Hyperscience

Hyperscience has successfully transitioned from a cool VC-funded AI start-up into a mature IDP company. Its latest release offers end-to-end IDP for every document class: structured, semi-structured, and unstructured. On its current trajectory, we think Hyperscience can challenge Kofax and ABBYY for market leadership as measured by revenue.

Founded 2014 | HQ New York City | >250 employees April 2023

The Company

Hyperscience, a pure-play intelligent document processing (IDP) software company, was founded in 2014 in Sofia, Bulgaria, and is now headquartered in New York City. The company employs over 250 people. It has raised nearly $300 million to date – by far the largest investment ever in an IDP company – with the latest round a Series E placement in 2021 of $100 million from Tiger Global and Bessemer Ventures, among others.

In a March 2023 briefing, Hyperscience reported strong growth year over year including a 100% increase in volume of pages processed and an 80% spike in annual recurring revenue for new unstructured data use cases. The team showed us an impressive go-to-market slide with their broad array of channel and technology partners worldwide.
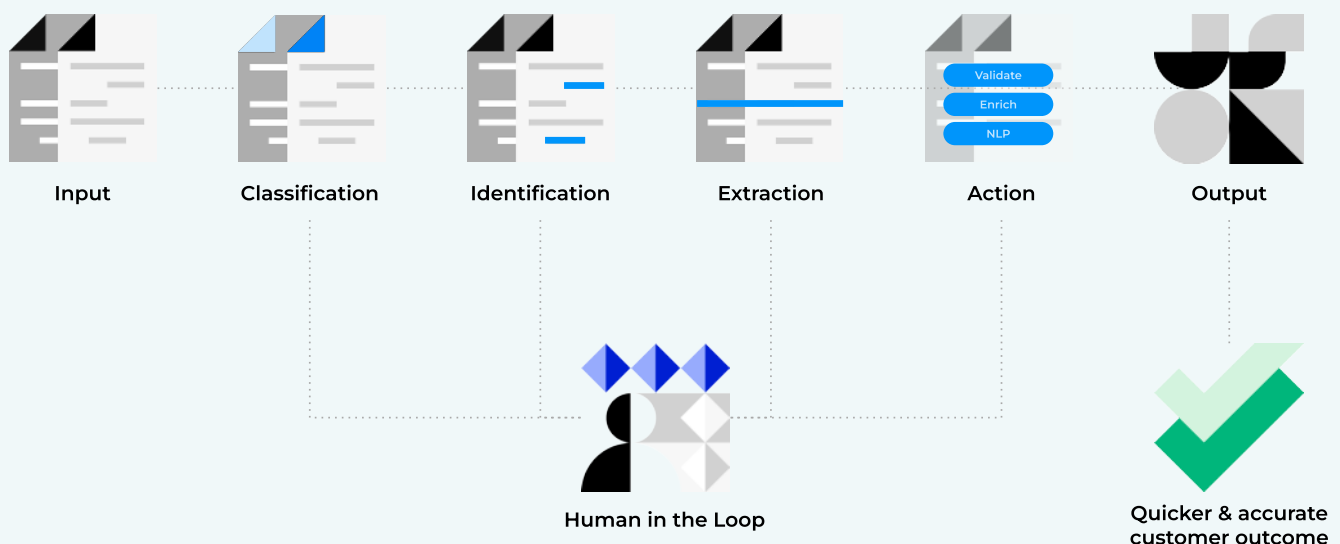
The Technology

Hyperscience is best known for developing a modern, AI-first approach to optical character recognition (OCR). The company developed its own OCR engine from the ground up, eschewing both the traditional engines developed in the 1980s and 1990s – used in Kofax, ABBYY, and other classic IDP vendors – and the stripped-down Tesseract open-source engine from Google that many other start-ups embraced. At the time, choosing to build versus buy was a bold and expensive decision. Today we think it gives Hyperscience a distinct competitive advantage.

However, Hyperscience does far more than OCR. The company has built a full-featured, end-to-end IDP platform using AI and ML, ranging from document ingestion through classification and quality control all the way to integrations with leading business applications for RPA, ERP, CRM, content management, and other processes.

Figure 1
The Hyperscience IDP Process Flow

Input   Classification   Identification   Extraction   Action   Output

Validate
Enrich
NLP

Human in the Loop

Quicker & accurate customer outcome

Hyperscience is also well known for its "hyper" focus on Human-in-the-Loop (HITL), shown in Figure 1. The company uses the term "Human Centered Automation," and this is more than just a catchy marketing term. To recall, almost all modern IDP systems (indeed, all back-office AI products) involve HITL, whereby when the machine is unsure of something or fails to reach a programmed threshold of confidence, it will flag the issue for a human to correct or clarify. Once the human does so, the system learns and uses that knowledge to automatically process future similar exceptions.

From its beginning, Hyperscience recognized that though AI and machines in general can work at speed, at volume, and with a high degree of accuracy, there will always be break points and limitations, for example when a machine reads handwriting. It has embraced the use of HITL far more extensively than others we have seen, and this dedication continues in its latest release, R36.

R36 has added significant enhancements to the version we reported on last year and supports Hyperscience's goal to expand beyond its recognized expertise with structured and semi-structured forms into the realm of unstructured documents.

R36 highlights include the following:

→ **Extracting data points from unstructured documents.** Hyperscience's Text Classification feature (first introduced in R34) has been significantly upgraded. Unstructured extraction is the process to extract data points from long documents that contain unstructured text, such as contracts or correspondence. This feature uses natural language processing (NLP) to analyze and organize unstructured text by user intent, sentiment, topic, or any custom labels based on custom business rules, and then generates alerts or emails based on these results. The most obvious first use case will be to assist customer service reps in any industry. Other use case examples are to classify the medical disorders listed in a patient intake form; to determine the sentiment expressed in a customer comment (e.g., "positive," "negative"); and to prioritize emails by customer intent (e.g., inquiries, complaints, requests to change account information).

Figure 2
Protecting a Personal Tax Identifier – Before and After
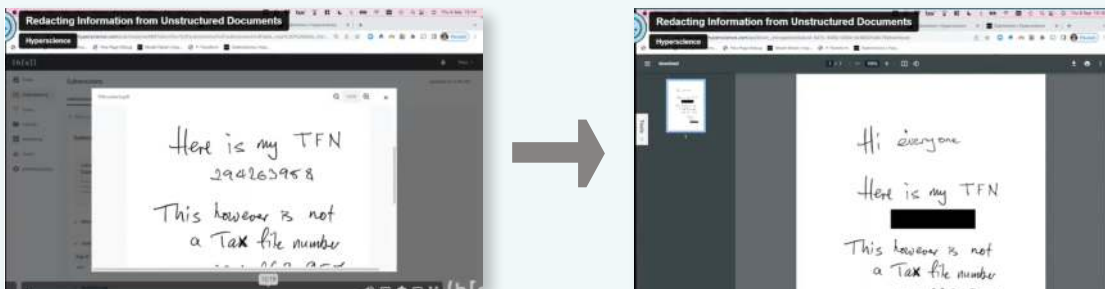
→ **Named entity recognition (NER).**
Using advanced NLP, Hyperscience NER can extract entities from unstructured documents without the need for third-party solutions or custom code.

→ **Redaction.** The redaction feature could be a game-changer for legal and healthcare use cases. Hyperscience can now detect and redact key personally identifiable information (PII) entities such as names, addresses, or organizations without having to pre-define a field list or train an extraction model through its built-in NER capabilities. In addition, users can specify custom entities using a mix of regular expressions and keywords that Hyperscience will look for in the document. After they are located, the keywords are redacted in the PDF output file. This can be applied to any text in any document, including handwriting. (Hyperscience focused much of its early R&D on training its system through deep learning to recognize handwriting, and it does as good a job as any we have seen. The handwriting recognition in their latest demo is exceptional.)

The demo showed the recognition and redaction of PII in a handwritten note. In the example in Figure 2, the tax ID has been redacted on the PDF file.

Our Opinion

Hyperscience has successfully transitioned from a cool VC-funded AI start-up into a mature IDP company working with top channel and technology partners and winning big deals with blue-chip customers. With its latest release, the company now offers IDP for every document class: structured, semi-structured, and unstructured. On its current trajectory, we think Hyperscience can challenge Kofax and ABBYY for IDP market leadership as measured by revenue.

We were curious to learn where Hyperscience is spending the investment money. The team's product roadmap reveals impressive ambition and vision behind the depth and breadth of their product plans. It is clear to us that the product development process will receive a significant portion of the investment. That bodes well for its customers.

Advice to Buyers

Hyperscience now has a highly advanced, end-to-end IDP platform for all document classes. If your IDP project is multi-faceted with a diverse group of documents including forms, bills, IDs, and contracts, Hyperscience is a viable one-stop shop. Insurance claims processing and customer onboarding/KYC are two obvious use cases where buyers should consider shortlisting Hyperscience. And with its focus on enterprise-level computing in the latest release, Hyperscience is also suitable for the largest deployments, where volume, speed, and accuracy are at a premium. With the impressive list of investors behind the company, buyers can be more confident of building a long-term partnership.

🔍 SOAR Analysis

Strengths

→ Exceptional core OCR and ML technology platform

→ Laser-focus dedication to Human-in-the-Loop user experience

Aspirations

→ Become the industry leader in IDP

→ Replace overly complex RPA deployments

Opportunities

→ Replace the legacy IDP leaders as the one-stop IDP shop

→ Invest in more industry-specific solutions

Results

→ Largest investment ever in any IDP company

→ May already be #3 or #4 in total IDP revenue

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

Contact us:

info@deep-analysis.net
+1 978 877 7915

Deep Analysis is an advisory firm that helps organizations understand and address the challenges of innovative and disruptive technologies in the enterprise software marketplace.

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

Led by Alan Pelz-Sharpe, the firm focuses on Information Management and the business application of Cloud, Artificial Intelligence, and Blockchain. Deep Analysis recently published the book "Practical Artificial Intelligence: An Enterprise Playbook," co-authored by Alan Pelz-Sharpe and Kashyap Kompella, outlining strategies for organizations to avoid pitfalls and successfully deploy AI.

Deep Analysis works with technology vendors to improve their understanding and provide actionable guidance on current and future market opportunities.

Yet, unlike traditional analyst firms, Deep Analysis takes a buyer-centric approach to its research and understands real-world buyer and market needs versus the "echo chamber" of the technology industry.