Everest Group®

Seven Must-have Innovations in Intelligent Document Processing

Anil Vijayan, Partner

Ashwin Gopakumar, Practice Director

Akash Munjal, Senior Analyst

Contents

Executive Summary

Organizations – and their employees and customers – are increasingly being buried under an avalanche of documents and data. As business and regulatory complexities climb sharply, the effort to process these documents is exponentially more labor- and time-intensive, with significant manual effort that is not only mind-numbingly dull, but also prone to error. While enterprises can use technologies such as Optical Character Recognition (OCR) and template-based capture solutions for document digitization and data extraction, these tools barely scratch the surface in addressing the larger problem.

Evolving Intelligent Document Processing (IDP) solutions are the toolset many organizations need to help them automate complex document-intensive business processes to efficiently classify documents and extract data from them, while also validating and enriching the data and feeding it into downstream applications. IDP solutions help significantly reduce human effort in document processing, increasing efficiency, quality, and employee/customer satisfaction.

IDP's many benefits have spurred its adoption, but in many cases its implementation and ongoing management complexities have hindered more widespread acceptance and scaling. IDP solutions are introducing capabilities that address these implementation/management challenges, including pre-packaged solutions, support for complex document types, enhanced data enrichment, and advanced model monitoring, among others. These more sophisticated capabilities enable enterprises to adopt IDP more quickly, speed ROI, and ease business user concerns.

In this report, we:
- Provide an introduction to IDP
- Explore the current state of IDP
- Detail seven must-have capabilities that enable sustainable IDP adoption

Our report is best suited for
- Executives responsible for optimizing business processes
- Chief Operating Officers and other business leads charged with improving operations
- Digital transformation and automation leaders

Evolution of document processing: what was previously only possible for standard templates and forms is now accessible for documents of any complexity

Documents and documentation have long been integral to business processes, and rising business complexity and increasing regulatory burdens are only exacerbating the challenge. Today's organizations are flooded with documents across their business processes in structured (such as customer onboarding forms), semi-structured (such as invoices), and unstructured (such as contracts) forms. A significant amount of time and effort is spent in processing these documents – classifying them into relevant categories for further processing, extracting meaningful data for analysis, reviewing for data validity, cross-referencing with data from various enterprise systems, and manually keying data into different systems. To make matters worse, the tedious nature of these tasks negatively impacts employee experience. For some time, enterprises have struggled to automate these document-intensive processes to increase efficiency and improve the data-driven customer and employee experience.

Traditional document capture tools, such as Optical Character Recognition (OCR) and template-based capture solutions, are able to solve only part of the problem. These solutions can extract data from structured documents but are not able to handle documents with variations in structure, nor can they automate complex business workflows.

A new category of solutions has recently emerged that addresses these challenges: Intelligent Document Processing (IDP) solutions address these challenges at scale by transforming not only structured, but also semi-structured and unstructured documents into usable information.

Exhibit 1 depicts an enterprise-grade IDP solution workflow.

EXHIBIT 1

Enterprise-grade IDP solution

Source: Everest Group (2022)
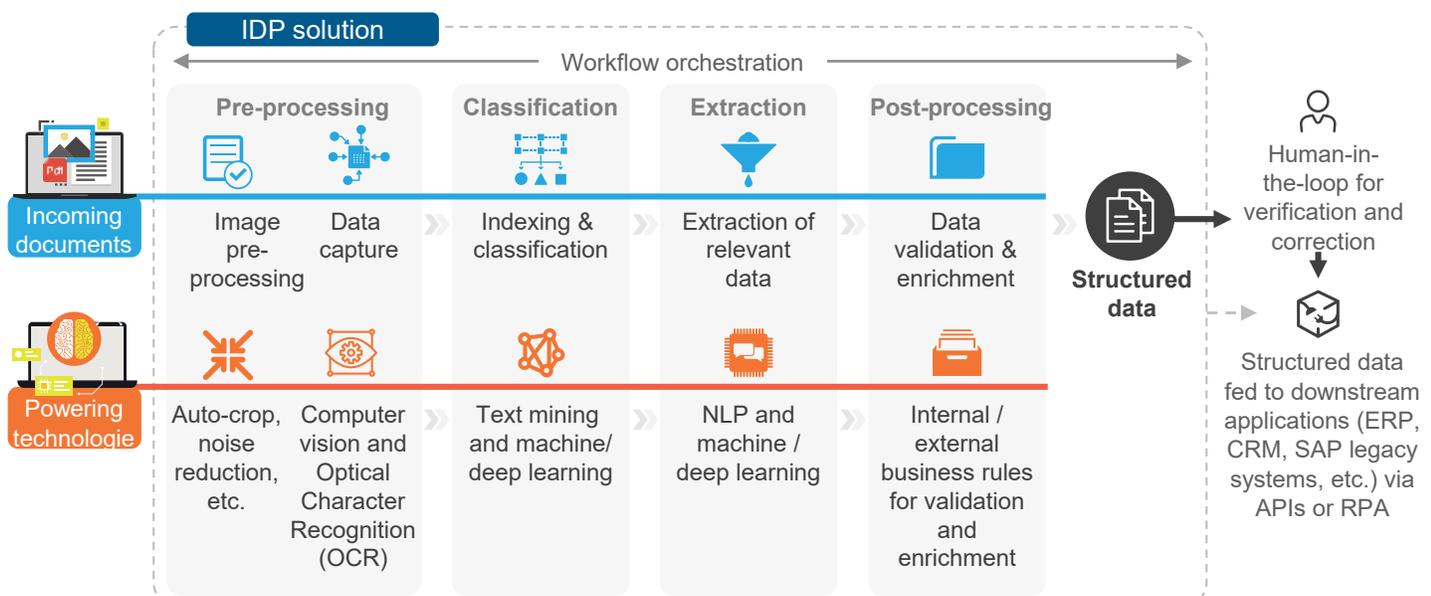
IDP software solutions automate complex document-intensive business processes – including digitization of documents – by leveraging AI to efficiently classify documents and extract data from them, validate and enrich the data from the documents, and once the data is structured, feed it into downstream applications. IDP solutions can process all types of documents – structured, semi-structured, and unstructured – efficiently.

IDP uses a multi-step process to automate document-intensive processes, assembling a broad set of capabilities to achieve automation outcomes. Combining these varied capabilities IDP is able to offer broader automation for complex document processing scenarios such as customer onboarding and KYC, mortgage loan processing, and healthcare claims processing.

Benefits of adopting IDP solutions

Through IDP solutions, enterprises can reduce their dependence on manual processing and achieve several operational, cost, and business benefits, as illustrated in Exhibit 2. For instance, IDP can help the Finance & Accounting function to automate invoice processing, extracting relevant data from invoices with varying structure, validating that data with existing data in ERP systems, managing validation and approval workflows, and updating enterprise systems, leading to high levels of Straight Through Processing (STP). Benefits include earlier invoice payments, which allow them to take advantage of early payment discounts, less manual effort, and a better employee and supplier experience. In customer-facing processes, the adoption of IDP can reduce turnaround time significantly, dramatically improving customer experience. For instance, the adoption of IDP in mortgage loan application processing can significantly reduce the time to approve loans significantly, resulting in a great customer experience and helping the mortgage provider to win more business.

EXHIBIT 2

Benefits of adopting IDP solutions

Source: Everest Group (2022)

Operational impact

- Streamlines document tracking
- Increases productivity and efficiency of the digital & non-digital workforces
- Speeds turnaround times by increasing Straight Through Processing (STP)
- Improves accuracy with minimum manual intervention
- Improves the employee experience
- Increases compliance and governance
- Improves customer satisfaction when used in customer-facing processes
- Enables business continuity and builds enterprise resilience

Cost impact

- Reduces the overall cost of processing large volumes of data
- Generates quick ROI

Top-line impact

Impacts top-line growth, for instance, by enabling the creation of new products

Through IDP solutions, enterprises can reduce their dependence on manual processing and achieve several operational, cost, and business benefits
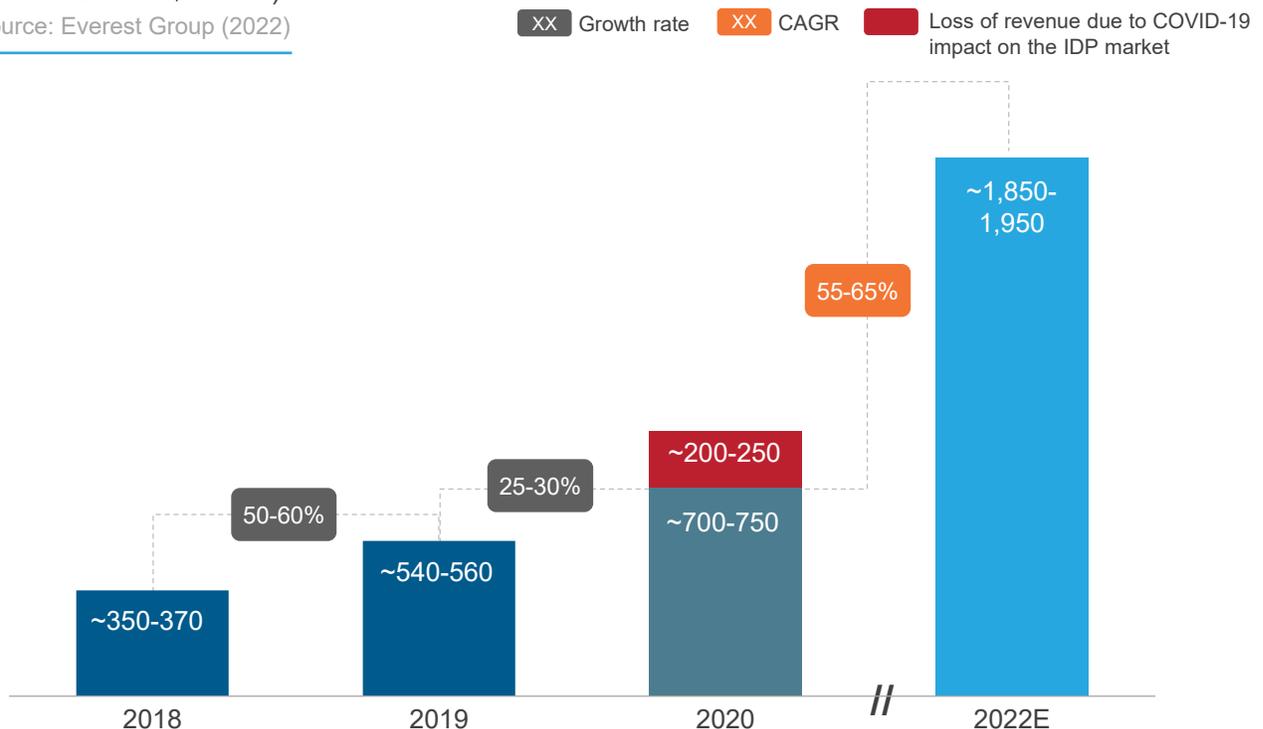
Enterprises are in dire need of IDP, as demonstrated by the adoption rates

Given its significant advantages, IDP adoption has been rising rapidly, with the IDP software license market growing at a rate of 25-30% from 2019 to 2020 (see Exhibit 3).

EXHIBIT 3

IDP adoption, 2018-22E
(Revenue in US$ million)
Source: Everest Group (2022)

| XX | Growth rate | XX | CAGR | | Loss of revenue due to COVID-19 impact on the IDP market |

~1,850-1,950

55-65%

~200-250

~700-750

25-30%

~540-560

50-60%

~350-370

2018   2019   2020   2022E

COVID-19 has further accelerated IDP adoption as enterprises looked to automation to solve new challenges and ensure business resiliency.

However, as IDP adoption has risen, enterprises face ongoing challenges in adopting IDP solutions, as described in Exhibit 4.

EXHIBIT 4

Key challenges in adopting IDP solutions

Source: Everest Group (2022)

Model training effort and complexity	• Some ML models require large volumes of sample data to train the solution to attain the desired accuracy levels
	• Training ML models depends not only on the volume of data, but also on the associated structured values for training purpose. These factors significantly affect the training approach. Variances in the data mix and image resolution used for training purposes can impact the accuracy level of solutions in production
	• Creating a repository of previous data and preparing it for training is a significant task
Data integration challenges	• IDP implementation projects often run into challenges with connecting to source and output systems
	• Building custom integrations to enterprise systems and third-party data sources for validations can delay implementation
Challenges with scaling up adoption	• Successful implementation of IDP solutions depends on the complexity of use cases. As enterprises attempt to scale adoption of IDP for different use cases, the inability of some solutions to process complex document types and multiple languages can hamper progress
	• The complexity of the IDP user interface itself can also lead to challenges with business user adoption and can cause disillusionment
Issues with model drift and degradation	• After they have been deployed, over time, ML models' prediction powers often degrade
	• If the model trains on poor quality inputs, it can learn from noise or inaccurate data entries, which will also degrade performance
	• Not identifying and addressing these issues early can lead to additional effort and rework
Managing continuous improvement	• IDP solutions technology is evolving; adopting enterprises may be challenged to keep pace with updates and take advantage of innovations
	• An inability to compare an organization's IDP performance against best-in-class solutions can also lead to underperformance and failure to realize the full potential of IDP

The complexity of training the ML models, integrating them with other data systems, and managing continuous improvement while addressing potential degradation in model performance makes IDP adoption challenging for many enterprises. Traditionally, long implementation times and the complexity of the user interface can also limit adoption among business users.

However, with the growing sophistication of the technology, some IDP solutions now offer capabilities to help address these challenges and ease adoption. These advanced IDP capabilities make it easier to adopt and manage these solutions while still taking advantage of their ML capabilities. In addition, the increased sophistication of these solutions helps enterprises address more difficult challenges such as complex unstructured document processing and multi-language document processing. In the next section, we take a look at these mission-critical capabilities that enable sustainable adoption of IDP across the enterprise.

Seven mission-critical elements of sustainable IDP

IDP solutions now offer a range of advanced capabilities (which we consider mission-critical) to solve enterprise challenges, ranging from packaged solution offerings to advanced data management and model monitoring features. Enterprises need to take into consideration these key capabilities when evaluating IDP solutions to ensure that they get the maximum ROI from their IDP programs.

Packaged
solutions

Continuous model
improvement

Complex document
processing capabilities

Enhanced data
enrichment

Advanced data
management

Flexible deployment
models

Model monitoring
and explainability

1. Packaged solutions

IDP solutions offer packaged solutions for specific horizontal or vertical processes that are customized to process specific document types such as invoices, claims, and KYC documents. While these solutions might often need to be customized to fit the enterprise specific requirements, they dramatically reduce the time to value for enterprises. Packaged solution offerings enable easier adoption of IDP, enhance the extent of automation possible, and ensure a quick ROI for enterprises. Certain packaged solutions offer high levels of document understanding in such a manner that they are closer to being digital workers rather than traditional software programs. As the market evolves, we expect an increase in the breadth and sophistication of packaged solution offerings in the IDP market. The capabilities offered through packaged solutions are explained below.

Pre-trained models
Packaged solutions are pre-trained on specific document types that are relevant for that process area, which improves the accuracy of extraction out-of-the-box while supporting continuous improvement in production.

Pre-built connectors
These solutions have pre-built connectors to the common enterprise systems that are relevant to that particular use case to speed deployment.

Workflow
The solutions have pre-built workflows for the specified process area, including workflows for review and approvals.

Process-specific capabilities
Packaged solutions are often powered with process-specific capabilities including business validation rules and AI prediction models that are required for specific use cases.

2. Complex document processing capabilities

As noted earlier, while IDP solutions have traditionally processed semi-structured documents such as invoices and purchase orders, some IDP solutions have expanded their capabilities to process more complex unstructured documents such as contracts, loan documents, emails, and annual reports. These unstructured documents do not conform to pre-defined data structures and lack keys to separate semantic elements. These capabilities help enterprises expand the scope of documents that can be processed with IDP, increasing their overall benefits.

Entity recognition from free-flowing text documents such as contracts and lease agreements

Classification of unstructured documents such as routing emails to appropriate teams

NLG capabilities to perform extractive and abstractive summary of lengthy and text-heavy documents

Search/comparison of documents based on meaning and intent

Object detection and classification capabilities for images using computer vision

Ability to process different data types such as handwritten text, logos, and signatures
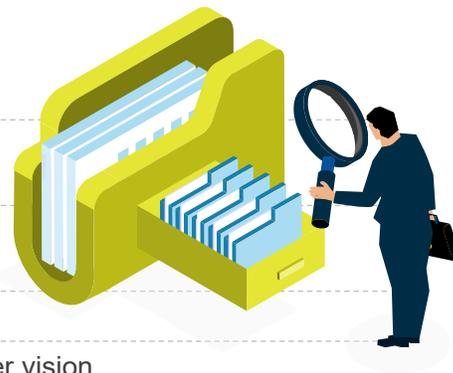
IDP solutions are also increasingly expanding their capability to support multiple languages for extraction. While IDP technology providers have most commonly supported English and Latin languages, many providers are also increasingly adding extraction support for Asian languages such as Chinese, Korean, and Japanese. IDP solutions are increasingly offering support for a wide set of languages and offer sophisticated capabilities such as support for extraction of multiple language content from the same document.

3. Continuous model improvement

Continuous model improvement in IDP involves the ongoing re-training of the Machine Learning (ML) models based on the document types encountered in production. IDP solutions leverage Human-In-The-Loop (HITL) capabilities to continuously train ML models with the latest data. IDP solutions are increasingly introducing capabilities to further increase the ease of use of continuous learning, including the advanced product capabilities for continuous improvement of ML models described below.

Better user interface for HITL
IDP solutions are offering better UI capabilities, such as taking operators directly to the fields that require review rather than having them to scroll through the entire list of fields, as well as offering autofill and suggestions speed and ease the correction process.

Benefits: These UI-based improvements increase ease-of-use of human-in-the-loop functionality to provide inputs for training the system.

Advanced controls to manage data quality and reliability
IDP solutions are offering advanced controls in the continuous learning process, including admin controls to review re-training datasets and models and controls to ensure the quality of the training datasets.

Benefits: These guardrails, which help manage the quality of data that goes into the training model, are a key component of continuous model development.

Federated learning to improve the model
Through the use of federated learning, IDP solutions are able to leverage the learnings across enterprises without sharing the underlying data. For instance, if one organization is testing a new use case, it can leverage the learnings in the model from other enterprise deployments instead of starting from scratch.

Benefits: With federated learning, enterprises' core data is protected, but they can still benefit from training other enterprises have already done, improving model accuracy and reducing the effort and time required to build the model.

Ongoing model updates to enterprises
IDP solution providers are also offering ongoing model updates to enterprises that allow them to make use of the latest advances in the ML algorithms and data sets. Enterprises can review the performance of new models made available by the IDP solution and choose if they want to switch to the new model or continue to use the existing one.

Benefits: Enterprises can take advantage of the ongoing improvements that IDP solutions are offering to improve IDP outcomes for different use cases on an ongoing basis.

4. Enhanced data enrichment

Data enrichment in IDP involves the use of data from different sources to validate the extracted fields and provide additional insights. IDP solutions are now offering more sophisticated data enrichment capabilities such as:

Pre-built integrations with third-party data sources
IDP solutions traditionally offer the capability to set up data validation rules, wherein data can be introduced via APIs from different existing systems within the enterprise to perform various validation checks, such as validation of data extracted from an invoice with the purchase order and goods receipts data from the ERP system. While validation with internal data sources is necessary for various processes, many processes also require validation with third-party data sources. Some IDP solution providers are introducing pre-built integrations within the software so that enterprises can pull in that data from other sources and ensure those additional validations are performed.

For example, as part of the KYC process, financial institutions are required to refer to a sanctions lists when conducting business with an individual or organization to ensure financial sanctions are observed and enforced. IDP solutions that offer pre-built integrations with these data sources and pre-configured validations drastically reduce the need for human intervention in the process. Enterprises can perform the validation checks out-of-the-box and ensure better compliance.

Industry benchmarking
IDP solution providers can also provide benchmarks to compare IDP solution performance for specific use cases across enterprises. Enterprises can benchmark their accuracy levels and other performance metrics with against industry peers and determine if their solution needs further training. Enterprises can also leverage broader process benchmark data through the product and measure overall process performance KPIs such as process time or failure rate and take necessary actions if there is significant difference between their process performance and industry standards.

5. Flexible deployment models

The impacts of the COVID-19 pandemic have compelled many enterprises to move their data and applications to the cloud to ensure business continuity and support remote working. With rising demand for cloud-based solutions, many cloud-native IDP solutions have emerged, that are built on microservices-based architecture; they can be deployed independently on elastic cloud infrastructures and can interact with each other through APIs. These microservices are containerized, which enables the solution to run efficiently across different computing environments without requiring much effort in porting. Cloud-native solutions also leverage Continuous Integration / Continuous Delivery (CI/CD) development methodologies to smooth the release process given frequent packaging and code releases to different environments. A cloud-native IDP architecture can provide significant advantages, as described below.

Product support benefits
- **Smooth upgrades:** Microservices breaks the whole application into smaller deployment units and smooths the rollout of upgrades and bug fixes without interrupting critical services
- **Quicker issue resolution:** The microservices architecture helps to quickly isolate issues and identify their root causes, and the CI/CD model allows for quick deployment of fixes for specific impacted microservice

- **Easier management of upgrades:** Cloud-native IDP solutions significantly reduce IT maintenance efforts for product upgrades

Operational benefits
- **Scalability:** Cloud-native architecture breaks the application into logical blocks, allowing rapid scaling up or down based on demand
- **Optimal resource utilization:** As containerized microservices can be scaled independently, they utilize fewer IT resources than traditional architectures, which replicate the entire application

Business benefits
- **Lower Total Cost of Ownership (TCO):** The lightweight architecture consumes fewer resources on the cloud, leading to cost savings and a lower TCO
- **Business continuity:** Cloud-native solutions help ensure business continuity through native support for remote access and strong disaster recovery capabilities
- **High availability and resilience:** These systems improve availability and resilience due to high resource availability on the cloud, coupled with the stability offered by loosely coupled, independent microservices

While cloud-based IDP solutions are gaining traction in the market, many enterprises still prefer on-premises deployments due to data security concerns and stringent industry regulations. In addition to on-premises and cloud deployments, some solutions provide hybrid deployment options, in which clients' data resides on-premises while other management functions, such as user authentication and updates, are hosted on cloud. As the market evolves, IDP solutions increasingly offer more flexibility in deployment models.

6. Advanced data management

A well-structured and centralized data management system is an essential component of the IDP solution to manage large datasets that are used to train the solution for various use cases. IDP solutions are advancing data management capabilities to offer end-to-end ML control as part of their IDP products, as described below.

The ability to collect data, normalize it, clean it, and perform quality checks to make sure data is reliable
Enabling the management of the entire ML workflow including separate environments, such as development, test, and production, directly through the IDP solution
Ability to generate synthetic training data for training models in scenarios where the available labelled training data is limited
Flexibility for enterprises to fine-tune ML models and run ML experiments to figure out how to achieve the best results
Allowing datasets to be divided into test sets, training sets, and validation sets as part of the ML workflow to enable proper training and validation of datasets with various ML models
Versioning capability for the training datasets and models for effective management

7. Model monitoring and explainability

Model monitoring

IDP solutions now offer advanced model monitoring capabilities through which enterprises can continuously monitor the performance of various ML models on an ongoing basis. Model performance may degrade over time. For example, if there are significant changes in document structure or the flow of unstructured text, models trained on older data may perform poorly, significantly increasing the need for human intervention. In these cases, IDP solutions can pinpoint the particular document variations or data elements that are leading to model degradation, making it easy for enterprises to identify and address these issues. Without dedicated production model monitoring, enterprises cannot identify root-cause of degradation. If model accuracy declines without detection, the results can severely impact business KPIs.

Model explainability

In AI-based solutions, the underlying ML models are often seen as "black boxes"; users input data and receive outputs but cannot comprehend or retrace how the algorithm produced the result. However, as adoption of AI-based solutions rises, more enterprises are interested in understanding the factors that drive outputs.

The model explainability capability is essential for enterprises to understand, appropriately trust, and effectively manage AI-based solutions. It allows the solution to describe an AI model, its expected impact, and potential biases. It helps characterize model accuracy, fairness, transparency, and outcomes in AI-powered decision-making. It is particularly useful in industries such as banking with stringent regulatory requirements that require transparency and control.

IDP solutions are increasingly offering explainability as part of their solutions. Explainability offers additional advantages as well; it:

- Helps identify the most influential data factors that affect the AI model
- Offers recommendations to make the AI model neutral and business supportive
- Optimizes model performance with constant evaluations and model prediction comparisons
- Increases confidence in the model output

Conclusion

Documents continue to be the mainstay of enterprise business processes. However, manual document processing is cost-intensive, error-prone, and slow – not to mention the impact it has on employee and customer satisfaction. Enterprises across industries are increasingly adopting Intelligent Document Processing (IDP) solutions for various document-intensive processes to solve these issues.

With the sophisticated capabilities that IDP solutions now offer, enterprises can adopt these solutions with relatively greater ease and more quickly realize an ROI. Advanced capabilities such as continuous model improvement, integration with third party sources, multi-language support, flexible deployment models, and model monitoring and explainability, also help enterprises address more complex document processing scenarios. With these advances in IDP, enterprises can unlock the next level of automation benefits in their organizations, while offering a better experience for their employees and customers.

Everest Group®
With you on the journey

Everest Group is a research firm focused on strategic IT, business services, engineering services, and sourcing. Our research also covers the technologies that power those processes and functions and the related talent trends and strategies. Our clients include leading global companies, service and technology providers, and investors. Clients use our services to guide their journeys to maximize operational and financial performance, transform experiences, and realize high-impact business outcomes. Details and in-depth content are available at **www.everestgrp.com**.