Hyperscience

Hyperscience recently launched its Hypercell R41 platform, marking a pivotal evolution in the company's strategy. R41's advanced AI capabilities address persistent challenges in intelligent document processing (IDP) and can orchestrate the entire document lifecycle from ingestion and classification to extraction, validation, and decisioning. Organizations with high-volume, complex document workflows will want to take a look.

Founded 2014 | HQ New York, NY | 250 employees July 2025

📋 The Company

Hyperscience is a leading intelligent document processing (IDP) software company founded in 2014 and headquartered in New York City. Known for its laser focus on and innovative approach to document processing, the company has carved out a strong position in high-volume, mission-critical document automation, particularly within government and financial services.

The recent launch of its Hypercell R41 platform marks a pivotal evolution in Hyperscience's strategy: introducing advanced AI capabilities that address the persistent challenges of onboarding new document types, orchestrating AI models, and enabling human-in-the-loop (HITL) decisioning. The company's $80 million contract with the US Social Security Administration, recognized as the "IDP Deal of the Year" in 2024 by Deep Analysis, underscores its momentum and credibility. With its recent US FedRAMP High certification, Hyperscience is well positioned for further expansion in the public sector.

⚙️ The Technology

The Hypercell platform is designed to orchestrate the entire document lifecycle from ingestion and classification to extraction, validation, and decisioning. The R41 release introduced innovations centered around three themes: understanding, speed, and modularity. The following summarizes those innovations.

ORCA: Vision Language Intelligence

ORCA (Optical Reasoning Cognition Agent) is a proprietary vision language model (VLM) purpose-built by Hyperscience to handle issues presented by unstructured and novel document types. It addresses one of the most persistent challenges in IDP: the "cold-start" problem, which occurs when there is insufficient data or prior knowledge for the AI model to make accurate data extraction decisions. This is particularly problematic when onboarding new document types or formats into an existing workflow; the entire workflow must stop until the ML model is trained. ORCA uses zero-shot learning to extract data from new documents without requiring extensive training data or ML expertise; this will reduce – and in some cases, eliminate – costly delays.

Built around a 7-billion-parameter open-source VLM, ORCA is designed for continuous benchmarking and improvement, allowing Hyperscience to rapidly incorporate new features from the open-source AI ecosystem. To provide transparency and compliance, ORCA operates within Hyperscience's supervised environment, inheriting the HITL workflows, accuracy scoring, and audit trails. ORCA is optimized for multi-field extraction tasks that require contextual understanding – such as identifying relationships between fields (e.g., borrower and co-borrower) or interpreting conditional clauses in contracts.

ORCA also includes few-shot fine-tuning. When customization is needed, ORCA can be fine-tuned with as few as 15 annotated documents using the no-code UI. This enables business analysts to adapt the AI model without data scientist or developer support.

VLMs like ORCA represent an evolutionary leap forward for IDP, which can now manage the most complex use cases containing diverse documents and forms (structured, semi-structured, and unstructured). For example:

→ Mortgage origination document intake (tax returns, banking statements, inspection reports, closing disclosures, building surveys, flood zone analysis, etc.)

→ Insurance claims processing (accident reports, accident scene photos, repair estimates, policy documents, medical bills, etc.)

→ Government benefits eligibility determination (personal identification, income verification, residency proof, etc.)

→ Healthcare prior authorization or medical necessity reviews (clinical data, administrative data, personal information forms, physician notes, lab reports, etc.)

OICR: Full-page OCR

While ORCA addresses the challenges of unstructured and novel document types, Hyperscience's proprietary OICR OCR AI model remains a foundational component for high-precision text extraction from structured and semi-structured documents. OICR (Optical Intelligent Character Recognition) is a deterministic AI model that provides consistent, rule-based extraction, essential for regulated industries that require predictable and auditable outcomes.

Unlike GPU-intensive LLMs and VLMs, OICR was engineered for high performance and low latency on a standard CPU infrastructure, which best suits on-premises deployments or environments with strict hardware constraints. The model is optimized for documents with fixed layouts (such as tax forms, invoices, and application templates) where field positions and labels are known or somewhat predictable.

In practice, Hyperscience workflows often combine both models: OICR handles the structured components of a document packet (e.g., a standardized form), while ORCA interprets accompanying unstructured content (e.g., handwritten notes or attachments).

Multimodal AI Orchestration

Hyperscience's architectural philosophy is that no single AI model or tool can address the full spectrum of document processing and decisioning challenges. The platform is designed as an open, composable orchestration layer, allowing organizations to dynamically combine proprietary models, third-party AI services, and enterprise data infrastructure. Hyperscience positions this modular approach as "The Kubernetes of IDP." (Kubernetes is an open-source platform for automating the deployment, scaling, and management of containerized applications. It acts as a container orchestration system, with tools to manage applications across a cluster of machines.)

This orchestration is enabled by the platform's modular "Blocks and Flows" architecture, which abstracts complex AI capabilities and skills into reusable components that can be assembled by business analysts into document automation workflows (see Figure 1).

## Core Orchestration Components

In addition to managing its own ORCA and OICR AI models, Hyperscience orchestrates the integration and management of third-party LLMs from providers such as OpenAI (GPT), Anthropic (Claude), DeepSeek, and Google (Gemini). Hyperscience provides pre-built blocks for these models, allowing customers to invoke them selectively within a workflow; for example, using GPT to summarize a document packet after ORCA has extracted the structured fields.

R41 also orchestrates integration with vector databases (such as Pinecone or Weaviate) to power semantic search across document repositories. This capability supports retrieval-augmented generation (RAG) workflows, where LLMs can ground their responses in enterprise-specific knowledge. Orchestrated workflows can also include checkpoints for human validation, exception handling, and audit logging.

Document Chat: Decisioning

R41 introduces Document Chat, a natural language interface embedded in its custom supervision UI. Powered by GenAI, this feature enables knowledge workers to query a packet (folder or file) of documents using plain language (for example, in a mortgage loan packet, ask "What is the condition of the roof?") and receive summarized answers with inline citations that highlight the source text on the

Figure 1

Blocks and Flows Architecture of Hypercell R41
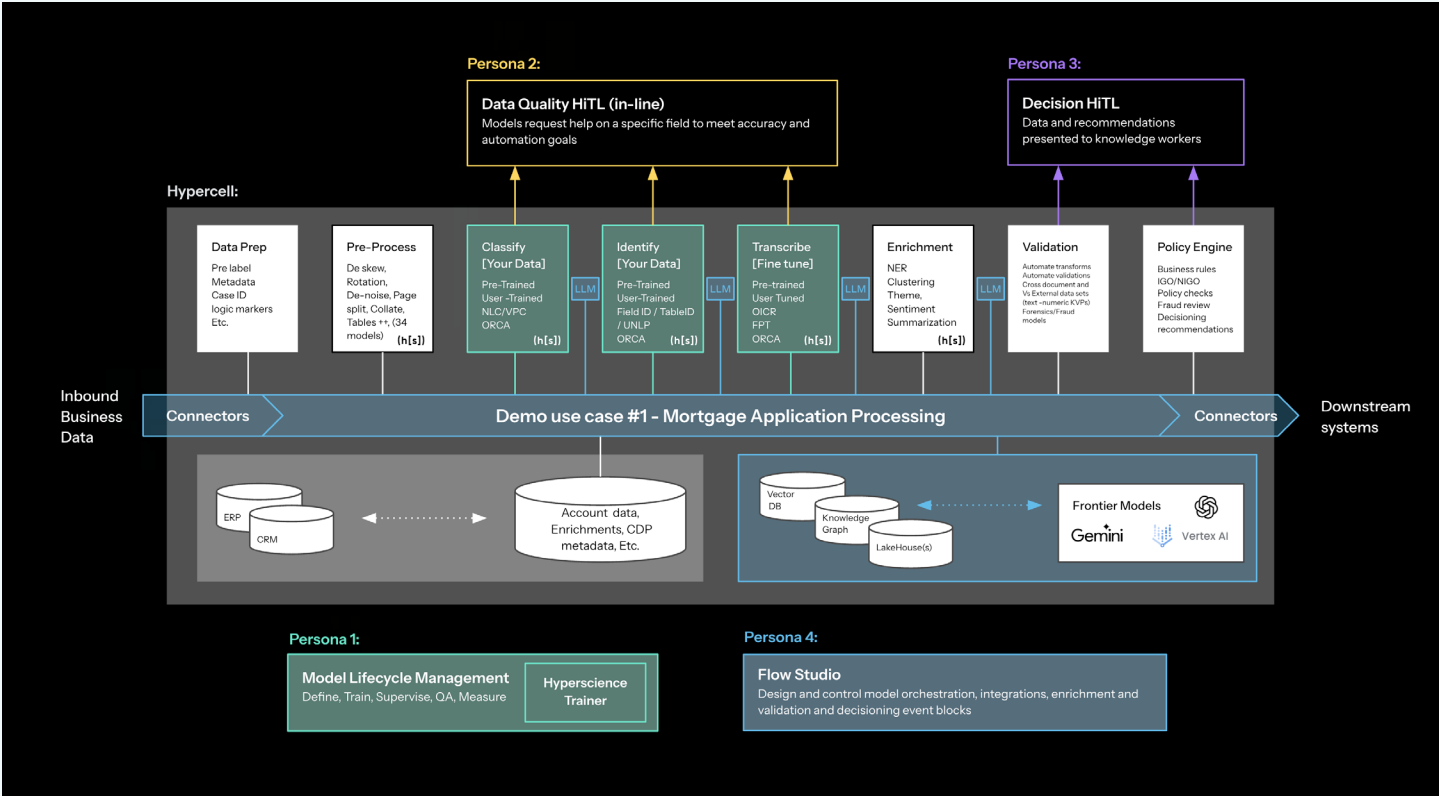
original document. This could reduce the need for a knowledge worker to constantly switch between applications and also accelerate decision-making by moving decisioning forward into the document automation environment. Other IDP vendors have offered this functionality for a while and Hyperscience has responded to the competition.

The Roadmap

The company's product direction seems clear: become an enterprise orchestration layer for document AI with enhanced agentic capabilities. The orchestration-first approach is foundational to Hyperscience's vision for agentic AI; the platform acts as a cognitive agent that can break down complex goals into discrete tasks, select the most appropriate model or tool for each task, monitor execution and escalate exceptions to human reviewers, and continuously improve the performance by learning from feedback. With this vision, Hyperscience can position itself beyond IDP and become a strategic AI infrastructure layer for deploying document intelligence at scale.

💬 Our Opinion

Hyperscience's R41 release is a timely response to the evolving AI landscape. The "Kubernetes of IDP" positioning is more than a clever metaphor; it is a strategic blueprint for modular, AI-driven document automation that will appeal to CIOs and CTOs. ORCA addresses a competitive market need for robust cold-start capabilities, while the document chat offering joins similar features from other vendors in extending the traditional scope of IDP to include real-time decision support.

However, as Hyperscience moves into the knowledge worker decisioning realm with features like document chat, it should connect with the domain-specific systems of record where these workers spend most of their time (e.g., loan origination systems such as Encompass, or healthcare management systems such as Epic and Cerner). Success will hinge on a clear go-to-market strategy that either partners with these companies or persuades enterprises to shift some knowledge worker decisioning tasks forward into the Hyperscience platform.

💡 Advice to Buyers

Organizations with high-volume, complex document workflows – whether modernizing legacy systems like Kofax, OpenText Captiva, or IBM Datacap, or launching new initiatives – should shortlist Hyperscience. With key wins such as the SSA and the VA, the company has proven its ability to handle the most demanding use cases at scale. Buyers should evaluate ORCA using their own "cold-start" documents to accurately assess the zero-shot performance. And with R41, Hyperscience has also introduced a new product offering (Developer Edition) that makes it easier to run a proof of concept (POC) before deployment.

🔍 SOAR Analysis

Strengths

→ Proven end-to-end processing platform for high-volume use cases

→ Modular "Blocks and Flows" architecture for AI orchestration

→ ORCA VLM delivers competitive zero- and few-shot capabilities

Aspirations

→ Become the "Kubernetes of IDP"

→ Lead in agentic AI automation for documents

→ Evolve into a decisioning platform

Opportunities

→ Partner with system-of-record vendors to embed decisioning tools

→ Use modular pricing to drive land-and-expand growth

→ Expand further into LLM use cases

Results

→ Secured $80 million US Social Security Administration contract

→ Raised $289+ million in funding

→ Achieved FedRAMP High certification

About Deep Analysis

We Research Innovation

Unstructured Data Management

ECM, KM, EX, etc.

Process & Task Automation

Intelligent Document Processing (IDP)

Due Diligence

Competitive Intelligence

Thought Leadership

Market & Product Strategy

Our Services

Deep Analysis is an advisory firm that helps technology vendors, buyers, and investors understand and address the challenges of innovative and disruptive technologies in the enterprise software marketplace.

The firm's work is built on decades of experience advising and consulting to global technology firms large and small, from SAP, Oracle, and HP to countless start-ups.

Led by Alan Pelz-Sharpe, Deep Analysis works with technology vendors, buyers of enterprise technology, and investors in the ECM and enterprise automation market to improve their understanding of the information management technology landscape and provide actionable guidance.

Deep Analysis' timely book, "Practical Artificial Intelligence: An Enterprise Playbook," outlines strategies for organizations to avoid pitfalls and successfully deploy AI.

Contact us:

info@deep-analysis.net

+1 978 877 7915