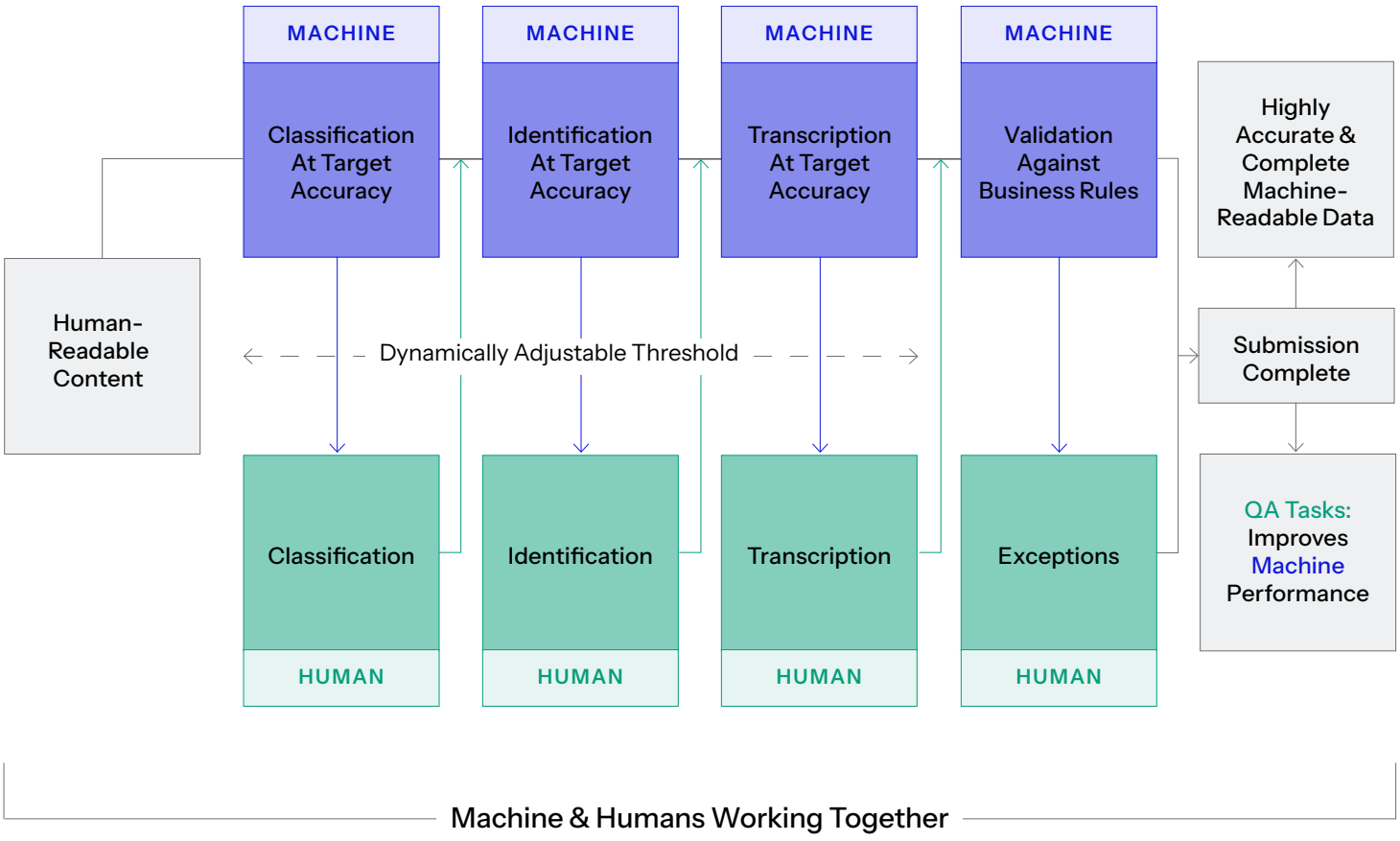
(Hyperscience AI Transparency Report)

How We Build Responsible Automation

Hyperscience Platform Overview and Goals_

At Hyperscience, we believe transparency is foundational to trustworthy AI. This report outlines how we develop, govern, and improve our AI models, ensuring alignment with our core principles: fairness, privacy, accountability, and human oversight.



AI Model Information_

Hyperscience employs over 30 distinct machine learning and deep learning models ("AI") to perform specific data extraction tasks. These models include both:

Model Type	Examples	Description
Pre-trained by Hyperscience	Rotation Detection, Optical Intelligent Character Recognition (OICR), Text Segmentation	General-purpose models trained on public or open-source datasets using proprietary methods
User-Inputted \ Fine-Tuned	Field Locator, Table Locator, Classification Model	Models trained or fine-tuned with data explicitly provided by the user

Human-in-the-Loop (HITL)

Hyperscience recognizes the critical role of human oversight in ensuring responsible AI deployment. Our "Human-in-the-Loop" (HITL) approach integrates human input at several key stages:

- **Manual Data Input**
  Examples: Account Opening Forms, Loan Application Forms, Credit Authorization Forms, Disclosure Forms, Consent Forms, etc.
- **Supervision of Machine Efforts**
  When a model's confidence in its prediction is low, the system prompts human validation.
- **Quality Assurance**
  Human validation efforts (separate from supervision) are used to verify model accuracy and can be configured to dynamically adjust the workload sent to human reviewers.

Model Performance Measurement

Model outcomes are primarily evaluated based on:

- **Automation Rate**
  The percentage of predictions performed by the machine compared to the total work executed (human + machine).
- **Accuracy Rate**
  The percentage of correct predictions out of the total number of predictions made.

Data Privacy, Security, and Compliancen_

Hyperscience is committed to maintaining the privacy and security of user data through configurable policies and robust data handling practices.

Data Handling & Redaction Pipeline

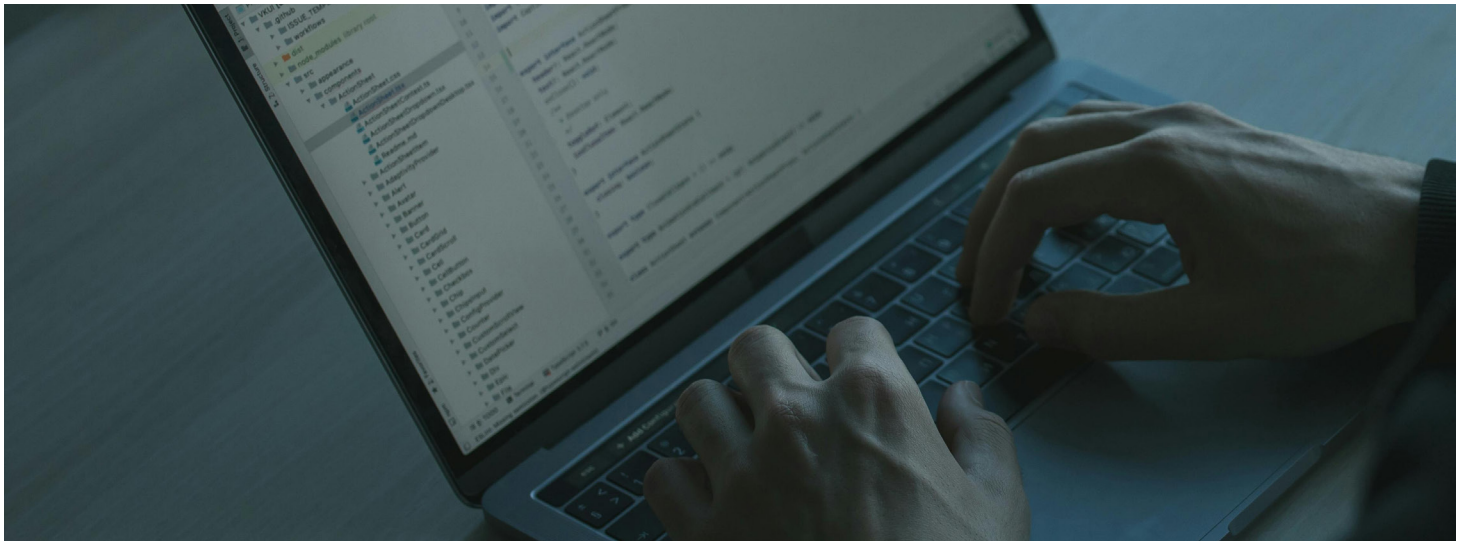
Customer data undergoes systematic redaction pipelines to remove Personally Identifiable Information (PII) before it is considered for model training. We only utilize data for which explicit permissions have been obtained (e.g., through data deals), ensuring that only de-identified, non-sensitive data is processed. For more information, please refer to our PII Data Deletion Policy documentation.

Model Training & Data Scope

Our AI models are developed and trained internally using only text segments and structured text-based data. We do not incorporate external or user-generated content outside of explicitly provided datasets for user-inputted models. This controlled approach significantly mitigates risks associated with large language model vulnerabilities, such as those outlined in the OWASP Top 10 for LLMs.

- **AI Bill of Materials (AIBOM)**
  We are taking strides in an AIBOM solution, providing a cutting-edge view of our AI infrastructure, including model details, dependencies, and training data lineage.
- **Regulatory & Contractual Compliance**
  Hyperscience adheres to GDPR and other relevant data protection regulations, ensuring secure data processing in accordance with contractual obligations and industry best practices.

By implementing these security measures and strict data handling protocols, Hyperscience ensures compliance, privacy, and robust AI security across our platform.



Oversight & Risk Mitigation_

Hyperscience mitigates potential AI-related risks through:

- **Targeted Model Application**
  Models are focused on specific data extraction tasks, avoiding broad or unintended consequences.
- **Explicit User Training Control**
  Users must actively opt-in to share data for training.
- **Deterministic Outputs**
  Models return consistent output for the same input.
- **Robust QA Process**
  Dedicated quality assurance workflows ensure accountability and safety.

Training Data Sourcing and Compliance_

Hyperscience utilizes a variety of datasets for training and evaluating our models, ensuring compliance and responsible data handling:

Dataset Category	Source	Use Case
Public Datasets	Research repositories (e.g., SROIE)	Research repositories (e.g., SROIE)
Synthetic & Internal	Generated by Hyperscience	Controlled testing and validation
Client & Partner Data	Via explicit agreements	Primary source for model training
CDRP Datasets	Anonymized customer data	Used for research and evaluation (not training)

By providing this detailed overview, Hyperscience aims to foster **[transparency]** and **[trust]** in our AI-powered platform. We are committed to continuous improvement in our AI practices and will update this report as needed.