP with Hyperscaler services after 5 years:**

**\$2,275,443**

### Dark Green Costs of Implementing Hyperscience

**Risk-adjusted present value (PV) costs include:**

Initial Implementation cost for Hyperscience Hypercell (Dark Green)

A one-time cost of \$100,000 for initial model training and workflow design.

Recurring Platform Costs for Hyperscience Hypercell (Dark Green)

An annual platform fee of \$153,639 for the *Hyperscience Hypercell* platform, based on processing one million pages per year. This cost is consistent across all five years and is accounted for once project implementation has taken place and the system is up and running.

**PV Total Cost of Hyperscience after 5 years:**

**\$682,413**

One of the challenges in calculating TCO for AI document automation is that many of the costs are difficult to quantify. *Hyperscience's* clients find it helpful when we separate the costs as dark green costs vs light green benefits.



#### Dark Green Costs

Hard costs that are represented by actual dollars spent or saved and easily measurable on spreadsheets.



#### Light Green Benefits

Business benefits that are harder to quantify but compound over time, often resulting in significant amounts through improved customer experiences, faster paths to revenue growth, higher employee satisfaction, and greater productivity.

Now that we can make an apples-to-apples cost comparison between deploying *Hyperscience* and building an in-house IDP solution with hyperscaler services, the resulting calculation reveals the total cost savings benefit of choosing *Hyperscience*. This figure represents one of the most critical inputs into the broader ROI model.

<p><b>[Present Value]</b></p> <p>Total cost of Hyperscience after 5 years:</p> <p style="font-size: 24pt; font-weight: bold;">\$682,413</p>	-	<p><b>[Present Value]</b></p> <p>Total cost of building in-house IDP with Hyperscaler services after 5 years:</p> <p style="font-size: 24pt; font-weight: bold;">\$2,275,443</p>	=	<p><b>[Present Value]</b></p> <p>Cost savings benefit of <i>Hyperscience</i> after 5 years:</p> <p style="font-size: 24pt; font-weight: bold;">\$1,593,029</p>
---	---	--	---	--

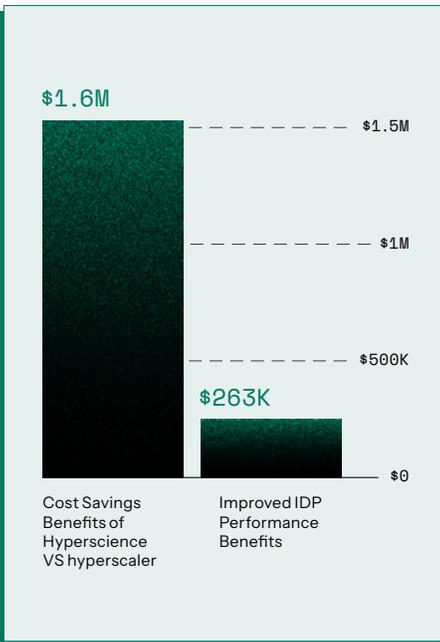
### Key Finding

By selecting *Hyperscience*, organizations can save \$1.6M over 5 years, for a medium complexity use case processing 1M documents per year. Cost avoidance is only one dimension of ROI. Equally important are the measurable performance benefits that *Hyperscience* unlocks, which translate into greater efficiency, accuracy, and scalability over time. These “light green” benefits form the second key input into the ROI model.



### Light Green Benefits

Benefits that are quantified due to improved document processing performance are slightly more speculative and depend on the exact complexity of the customer’s document estate and use case.



Improved document processing performance for years 1–5	<b>\$69,560</b>
Cost savings due to reduced model training time and effort required with the <i>Hyperscience</i> ORCA (Optical Reasoning and Cognition Agent) VLM model framework:	\$24,000
Reduced cost of manual document processing through sophisticated human-in-the-loop thresholding and accuracy harnessing:	\$20,000
Model accuracy performance improvement through iteration and fine tuning:	\$19,200
Higher accuracy levels than LLMs and other models commonly used by Hyperscalers for IDP:	\$6,000
<i>Hyperscience</i> ability to carry out least cost routing to the optimal model for each IDP task:	\$360

PV total benefit from improved document processing performance with *Hyperscience* after 5 years:

**\$263,687**

To calculate the Return on Investment (ROI) of choosing *Hyperscience* instead of building an IDP solution with hyperscaler services, we have followed the following formula:

**Costs:**

Cost of deploying *Hyperscience*

**\$682,413**

**Cost savings benefits:**

Cost difference between *Hyperscience* (\$682,413) and building with hyperscaler services (\$2,275,443)

**\$1,593,029**

**Improved document processing performance benefits:**

**\$263,687**

**Net benefits:**

Cost savings (\$1,593,029) + performance benefits (\$263,687) =

**\$1,856,716**

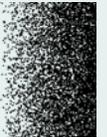
**ROI:**

Net benefits (\$1,856,716) / Cost of deploying *Hyperscience* (\$682,413) =

**272%**

**Key Finding**

The financial analysis in this TCO model found that the organization experiences Benefits Net Present Value (NPV) of \$1.9M in their IDP deployment over five years versus Costs NPV of \$682,413, with an ROI of 272%.



ROI

Benefits NPV

**272%**

**\$1.9M**

Costs NPV

Payback

**\$682K**

**<6 Months**

## The Bigger Picture of IDP Success

While we provide a data-driven approach to the main cost elements of an IDP project, there are many other qualitative factors that must be considered to ensure any IDP project is successful. This paper now shifts from cost analysis to a strategic roadmap designed to help organizations align goals, prioritize use cases, and achieve measurable results with IDP.

### Applying AI to Solve Your Biggest Document Problems

Each time an organization manually processes a physical or digital document, it slows down operations, introduces compliance risk, and directly impacts customer satisfaction. When repeated at enterprise scale, these document touchpoints evolve from minor slowdowns into systemic risks that threaten organizational performance and outcomes.

What begins as inefficiency quickly becomes a barrier to growth, innovation, and long-term competitiveness. To break through, organizations are embracing AI platforms like *Hyperscience* that reimagine how documents are processed end-to-end.

## 5 Key Questions to Answer Before Making AI Technology Decisions

Getting started with AI for document processing represents a significant business transition. Many organizations jump straight to purchasing and deploying technology. However, without first understanding the goals for using AI and creating an actionable strategy, these efforts yield lackluster results, or fail entirely. Without knowing what they want to achieve with AI, organizations will also likely make costly mistakes in their technology decisions.

Here are five key questions to answer before making any decisions:

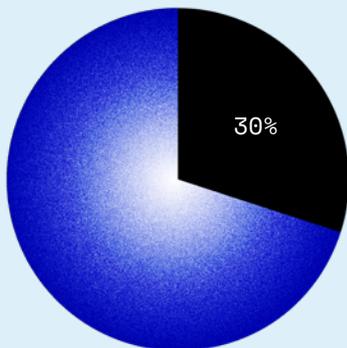
### /1 What Is the Goal?

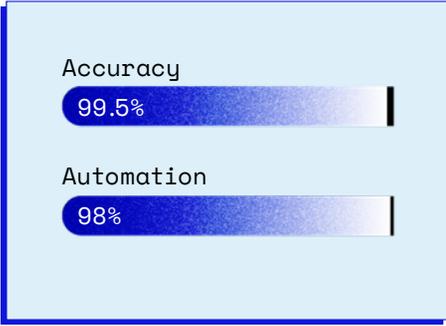
By defining a specific goal, an organization can create a strategy and select the right technology to achieve it. When determining a goal, use KPIs and define measurements of success as well as milestones. For insurance, the goal framework might look like:

- **Goal:** Reduce claims processing times from 8 weeks to 5 days
- **Measure of success:** Average human handling time per document reduced to 6 seconds
- **Milestone:** Achieve 95% automation on claims intake within the first 6 months

"Through 2026, **30%** of generative AI projects will be abandoned after proof of concept due to poor data quality, inadequate risk controls, escalating costs or unclear business value"

Gartner: A Journey Guide to Delivering AI Success Through 'AI-Ready' Data





*Hyperscience* customers routinely achieve 99.5% accuracy and 98% automation, which directly supports goals like faster claims turnaround. These targets can be set at the document type and even field level to ensure every stage of the claims process contributes to the broader objective. Other solutions require advanced development and architecture design to even reach a 50–60% accuracy level, and a lot of custom build out at significant cost to create any sort of accuracy harnessing, a standard feature in *Hyperscience* that we'll discuss further as part of our unique agentic approach to document processing.

## /2 What Documents Do I Need to Process First?

Look at the documents that are currently causing the most issues, in terms of overall volume, wasted time or the impact on customer satisfaction. Prioritize those that combine relatively straightforward processes with high business impact. Common documents to start with for different industry use cases include:



**Financial Services:** Paystubs for mortgage applications



**Healthcare:** Explanation of Benefits (EOBs) for patient billing reconciliation



**Government:** ID verification documents for benefits enrollment



**Transportation & Logistics:** Bills of lading for freight tracking and customs clearance

## /3 What Business Units Do I Start With?

Choosing the right business unit isn't just about who first touches a document—it's about who is accountable for its accuracy and outcomes. While a single document may move through multiple teams, one unit is usually the primary owner responsible for driving value from automation. Identifying that unit, while also engaging the adjacent teams who consume or rely on the document's data, ensures alignment and shared accountability. Strong starting points often include:

- **Accounts Payable** as the primary owner of invoices, with Finance and Procurement as key stakeholders.
- **Claims Operations** as the primary owner of insurance claims, with Risk and Payments as stakeholders.
- **Logistics Operations** as the primary owner of bills of lading, with Transportation Management and Finance as stakeholders.

By anchoring the initiative in a single accountable business unit and bringing adjacent teams into the conversation early, organizations can accelerate adoption and demonstrate value without overcomplicating the first deployment.

---

#### /4 Which Processes Am I Disrupting?

Documents rarely live in isolation, and neither do the processes they drive. Automating one document workflow almost always has downstream consequences: the data extracted may feed into a decisioning system, trigger a payment, update a customer record, or initiate a compliance check. What looks like a single improvement in one department can quickly ripple across others. For example:

- **Invoices:** Faster processing in Accounts Payable accelerates payment cycles but also affects cash flow management in Finance and vendor relationship management in Procurement.
- **Insurance claims:** Automating claim intake shortens resolution times, which impacts Risk Adjudication and downstream payment execution.
- **Bills of lading:** Streamlining verification speeds up Operations, but also affects Transportation Management workflows and financial reconciliation.

This is why mapping the end-to-end workflow is critical. Identifying which downstream processes intersect with the selected document ensures that automation creates efficiency and intended outcomes, rather than introducing new gaps or bottlenecks. Close collaboration with stakeholders across departments allows organizations to realize the full value of IDP while mitigating unintended risks.

---

#### /5 Who Needs to Buy into the Discussion?

Even the most promising initiatives stall without the right champions. Document automation impacts multiple teams, so building alignment early is critical. Start by identifying every leader and employee connected to the use case—from frontline processors to the executives accountable for outcomes. Engage them in a discussion not only about the inefficiencies of the current process, but also about the opportunities unlocked when automation is applied: faster turnaround times, stronger compliance, and better customer outcomes.

Equally important, address the human element directly. Successful IDP adoption depends on employees understanding that automation is not a replacement for their roles, but an enabler—freeing them from repetitive work so they can focus on higher-level decision-making, customer engagement, and innovation. When people see themselves as beneficiaries of change rather than casualties of it, they become advocates who accelerate transformation.

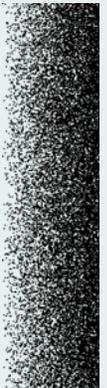
## Determining the Technical Infrastructure

The next step is selecting the technical infrastructure and tools to implement AI, which is a foundational decision with long-term impacts. Budget is, understandably, a driving force in most business decisions, and organizations often use it as the first benchmark in the decision-making process. Most companies mistakenly use the initial price of deployment to inform their purchasing decisions. Instead, use the Total Cost of Ownership (TCO) approach to compare the long-term costs of implementation, ongoing technical resources, system maintenance, employee training, infrastructure, and compliance, and make smarter business decisions that set up the organization for future success and growth.

### Key Finding

Use the Total Cost of Ownership approach to make smarter business decisions that set up the organization for future success and growth.

This is an important business decision that can affect the direction of the organization for years to come. While the choice requires upfront investment, the right technology delivers exponential returns by streamlining operations and unlocking new opportunities for growth. By taking a TCO approach, organizations can quantify the long-term ROI of their technology choices and ensure the investment translates into measurable business value.



Too often, organizations anchor their decision on short-term price-per-page calculations or license fees, overlooking the many factors that drive the true cost of Intelligent Document Processing (IDP). The reality is that TCO in IDP extends well beyond technology licensing—it includes the engineering hours to stitch tools together, the infrastructure required to scale, the cost of ongoing model training and maintenance, employee onboarding and training, and the compliance or audit frameworks necessary to keep data secure. Ignoring these inputs leads to underestimated costs, lost opportunities, reputational and financial risks, and higher spend over time.

Consider a business that sees a hyperscaler quoting document automation at \$0.10 per page and assumes it delivers the same outcomes as running *Hyperscience* on AWS, Azure, or Google Cloud Platform. What that calculation leaves out is substantial: the additional engineering resources needed to configure and maintain hyperscaler IDP tools and services, the delays in reaching production readiness, the lack of built-in accuracy guardrails, and the downstream inefficiencies created by lower automation rates. These overlooked costs compound quickly, making the comparison inaccurate and the long-term costs far higher than anticipated.

Hyperscalers can be a reasonable fit for simple, standardized scenarios—such as extracting data from a single government form (e.g., W-2s) or processing small volumes of uniform invoices. But once documents involve variability, exceptions, or downstream dependencies, hyperscaler tools quickly run into limitations in both accuracy and cost. That’s why many organizations—from Fortune 500 companies to the largest federal agencies in the United States—see the greatest value in pairing *Hyperscience* with hyperscaler infrastructure. These organizations benefit from market-leading accuracy, automation, and explainability combined with the scale and flexibility of the cloud.

Additionally, many businesses overlook time to value when selecting a solution. Companies that attempt to build their own IDP platform with hyperscaler tools often face 12–18 months of development before realizing any measurable benefits—and that’s before accounting for the ongoing burden of maintenance and model updates. By contrast, *Hyperscience* customers routinely achieve production-ready deployments in a fraction of the time, with accuracy harnessing, orchestration, and human-in-the-loop validation available out of the box. The real cost isn’t just the engineering hours—it’s the lost opportunities to reduce cycle times, serve customers faster, and free employees to focus on higher-value work.

### Each industry sets and measures AI goals differently.

Every industry uses a different category of metrics to measure the effectiveness of processing paperwork. Organizations that use the wrong success metrics typically do not achieve the desired results in the long term.

By starting with the right metrics, you can design a strategy that gets you to the right place.



For example, government agencies focusing on improving the passport application process typically focus on the number of days it takes for citizens to receive their passport. This may include tracking how many citizens receive their passport outside the 4-6 week stated standard service, as well as how many people pay extra to receive their passport in 2-3 weeks with expedited service. However, citizen satisfaction – which is more nebulous and likely requires an additional step such as a survey – must also be considered and measured.



A freight transportation and logistics company may focus on the number of days or hours it takes to process bills of lading, lumber receipts, as well as fuel and truck driving receipts in order to be able to invoice their customers faster. The process improvements from processing these individual documents faster and more accurately can greatly improve cash flow and profit margins.

## Understanding Business Factors Affecting the Total Cost of Ownership (TCO) for AI

Understanding Business Factors Affecting the Total Cost of Ownership (TCO) for AI

### Diversity of the Document Estate

The “document estate” refers to the full scope of documents an organization manages across its business processes. It’s a holistic term that covers every format—structured, semi-structured, and unstructured—and every type, from invoices and contracts to claims, applications, and government forms. Thinking in terms of the entire estate, rather than isolated document needs, is essential for building a scalable automation strategy.

When evaluating TCO, it’s not just the number of document types that matters—it’s the degree of variability within those types. An organization processing a single document type with a consistent layout will incur far lower costs than one handling dozens of variations. For example, invoices may represent a single category, but the layouts, fields, and formats can vary dramatically across suppliers. A business managing 50 different invoice formats is facing far more complexity—and higher TCO—than one working from a standardized template.

### Extraction Capability

It’s also critical to look beyond the document itself and assess the data that must be extracted to support downstream processes. The scope of extraction can vary widely: some workflows require only a handful of fields, while others demand every value on the page. The denser the data requirements—meaning more fields per document and more variability in how that data is presented—the greater the complexity and cost.

For instance, an accounts payable workflow may only need a vendor name, invoice number, and total amount to move forward, keeping extraction relatively simple. By contrast, an insurance claims workflow may require dozens of fields—policy numbers, claimant details, dates of service, line items, and more—each of which must be captured accurately to trigger adjudication, risk review, and payment downstream. The more extensive and detailed the extraction requirements, the higher the TCO and the more critical advanced extraction capabilities become.



## Transferability and Transparency

Review the processes and consider where the data is going. Organizations commonly take a piece of data from one system and transfer it through data integrations to another. By automating the process and ensuring transparency and transferability, organizations can materially improve their TCO. While building in-house might seem to offer more control, relying on third-party models without transparency and explainability can paradoxically lead to less control and make fixing errors at scale difficult.

## Operational Effectiveness

Additionally, the benefits and cost of improving effectiveness are not consistent as you move from zero to 100. In other words, going from nothing to something is very different than going from something to near perfection. Early gains—moving from 0% to 50% effectiveness—are usually achieved by eliminating obvious inefficiencies, automating repetitive tasks, and reducing manual handling. These “low-hanging fruit” improvements deliver quick wins at relatively low cost.

However, progress becomes progressively harder and more expensive as organizations aim to achieve end-to-end digital transformation that leads to business impact, competitive advantage, and industry disruption. Reaching 70–80% effectiveness requires tackling more complex workflows, integrating multiple systems, and dealing with exceptions. Breaking through to 90% or higher almost always demands advanced capabilities—such as accuracy harnessing, sophisticated human-in-the-loop controls, and model lifecycle management—that are costly and resource-intensive to build from scratch.

Many IDP projects assembled from hyperscaler components stall here and fail to deliver on the promises of digital transformation, because the final stretch requires exponentially greater investment and precision to achieve exceptional gains.

Although it is challenging to calculate or estimate the cost to achieve the required level of operational effectiveness, organizations should review their operations output through this lens to determine improvements.

Common examples to measure operational effectiveness include:

- **Error remediation time:** how quickly downstream teams can identify and correct incorrectly extracted data.
- **Rework or exception rate:** percentage of documents requiring manual reprocessing.
- **End-to-end cycle time:** from document receipt to downstream action, such as claim payout or payment posting.
- **Throughput capacity:** documents processed per hour / day with existing resources.

## Unique Agentic Approach to Extraction

Business documents only create value when they can be accurately read, understood, and acted upon. Yet each document type—whether a contract, medical record, or bill of lading—presents its own challenges: rotation, skew, handwriting, multiple languages, or massive attachments. Many organizations underestimate this complexity and rely on hyperscaler OCR and layout models, which quickly break down on unstructured or highly variable content.

With hyperscalers, oversight and accuracy safeguards must be cobbled together from separate APIs—HITL review steps, verification logic, and model supervision stitched into custom workflows. What looks like a simple add-on often turns into a major engineering project, driving up cost and complexity without delivering the accuracy enterprises need.

*Hyperscience* takes a different path. Our agentic extraction approach can be thought of as an ensemble of specialized models working together under the guidance of a goal-oriented agent. Instead of asking a single model to do everything, the agent orchestrates multiple proprietary models, vision language models (VLMs), and even LLMs across preprocessing, classification, identification, and extraction. Each model is chosen, sequenced, or swapped based on the desired business outcome—accuracy, speed, cost, or compliance thresholds set by the customer.

This agentic approach is what transforms document processing from static automation into a dynamic system. The agent doesn't just run models; it understands the objective. For example, if the goal is to maximize compliance, the agent may prioritize accuracy harnessing and route uncertain fields to HITL. If the goal is faster turnaround with lower cost, the agent can lean on zero-shot VLMs to handle more fields automatically. Not all documents—or fields within documents—are equal, and our approach allows fine-grained control to reflect that reality.

Because the agent is orchestrating across a modular, composable platform of blocks and flows, customers gain pipelines that are optimized end-to-end: combining traditional AI, proprietary models, and generative AI as needed. The result is higher accuracy, lower compute cost, and minimal human intervention—achieved not through patchwork, but through a flexible, intelligent system that continuously adapts to business goals.

The agent doesn't just run models; it understands the objective.

## For most organizations, the optimal path is not deciding between “Hyperscience versus hyperscalers,” but Hyperscience running on hyperscaler infrastructure.

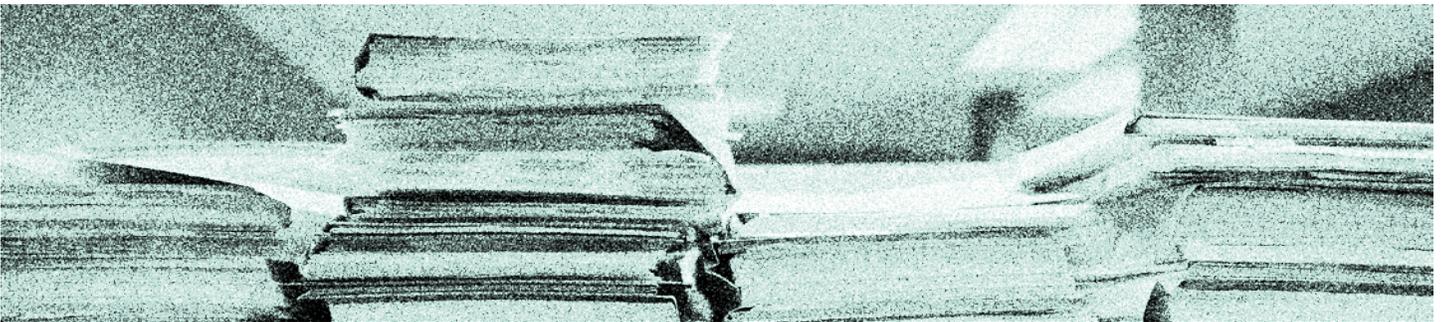
Many enterprises have already standardized on AWS, Microsoft Azure, or Google Cloud Platform for their core compute and storage. *Hyperscience* meets them where they are, natively deploying Hypercell within their existing cloud environment. This allows organizations to extend the value of their current investment while reducing the incremental compute, storage, and engineering resources that DIY builds demand.

With *Hyperscience*, setup is straightforward: our team collaborates with stakeholders to define goals, then deploys the platform directly into the customer’s hyperscaler environment—whether AWS, Azure, or GCP. Engineering resources are bundled into the service, eliminating hidden fees or long lead times. Organizations are able to begin automating document processes and delivering measurable business value in days, not the 6–12 months common with DIY approaches.

The result is a “best of both worlds” solution:

- Hyperscaler scale and flexibility combined with *Hyperscience*’s accuracy, automation, and compliance controls.
- Higher precision and resilience through specialized, customer-trained small language models from *Hyperscience*, rather than generic off-the-shelf OCR or form recognizers.
- Continuous improvement through model drift management—something hyperscalers do not natively offer. For example, if a vendor changes the format of purchase orders, *Hyperscience* can quickly retrain models to restore accuracy, while hyperscaler-only solutions stall or require heavy custom development.
- Transparency, explainability, and trust via human-in-the-loop review, audit reporting, and built-in accuracy harnessing.

In practice, this means enterprises gain faster time to value, lower TCO, and superior accuracy compared to hyperscaler-only solutions, without abandoning their existing cloud infrastructure.

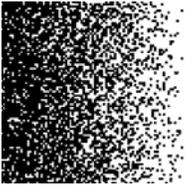


## Conclusion

\\ Your decision about which model to use for AI-based document processing makes a significant impact on your business today and years to come. Businesses that look beyond the initial costs to see the savings found by purchasing *Hyperscience* better position their organization for success.

*Hyperscience* provides significant benefits in terms of cost savings as well as the ability to capitalize on opportunities today instead of waiting a year. By choosing *Hyperscience* to manage hyperscaler capabilities and add significant value above that, you can realize the greatest business benefits and TCO value.





# Calculating the TCO of a Hyperscience Deployment for 1 million docs per annum

## Summary Table

@ 1M pages / year	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
<b>Cost of implementing Hyperscience</b>	<b>\$(100,000)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(868,196)</b>	<b>\$(682,413)</b>
Initial model training and workflow design	\$(100,000)						\$(100,000)	\$(100,000)
Hyperscience infrastructure and orchestration service costs		\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(768,196)	\$(582,413)
<b>Cost of implementing hyperscaler</b>	<b>\$587,500</b>	<b>\$445,275</b>	<b>\$445,275</b>	<b>\$445,275</b>	<b>\$445,275</b>	<b>\$445,275</b>	<b>\$2,813,875</b>	<b>\$2,275,443</b>
<b>Technical labor costs</b>	<b>\$587,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$2,275,000</b>	<b>\$1,866,891</b>
Infrastructure Engineering	\$300,000	\$150,000	\$150,000	\$150,000	\$150,000	\$150,000	\$1,050,000	\$868,618
Model Engineering	\$200,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$700,000	\$579,079
Security Engineering	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$225,000	\$179,655
Project / Program Management	\$50,000	\$50,000	\$50,000	\$50,000	\$50,000	\$50,000	\$300,000	\$239,539
<b>Technical infrastructure costs</b>		<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$538,875</b>	<b>\$408,552</b>
AI API service costs		\$101,500	\$101,500	\$101,500	\$101,500	\$101,500	\$507,500	\$384,765
Infrastructure monitoring		\$5,075	\$5,075	\$5,075	\$5,075	\$5,075	\$25,375	\$19,238
Deployment pipelines		\$1,200	\$1,200	\$1,200	\$1,200	\$1,200	\$6,000	\$4,549
<b>Cost savings benefits of Hyperscience versus hyperscaler</b>	<b>\$487,500</b>	<b>\$291,636</b>	<b>\$291,636</b>	<b>\$291,636</b>	<b>\$291,636</b>	<b>\$291,636</b>	<b>\$1,945,679</b>	<b>\$1,593,029</b>
<b>Improved IDP performance benefits of using Hyperscience</b>		<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$347,800</b>	<b>\$263,687</b>
Reduced training and cost with ORCA		\$24,000	\$24,000	\$24,000	\$24,000	\$24,000	\$120,000	\$90,979
Reduced cost of manual document processing		\$20,000	\$20,000	\$20,000	\$20,000	\$20,000	\$100,000	\$75,816
Model performance improvement		\$19,200	\$19,200	\$19,200	\$19,200	\$19,200	\$96,000	\$72,783
Increased accuracy		\$6,000	\$6,000	\$6,000	\$6,000	\$6,000	\$30,000	\$22,745
Least cost routing for model execution		\$360	\$360	\$360	\$360	\$360	\$1,800	\$1,365
<b>Net benefits</b>	<b>\$487,500</b>	<b>\$361,196</b>	<b>\$361,196</b>	<b>\$361,196</b>	<b>\$361,196</b>	<b>\$361,196</b>	<b>\$2,293,479</b>	<b>\$1,856,716</b>
<b>ROI</b>								<b>272%</b>
<b>Payback</b>								<b>&lt;6 months</b>

## Analysis of Costs - Implementing [Hyperscience]

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Initial model training and workflow design	\$(100,000)						\$(100,000)	\$(100,000)
Hypercell infrastructure and orchestration service costs		\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(768,196)	\$(582,413)
<b>Total costs (risk-adjusted)</b>	<b>\$(100,000)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(153,639)</b>	<b>\$(868,196)</b>	<b>\$(682,413)</b>

This table summarizes the costs of running Hyperscience Hypercell, both from initial setup costs to recurring costs throughout year 1 – year 5, and all discounted at a rate of 10%, resulting in a Net Present Value of \$682,413 over five years. The two lines of costs are detailed below:

### Initial model training and workflow design:

The implementation of Hyperscience includes an average initial cost of \$100,000 for two key sets of services: **XDS (Expert Data Services)** and **XFS (Expert Flow Services)**.

**XDS (Expert Data Services)** is a professional services offering from Hyperscience where our team builds layouts and trains machine learning (ML) models on behalf of the customer, using the customer’s specific data and documents.

The primary goal of XDS is to accelerate a customer’s time-to-value by leveraging Hyperscience’s specialized expertise. It lowers the initial time and effort required from the customer to train models and prepare the system for production, which is particularly beneficial when replacing legacy systems. The service helps customers build expertise on their unique datasets by having Hyperscience experts lead the initial model development.

XDS includes the following services:

- **Layout and Model Creation:** The creation of layouts and the training of corresponding models for specific document types, particularly for semi-structured documents. This includes training field identification models and, if applicable, table identification models.
- **Transcription Threshold Optimization:** A key part of the service is optimizing transcription thresholds to enhance accuracy and efficiency for various data fields within the customer’s documents.

Key deliverables from an XDS engagement typically include:

- An **Implementation Solution Document (ISD)** summarizing the layout requirements and design.
- A demonstration of the developed layouts and models.
- The final, trained layouts and models provided to the customer for deployment.

**XFS (Expert Flow Services)** is a professional service offering where Hyperscience experts design and build custom workflows, also known as “flows,” on behalf of a customer.

The service provides customers with expertly built, custom workflows designed around their unique business processes. It serves as an alternative to customers building flows themselves using the no-code interface in the Flow Studio or programmatically with the Flows SDK. This is particularly useful for implementing custom business requirements like data validation or redaction.

- **Process:** The process begins with a discovery phase, typically led by Hyperscience’s sales engineering or implementation consulting teams, to thoroughly understand the customer’s business problem and requirements. Hyperscience then brainstorms a solution and designs a custom flow using the platform’s available “blocks” to meet those needs. The development follows a structured approach of design, build, and deploy.
- **Scope:** XFS focuses on creating the entire end-to-end business process workflow. The service can be used to create a variety of flows, such as those for case collation or combining full-page transcription with standard document processing.

**Hypercell infrastructure and orchestration license costs:**

The Hyperscience Hypercell platform has an annual list price cost of \$153,639 for processing 1 Million pages / year. This platform and orchestration license cost includes the following key platform elements:

**ML Models** - Composable pre-trained models (OICR - Optical Intelligent Character Recognition), Vision Language models (ORCA - Optical Reasoning and Cognition Agent), Specialized models, Task-based ML.

**Blocks** - Modular and reusable components within a document processing workflow. Options include Standard Blocks, Advanced Blocks (include External API calls, Knowledge Store and Custom Supervision), Gen AI blocks (leverage LLMs and Vector Stores for natural language querying and RAG. Knowledge Store (for storing common artifacts for document processing and accuracy harnessing of models).

**Page volume** - Platform is priced at different tiers to reach the annual page volume required.

**Infrastructure** - Deploy on-premises, SaaS (Hyperscience manages all infrastructure deployment of Hypercell), or FedRAMP High (for critical government use cases and workflows).

*This actual cost could vary depending on the specific Hypercell components chosen as well as the use case details addressed for each organization.*

**Analysis of Costs - Implementing [a hyperscaler]**

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Technical labor costs	\$587,500	\$337,500	\$337,500	\$337,500	\$337,500	\$337,500	\$2,275,000	\$1,866,891
Technical infrastructure costs		\$107,775	\$107,775	\$107,775	\$107,775	\$107,775	\$538,875	\$408,552
Total costs (risk-adjusted)	\$587,500	\$445,275	\$445,275	\$445,275	\$445,275	\$445,275	\$2,813,875	\$2,275,443

This table summarizes the cost of implementing an IDP solution with a hyperscaler. This is due to the **technical labor costs** required for implementing a hyperscaler, as well as the **technical infrastructure costs**. When the initial costs and recurring costs are totaled over five years and discounted at 10%, the Net Present Value of these costs are \$2,275,443.

The next two tables detail the calculation of each of these costs in detail.

### Technical labor costs - implementing a hyperscaler

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Infrastructure Engineering	\$300,000	\$150,000	\$150,000	\$150,000	\$150,000	\$150,000	\$1,050,000	\$868,618
Model Engineering	\$200,000	\$100,000	\$100,000	\$100,000	\$100,000	\$100,000	\$700,000	\$579,079
Security Engineering	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$225,000	\$179,655
Project / Program Management	\$50,000	\$50,000	\$50,000	\$50,000	\$50,000	\$50,000	\$300,000	\$239,539
<b>Technical labor costs (risk-adjusted)</b>	<b>\$587,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$337,500</b>	<b>\$2,275,000</b>	<b>\$1,866,891</b>

Based on a detailed analysis of hyperscaler model architectures, the following costs have been identified. The technical labor roles required for deploying an IDP solution on a hyperscaler include, but are not limited to, the following:

- 1/ **Infrastructure Engineering:** estimated to be 2 FTE infrastructure engineers in the initial one-year design, build, and deployment phase. These FTEs have an annual salary cost of \$150,000. After the initial setup and project go live, the requirement for infrastructure engineering drops to 1 FTE / year, at the same cost. The total cost is \$1,050,000, discounted at 10% over 5 years, which yields a Net Present Value of \$868,618.
- 2/ **Model Engineering:** estimated to be 1 FTE machine learning model engineer in the initial one-year design, build, and deployment phase. This FTE has an annual salary cost of \$200,000. After the initial setup and project go live, the requirement for model engineering drops to 0.5 FTE / year, at the same cost. The total cost is \$700,000, discounted at 10% over 5 years, which yields a Net Present Value of \$579,079.
- 3/ **Security Engineering:** estimated to be 0.25 FTE security engineer in the initial one-year design, build, and deployment phase. This FTE has an annual salary cost of \$150,000. After the initial setup and project go live, the requirement for security engineering remains constant at 0.25 FTE / year, at the same cost. The total cost is \$225,000, discounted at 10% over 5 years, which yields a Net Present Value of \$179,655.
- 4/ **Project / Program Management:** estimated to be 0.5 FTE program manager in the initial one-year design, build, and deployment phase. This FTE has an annual salary cost of \$200,000. After the initial setup and project go live, the requirement for project / program management remains constant at 0.5 FTE / year, at the same cost. The total cost is \$300,000, discounted at 10% over 5 years, which yields a Net Present Value of \$239,539.

### Technical infrastructure costs - implementing a hyperscaler

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
AI API service costs		\$101,500	\$101,500	\$101,500	\$101,500	\$101,500	\$507,500	\$384,765
Infrastructure monitoring		\$5,075	\$5,075	\$5,075	\$5,075	\$5,075	\$25,375	\$19,238
Deployment pipelines		\$1,200	\$1,200	\$1,200	\$1,200	\$1,200	\$6,000	\$4,549
<b>Technical labor costs (risk-adjusted)</b>		<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$107,775</b>	<b>\$538,875</b>	<b>\$408,552</b>

Based on an analysis of hyperscaler reference IDP architectures, the following costs have been identified:

**1 / API Service costs** - estimated at \$101,500 per year, based on cost from the hyperscaler of \$0.1015 / page, times 1 Million pages processed per year. The total cost is \$507,500 discounted at 10% over 5 years, which yields a Net Present Value of \$384,765.

This cost of API Services is a composite calculation across the different hyperscaler services invoked during Document Ingestion and Text Extraction, Document Classification, Extraction, and Results Validation, while also taking into account an average of three prompts submitted during the extraction phase, for a medium complexity use case.

**2 / Infrastructure monitoring** - estimated at 5% of API service costs. The total cost is \$25,375 discounted at 10% over 5 years, which yields a Net Present Value of \$19,238.

**3 / Deployment pipelines** - estimated at \$99 / month (\$1,200 / year). The total cost is \$6,000 discounted at 10% over 5 years, which yields a Net Present Value of \$4,549.

### Cost savings benefits of Hyperscience versus hyperscaler

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Cost of implementing Hyperscience	\$(100,000)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(1,868,196)	\$(682,413)
Cost of implementing hyperscaler	\$587,500	\$445,275	\$445,275	\$445,275	\$445,275	\$445,275	\$2,813,875	\$2,275,443
Cost savings benefits of Hyperscience versus hyperscaler	\$487,500	\$291,636	\$291,636	\$291,636	\$291,636	\$291,636	\$1,945,679	\$1,593,029

The total difference in cost between implementing Hyperscience and a hyperscaler is \$1,945,679 discounted at 10% over 5 years, which represents a Net Present Value of \$1,593,029.

### Improved document processing performance - implementing Hyperscience instead of a hyperscaler

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Reduced training and cost with ORCA		\$24,000	\$24,000	\$24,000	\$24,000	\$24,000	\$120,000	\$90,979
Reduced cost of manual document processing		\$20,000	\$20,000	\$20,000	\$20,000	\$20,000	\$100,000	\$75,816
Model performance improvement		\$19,200	\$19,200	\$19,200	\$19,200	\$19,200	\$96,000	\$72,783
Increased accuracy		\$6,000	\$6,000	\$6,000	\$6,000	\$6,000	\$30,000	\$22,745
Least cost routing for model execution		\$360	\$360	\$360	\$360	\$360	\$1,800	\$1,365
<b>Total performance benefits (risk-adjusted)</b>		<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$69,560</b>	<b>\$347,800</b>	<b>\$263,687</b>

**Reduced training and cost with ORCA:** ORCA is a VLM or Vision Language Model that excels at multi-modal document extraction (images and text). As such, it provides a

zero-shot experience. That is, it does not require training with sample documents in order to be up and running. It is a high performance model which runs on GPUs, ideal for high document variability and lower document volume processing needs. The cost savings here represent reduced time and effort training ML models for a subset of the 1 million pages per annum.

The total benefit of this reduced training time and effort is \$24,000 per year discounted at 10% over 5 years, which yields a Net Present Value of \$90,979.

**Reduced cost of manual document processing:** Hyperscience offers an optimal AHT (average handling time) for manual intervention from Human-in-the-loop supervisors. Our unique accuracy thresholding, QA for accuracy measurement and Custom Supervision for advanced knowledge worker analysis, review and sign off ensure much higher efficiency than standard business process outsourcing / manual keying and re-work required by hyperscaler IDP solutions. The total benefit of this reduced manual document processing cost is \$20,000 per year discounted at 10% over 5 years, which yields a Net Present Value of \$75,816.

**Model performance improvement:** Hyperscience models are continually tuned and refined as they process additional documents from your business, driving higher accuracy and corresponding cost savings. The total benefit of model performance improvement is \$19,200 per year discounted at 10% over 5 years, which yields a Net Present Value of \$72,783.

**Increased accuracy:** our extensive benchmarking of our ML models compared to the main LLMs and commercially available model shows on average a 10%+ higher accuracy level versus hyperscaler IDP solutions. The total benefit of increased document processing accuracy is \$6,000 per year discounted at 10% over 5 years, which yields a Net Present Value of \$22,745.

**Least cost routing for model execution:** The Hyperscience Hypercell's unique agentic approach and modular and composable architecture allows companies to leverage the models needed for each use case, taking into account cost of execution on CPU or GPU, accuracy level, throughput, and other key factors, in order to optimize cost over time. For example, customers may start their document processing with a higher cost VLM (ORCA) running on GPUs, and then switch to a smaller specialized model on CPUs once the document processing workflow has been built, tested, fine tuned and is running at scale. The total benefit of least cost routing is \$360 per year discounted at 10% over 5 years, which yields a Net Present Value of \$1,365.

## Financial Summary

Cash Flow Analysis (Risk-Adjusted Estimates): ROI for using Hyperscience instead of hyperscalers.

Metric	Initial	Year 1	Year 2	Year 3	Year 4	Year 5	Total	Present Value
Total costs	\$(100,000)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(153,639)	\$(768,196)	\$(682,413)
Total benefits	\$587,500	\$514,835	\$514,835	\$514,835	\$514,835	\$514,835	\$3,161,675	\$2,539,130
Net benefits	\$487,500	\$361,196	\$361,196	\$361,196	\$361,196	\$361,196	\$2,293,479	\$1,856,716
ROI								272%
Payback (months)								<6

So to summarize the outcomes of this analysis, the total cost of setting up and running Hyperscience for five years, to process 1 Million pages per annum, is \$768,196, which discounted at 10% over 5 years yields a Net Present Value of \$682,413.

The Total benefits experienced by running Hyperscience and avoiding the higher technical labor costs and technical infrastructure costs as well as harvesting the higher performance levels of accuracy and reduced model training costs over five years is \$3,161,675, which discounted over 5 years yields a Net Present Value of \$2,539,130 and a Net Benefit (Benefits - Costs) of \$1,856,716.

This yields an ROI of 272% with a payback period of less than 6 months.

