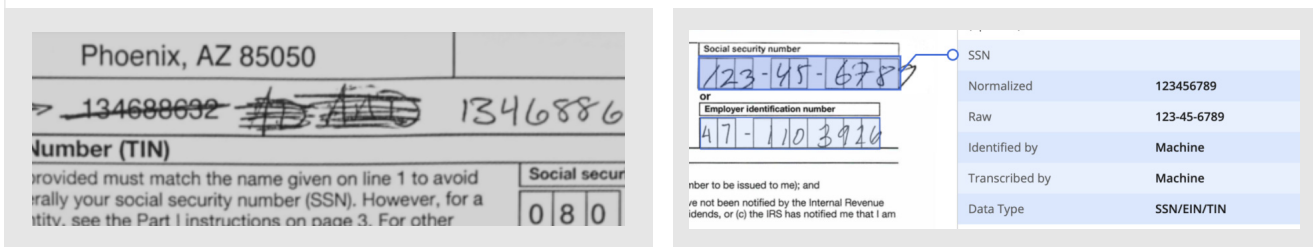


# DRIVING FEDERAL INNOVATION WITH INTELLIGENT DOCUMENT PROCESSING

## The Data Problem

Every day, public sector organizations must efficiently and effectively process millions of forms, applications and images to meet the needs of mission-critical workflows, supporting both warfighter and citizen-facing applications. These documents come in a wide variety of formats, oftentimes with poor readability (e.g. messy handwriting, fax marks, low resolution) and high variability, making it nearly impossible to reliably and efficiently process and extract data for downstream usage.



*Hyperscience is built to tackle messy, handwritten, real-world documents without sacrificing efficiency or quality. On the right, you can see the Hyperscience output of a structured tax form.*

Some agencies attempt to fix the problems by patching together various legacy systems to index and transcribe the documents. Even with investment, these systems struggle to support new use cases, unexpected surges in volume, or changing regulatory conditions.

**These outdated, manual workflows and legacy approaches are contributing to a massive data backlog and leading to an information bottleneck that affects all downstream processes. The result is strained systems, overworked employees and frustrated citizens who are left waiting for answers.**

Fortunately, however, advances in the areas of Artificial Intelligence (AI) and Machine Learning (ML) are causing organizations across the public and private sectors to reevaluate their existing document processing workflows. Government agencies at all levels are increasingly turning to intelligent automation to unlock and parse through data, gain efficiencies, improve constituent and employee satisfaction, and drive better agency outcomes.

# Intelligent Document Processing

Intelligent Document Processing, or IDP, is a powerful example of intelligent automation in action, providing tangible value for organizations that are looking to increase efficiency and accuracy when dealing with huge volumes of data.

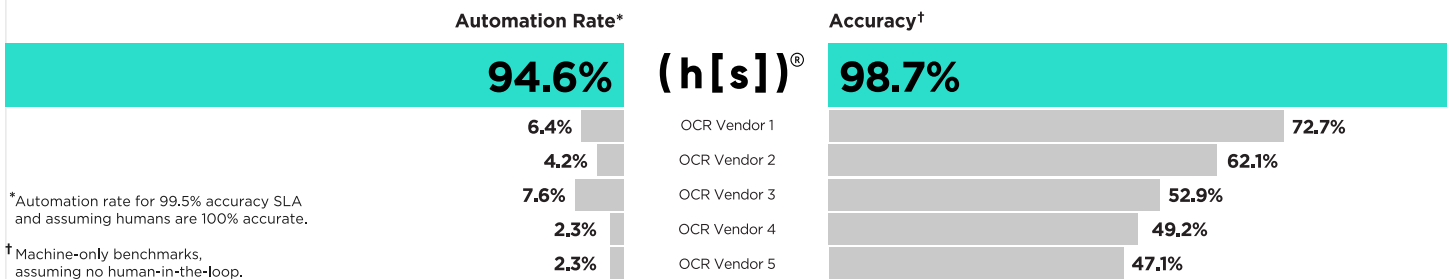
## IDP Processes Documents With Greater Speed And Accuracy Than Traditional Optical Character Recognition (OCR) Technology.

While OCR converts scanned images into machine-encoded text, typically transcribing it character by character, IDP captures data directly from diverse document types (e.g., email, text, pdf, and scanned pages), categorizes, and extracts relevant data for further processing using AI.

Leading IDP software has the intelligence to classify and extract increasingly diverse data inputs and uses advanced Machine Learning techniques to continue to learn on the data it's exposed to, driving lower error rates and greater automation over time.

## The Hyperscience Difference

The Hyperscience Intelligent Document Processing solution automatically classifies and extracts data across diverse documents, including handwritten forms, PDFs and low quality images with higher accuracy and automation than existing solutions, as demonstrated in the chart below:



Hyperscience captures data at the field level and measures accuracy accordingly, in contrast to outdated OCR tech. While legacy products might claim 90% accuracy at the character level, at Hyperscience, we know that a Social Security Number with one incorrect digit is 0% useful and consider it 0% accurate and 0% automated as a result.

Electronic documents and images are ingested by Hyperscience using a variety of standard input and output connection options including a REST API, MQ-based integrations, RPA software connectors, direct connections to file stores and more. Once ingested, the system classifies documents, maps them to a configured layout that contains the names, type, and structure of the desired output, and then automatically extracts data from each field. Once the document is fully transcribed, a structured output is available for downstream system consumption.

## Core Hyperscience product capabilities include:

- **Classification:** Automatically classify and separate incoming documents into appropriate document types given user-defined taxonomies.
- **Collation:** Group and order document packets, keeping shared context across separate documents within a transaction, to facilitate further processing or storage.
- **Deduplication:** Find duplicate pages among a document submission.
- **Extraction:** Handwritten, cursive and machine-printed text is transcribed and structured into JSON format.

## How Hyperscience is Driving Better Outcomes at the Federal & State Level:

Claims processing for a federal cabinet-level agency that provides integrated life-long healthcare services at nationwide medical centers and outpatient clinics.

Mortgage application packet processing and classification for FHA, single family low income housing.

Deduplication of incoming disability claim applications for largest non-defense federal agency.

Reducing manual work and enabling state-level case workers to move through backlog of processing Periodic Report forms for SNAP program.

Processing forms for the Shared Work CT Program, which is experiencing an influx of Covid-related hardship requests.

## Extraction Performance

Customers who choose Hyperscience can expect roughly 80% automation at over 98% accuracy on Day 1 with additional performance improvements over time. As the solution continues to learn on an organization's data, Hyperscience delivers up to 95% of data entry with over 99% accuracy, far surpassing the average industry accuracy rate which hovers around 55%.

Hyperscience leverages state-of-the-art Artificial Intelligence and Machine Learning techniques to achieve these performance breakthroughs. Its proprietary extraction engine combines dozens of different ML algorithms and does not rely on any off-the-shelf or third-party ML or data sets.

### **HYPERSCIENCE IS BUILT TO HANDLE CHALLENGING DOCUMENT CONDITIONS THAT CAUSE LEGACY PRODUCTS AND OCR TECH TO FAIL, INCLUDING MESSY HANDWRITING, TEXTURED BACKGROUNDS, LOW RESOLUTION IMAGES AND OTHER DISTORTIONS.**

Our models have trained to interpret the intent of fields, so we can tell the difference between what is a response and what should be "dropped out" like a field name or crossed out text. There is no need for separate dropout scans or red forms. There's also no need to programmatically exclude or filter out the label text or other form information or text that legacy technology may inadvertently capture.

Hyperscience uses a single extraction model for handwritten & machine typed text, versus most legacy solutions which require an operator to select which model to use, limiting scale and automation rates.

Hyperscience knows that addresses usually have numbers upfront and street names toward the middle and processes pages accordingly with separate models for each data type.

This avoids ambiguity when the system thinks something could be a “5” vs. an “S” and improves data extraction quality. In addition, our proprietary tech chases text outside the box to deliver best-in-class extraction. If the machine is only seeing 8 digits of a Social Security Number, it’ll gradually expand the crop to find the 9th digit.

## Accuracy And Machine Learning Improvements

Hyperscience takes a fundamentally different approach to document processing and data extraction.

### Compared to legacy OCR vendors or alternative automation solutions, Hyperscience treats accuracy as paramount.

Clients set their desired accuracy target (based on internal Service-Level-Agreements or other compliance criteria) and Hyperscience extracts data accordingly, automating as much of the work as possible.

The platform will automate where it is confident it can reach the target accuracy. When the machine is not confident it can achieve that accuracy, it will flag tasks for human supervision. In general, a higher accuracy target will lead to lower initial automation (i.e., increasing the accuracy target from 99% to 99.5% can impact automation STP rates by 5-10% or more). Customers “dial” the necessary accuracy target (e.g., 98%, 99% , 99.7%) and Hyperscience automates against that, using human supervision and consensus-driven quality assurance to drive lower error rates and higher automation over time.

#### ON DAY 1, CLIENTS CAN EXPECT:

> **80%** automation & > **98%** accuracy

Hyperscience continues to learn on an organization’s documents over time, driving lower error rates and higher automation.

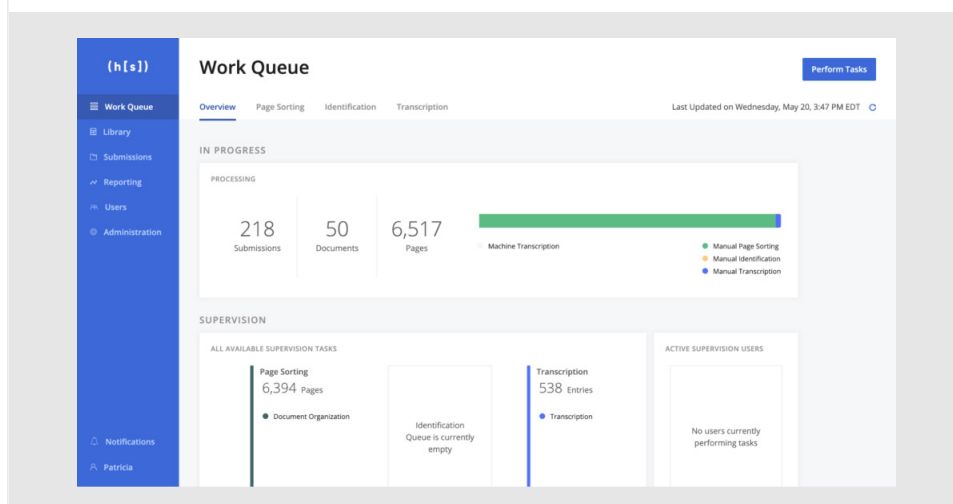
Hyperscience also measures accuracy at the field level, not the character level, since a Social Security or Account Number with one digit off is useless. While a legacy OCR product might extract “John” as “Jonn” and consider that extraction to be 75% accurate, Hyperscience knows that is 0% useful and would mark it as 0% accurate. This more approach to field-level accuracy ensures high-quality data is sent downstream for more reliable, streamlined processing.

## Intuitive User Interface

The Hyperscience Platform is designed for business users with no technical experience required, allowing for widespread adoption and workflow improvements. **HYPERSCIENCE DOESN'T REQUIRE CUSTOM BUILDS, HUNDREDS OF HOURS OF DEVELOPER RESOURCES AND COSTLY, TIME-INTENSIVE TRAINING JUST TO GET SET-UP AND BEGIN USE.** In addition, built-in reporting helps users understand the volume of pages processed by the system, the frequency at which human supervision is needed, and the accuracy of the work done in the system.

Below are some key reporting metrics, most of which can be filtered by date, by task and by document types, and can be exported as a .csv file for further analysis:

- **Automation:** The percentage of fields automatically processed by Hyperscience. The platform also reports the raw count of automatically processed fields and manually processed fields.
- **Field Output Accuracy:** The overall accuracy of the final output at the field level, including both machine processed fields and human processed fields. The margin of error and the number of fields considered in the calculation are included.
- **System Throughput:** The count of objects processed in the platform. Objects can be submissions, documents, pages, fields, and table cells.
- **Manual Accuracy vs. Machine Accuracy:** Manual accuracy is the field-level accuracy of manually processed human supervision tasks. Machine accuracy is the field-level accuracy of fields processed by Hyperscience. Hyperscience reports on both of these metrics.
- **System Transcription Sampled Errors:** To support optional application configuration, Hyperscience provides complete information on transcription errors, including a crop of the field, the machine transcribed value, and any manually transcribed values.
- **Time to Completion:** The average processing time from submission into the application to the completion of all Supervision tasks.
- **Supervision Volume:** The count of total supervision tasks, consensus tasks, and quality assurance tasks performed manually in the application. These KPIs can be filtered by date, by task type, and by document type.
- **Performance Distribution:** A visual representation of each human keyer's performance, including their average accuracy and the number of tasks they complete.
- **Usage Report:** Machine transcriptions of signatures, checkboxes, and entry fields; Supervision tasks; system configuration and environment-level variables; and application usage.



Example of the Hyperscience user interface, which includes built-in reporting dashboards.

## On-Premise Deployment

Hyperscience deploys on-premise in your private Cloud or physical servers, ensuring complete control over data. The application operates with standard ports and protocols, maintaining compliance, and utilizes baselined Linux VM images, enforcing existing security posture.

At this time, Hyperscience is offered exclusively as a software bundle to be installed on-premises in your datacenter or Cloud infrastructure, and is not available as a hosted Cloud capability. Therefore FedRAMP certification does not apply to Hyperscience.

In addition, Hyperscience releases software updates using an Agile Methodology and ships the latest version (including relevant patches) at the time of engagement

### FAQs:

- ✓ Hyperscience complies with HSPD-12 (smart card, CAC, PIV, two-factor authentication). Once configured, an external authentication service (SAML, AD, OIDC, etc.) handles all auth concerns, such as Multi-Factor Auth and Smart-card based Auth, which can be used for any or all users.
- ✓ The Hyperscience platform can be configured with standard secure ports and protocols that comply with FIPS 140-2 encryption algorithms.
- ✓ The Hyperscience software bundle runs in a set of containers that can be deployed on 1 or more Linux-based VMs. For GovCloud deployment, AWS EC2 instances can be provisioned as the underlying Linux VM to install the Hyperscience platform.
- ✗ There are no published COMPUTER SECURITY INCIDENT RESPONSE CAPABILITY (CSIRC) advisories for Hyperscience.
- ✗ There are no known “zero day” defects. Hyperscience can also provide automated scan results intended to detect common application security vulnerabilities, including those listed in OWASP Top 10.
- ✗ As a proprietary software Hyperscience is not listed in the NIST vulnerability database.

**Additional notes:** There is no specific STIG or SRG for Hyperscience, however the product can be installed on a STIG-compliant build of RHEL 7 or Ubuntu 18.04. Per ARCYBER OPOD 2018-097, published April 20, 2018, the RMF Assess Only process has replaced the Army CoN process.

**Current ATOs include:** FedRAMP IL2-Moderate within IBM’s Automation offering inside AWS GovCloud.

## About Hyperscience

*Hyperscience modernizes mission-critical processes and operations for Global 2000 organizations and government agencies. Since 2014, Hyperscience’s automation technology has helped data-centric companies parse through vast amounts of unstructured inputs and raw information to get to swifter and smarter business outcomes. Through the Hyperscience Platform, enterprises are empowered to transform their operations, and drive operational efficiency as well as human productivity by fully unlocking the power of their data.*

*The company has raised over \$190 million from investors including Tiger Global, BOND, Bessemer Venture Partners, Stripes, and FirstMark and has more than 300 employees with offices in New York City, Sofia, Bulgaria, London, UK, and Toronto, Canada.*

For more information, please visit [www.hyperscience.com](http://www.hyperscience.com)