Go to Gartner Home

Gartner.

Licensed for Distribution

Market Guide for Intelligent Document Processing Solutions

Published 16 February 2022 - ID G00757528 - 23 min read

By Shubhangi Vashisth, Anthony Mullen, and 2 more

Intelligent document processing solutions are used to extract information from textual data, primarily documents, for analysis and workflow automation. Data and analytics leaders should use this research to evaluate IDP solutions for creating operational efficiencies and/or deriving insights.

Overview

Key Findings

- The intelligent document processing (IDP) market is dense with vendor offerings. New and commoditized natural language technology (NLT) techniques have lowered the barrier of entry. However, domain knowledge (such as industry-specific document training sets) and emerging technology capabilities differentiate vendor solutions.
- The use cases for IDP span across the spectrum of structured, semistructured and unstructured data. However, the majority of vendors are able to efficiently support only the use cases that deal with structured and semistructured data.
- Organizations often approach their IDP use cases as point solutions within specific departments. Increased awareness of vendor offerings and consolidation of product features will lead to a greater demand for platforms that can serve multiple use cases.

Recommendations

Data and analytics leaders looking to leverage Al-enabled IDP capabilities across business processes must:

 Decide between a point solution and a platform by articulating current requirements and also by considering future technological development and organizational growth (for example, support for written languages, unstructured data, domain-specific workflows and internal

departmental use cases). This will help not only select the right vendor but also avoid situations where you need to use alternative tools to plug gaps or add more capabilities.

- Evaluate vendor solutions by conducting detailed proofs of concept (POCs). Then, assess the
 overall quality and efficiency of the solution, the flexibility of the solution to efficiently handle
 diverse formats, ease of training and retuning, support for "human in the loop" workflows, and
 user experience.
- Ensure that labels and classifications are harmonized with broader master metadata management by using shared master metadata frameworks such as semantic approaches.

Market Definition

Intelligent document processing (IDP) solutions extract data to support automation of high-volume, repetitive document processing tasks and for analysis and insight. IDP uses natural language technologies and computer vision to extract data from structured and unstructured content, especially from documents, to support automation and augmentation.

Market Description

Although interest in IDP from enterprises has markedly increased in recent years, automation with document processing isn't new, and it has been signaled by many terms, including data capture, document AI, capture automation and more. Rapid capabilities in the last few years (transformers, more labeled data, advances in NLU and others) have fast-tracked document processing capabilities of solutions. Gartner estimates that intelligent document processing as a market is projected to reach \$4.8 billion in 2022.

The technological improvements are driven by advances in artificial intelligence, specifically:

- Computer vision can not only recognize characters that are both font-based (typed text) and
 cursive (handwritten) but also help decompose a document into chunks, providing contextual
 cues to the content within them. Recent IDP solutions retain spatial information related to
 characters and can also identify images.
- Natural language processing enables extracted characters to be treated as content that can be
 processed to identify and extract the factual elements (e.g., dates) within it, and that can be
 mapped to a data model (e.g., month/day/year).
- Advanced capabilities such as text analytics and knowledge graph provide insights into the data.

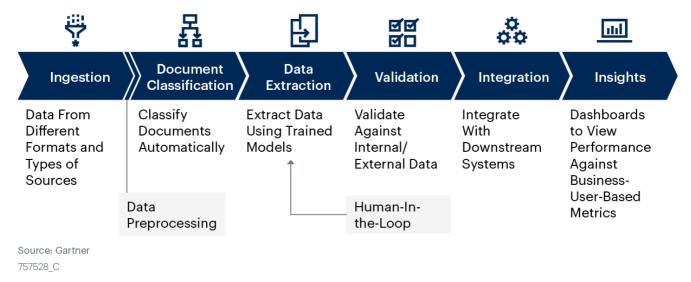
IDP platforms are therefore distinct from other document processing products in that they enable the processing of documents created exclusively for humans, as well as those for machines; they can adapt — or be easily adapted — to formats not previously processed; and they extract data in various models. As such, organizations can transfer document processing work currently done by people and other machines (where automation is the goal) to IDPs.

Figure 1 shows a typical IDP workflow. Some organizations may want to stop at data extraction while others may want to leverage a full workflow.

Figure 1: Different Phases in a Baseline IDP Workflow



Different Phases in a Baseline IDP Workflow



Gartner.

For a detailed understanding of the various phases and some differentiating capabilities of each phase, see Infographic: Understand Intelligent Document Processing.

Vendors in this market come from multiple starting points, including automation (e.g., AntWorks, Automation Anywhere and Blue Prism), document processing services (e.g., smartlayers and Applica), pure-play IDPs that include established vendors incorporating AI (e.g., Ephesoft) and new entrants starting with AI (e.g., Infrrd), and those from a long tail of other starting points, such as search (e.g., IntraFind). Here, we focus on vendors that sell a separate IDP product that can be integrated with other vendors' products to complete hyperautomation pipelines (see Competitive Landscape: Intelligent Document Processing Platform Providers).

Market Direction

Documents represent a trove of useful "information." Unlocking this data for language and process automation has seen demand for IDP solutions accelerate rapidly in the last 12 months. However, while IDP solutions are in high demand today, vendors and buyers should plan for both consolidation and specialization, not just within the IDP market, but also in the tightly clustered language automation ecosystem.

Gartner foresees the following drivers and restraints shaping IDP:

Competition from adjacent platforms. Insight engines, conversational AI platforms, text
analytics, robotic process automation (RPA), business process automation (BPA), no-code/lowcode application platforms (LCAPs) and content service platforms are all broadening both their
process and language automation offerings to include document workflow and mining tools.

Democratization of technology diminishing market share. As the use of large-scale language
models like BERT, GPT and T5 are commoditized (and less of a dark art), technical barriers to
this market will reduce and therefore increase vendor supply. Today, that additional supply,
while not able to compete on function, will likely compete on price. However, over time, buyers
will become more demanding of solutions, and this will cause consolidation.

- Multiform and multimodal solutions. Today, it is difficult to ensure that efforts for email
 automation can be used with conversational AI systems. The same is true when attempting to
 integrate document processing solutions with other AI initiatives. Over time, we expect to see
 more comprehensive language platforms that are able to integrate conversations, documents,
 emails and other combinations of form factor/modality. Generic providers without vertical
 specialization will find it increasingly difficult to compete against the synergy and reuse of
 multimodal and multi-form-factor platforms.
- Prioritization of information architecture vendors and semantic AI. Insights on concepts,
 objects and relationships in language must be realizable in various views and use cases and
 across different channels and modality. IDP vendors that are able to use insights for non-IDP
 use cases will command more of a strategic relationship with organizations by becoming more
 deeply embedded in their information architecture and comprehensive language services.

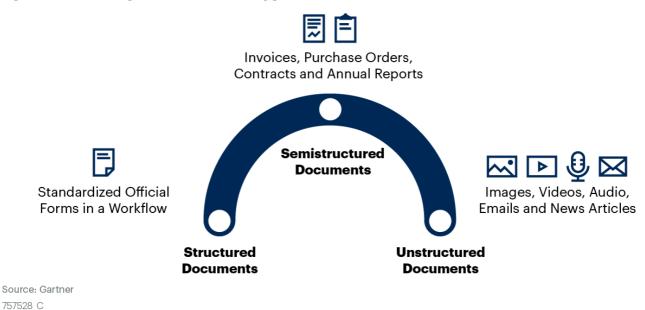
Market Analysis

The market for IDP has a diverse set of vendor offerings, including those from dedicated solution providers, optical character recognition (OCR) vendors, RPA vendors, cloud services, semantic platform providers and system integrators. Solutions for BPM, knowledge management (KM), insight engines and others may also offer IDP as embedded capabilities. As organizations deal with an ever-increasing volume of documents, it is imperative to use IDP for improving automation and process efficiencies. There exists a spectrum of use cases that organizations might want to leverage (see Figure 2).

Figure 2: Spectrum of Input Document Types and Use Cases



Spectrum of Input Document Types and Use Cases



Gartner

As a result, there are different approaches for organizations to invest in these solutions:

- Domain Specialist: Using IDP product/solution from vendors whose product/solution has specialization in particular industry or a particular use-case, like mortgage, loans or know your customer (KYC)
- Software Suites: Capability built into an existing enterprise software suite like RPA, ERP or other software platform
- Developer Platforms: Built and maintained using code or with custom development on top of APIs/SDKs
- No-Code/Low-Code Platforms: Using IDP/RPA technologies that allow the user to address a wide range of content types and to support a wide set of use cases

Table 1 provides a list of the core (minimum), differentiating and emerging capabilities of IDP solutions.

Table 1: Capabilities of IDP Solutions

Core	Ingest any type of structured and/or semistructured documents
	 Ability to extract data and insights from different elements (e.g., tables, charts and images) and format types (e.g., title, heading and bold)
	Support for multiple languages
	End-to-end process workflow automation
	Ease of use (low-code/no-code platform)
Differentiating	Process unstructured content, including handwritten content and signatures
	Pretrained off-the-shelf solutions for an industry or domain
	Small data approaches that minimize training data requirements
	 Professional services and software capabilities for human-in-the-loop validation
	Redaction of personally identifiable information (PII) and sensitive data
Emerging	Ability to integrate with taxonomies/ontologies/tags
	Search capabilities
	Natural language question answering (NLQA)
	Use of knowledge graphs for validation and business insights
	Data enrichment

Source: Gartner (February 2022)

Given the scope of this market and breadth of use cases, it can be difficult to identify which approach and tools are most appropriate. Most solutions, while useful, are tactical.

The following considerations can be applied to assist with solution selection:

- In the case of a single business function with an immediate need, consider application vendors with specific functionality that aligns directly with the business outcomes. Cloud-based solutions may offer additional flexibility and time to value over on-premises offerings.
- Keep in mind future growth and expansion in terms of support for language, diverse data types and formats. Check if you require only data extraction versus a workflow to fully process and

store the data.

• Check if the solution stores data for processing and what security controls are offered beyond data anonymization and encryption. Certain vendors comply with specific industry or geographic regulations such as the General Data Protection Regulation (GDPR).

Pricing Models

Pricing of IDP solutions is variable and is often not published. The dominant pricing model is annual subscription followed by usage-based pricing (e.g., per page or per document). Some vendors also offer workflow-based pricing, perpetual licenses, enterprise-level agreements, outcome-based pricing and professional services. The pricing range often starts from a few cents (per page or per document) for hundreds to thousands of pages/documents (threshold varies across vendors). Tiered pricing helps control costs as volumes go up.

Important considerations for IDP use cases include data type (structured or unstructured), language support, training, retraining requirements and professional services (for example, for human-in-the-loop validation) that impact the total cost of ownership (TCO). Vertical packs and complementary capabilities from adjacent submarkets such as text analytics also lead to price variations. When considering solutions, note that price may not be as important as quality. Additionally, quality evaluations may change significantly depending on how well a vendor can accommodate customization.

Representative Vendors

The vendors listed in this Market Guide do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings.

Market Introduction

This section provides a list of 40 IDP vendors (see Tables 2 and 3).

In selecting the right fit for your needs, you should consider the use case, vendor expertise in the use-case type, languages supported and other factors.

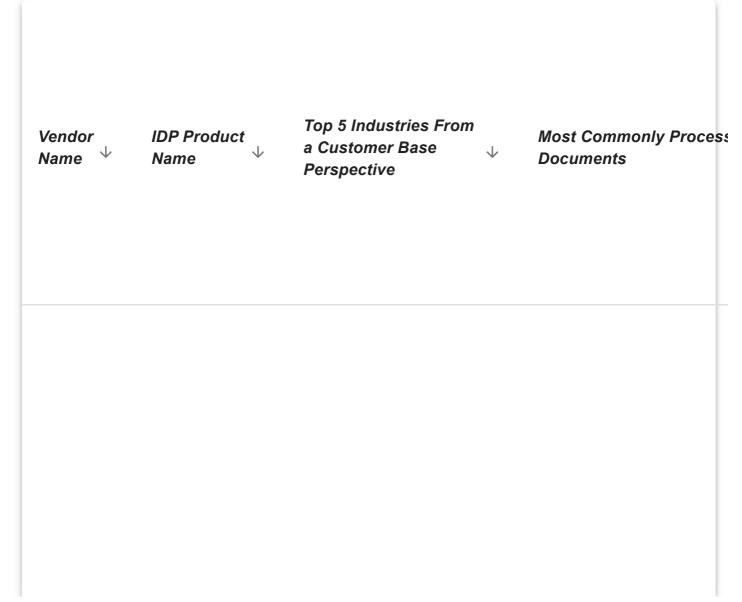
Table 2 lists the most common industries, documents processed, languages supported and deployment modes.

The following columns have given input categories:

- Vendor background:
 - RPA
 - KM/CM/semantic platforms
 - Pure-play IDP
 - System integrators/service providers

- OCR/data capture
- Model development and calibration:
 - Vendor-led
 - Client-led
- Input (level of structure): Most vendors offering capabilities to deal with unstructured data are, by default, able to cater to structured and semistructured data:
 - Structured (tabular)
 - Semistructured (outline)
 - Unstructured (free text)
 - Unstructured (free text + rich media)

Table 2: IDP Vendors by Industry, Document Types, Supported Languages, Model Development and Calibration, and Input



ABBYY	ABBYY Vantage, ABBYY FlexiCapture (including SDK, Cloud, FlexiCapture for Invoices), ABBYY FineReader Server, ABBYY FineReader Engine, ABBYY FineReader desktop and mobile applications, ABBYY mobile capture products	Banking, government, insurance, transportation and logistics, manufacturing	Invoices, purchase orders, re of lading, customs declaratic account opening and servicir processing, customer onboar proof-of-identity docs), claim contracts, lease agreements, immigration services forms, processing (K1 files, 1099s, 2, 1040s)
Alkymi	Alkymi Data Inbox and Alkymi Patterns	Banking, securities, insurance, communications, manufacturing	Capital call and distribution r account and transaction stat identification documents, inv
Altilia	Altilia Intelligent Automation Platform	Banking, insurance, utilities, retail, securities	Annual/sustainability/interim reports, balance sheets, Invo orders, Social media posts, n bank statements, paychecks
Appian	Appian Intelligent Document Processing	Banking, securities, manufacturing (life sciences and heavy industry), telecom	Account opening/closure, invalues orders, insurance claim education applications, medidevice licensing, government identity verification
Applica	Applica	Banking, securities, financial services, health services, insurance	Property sales contracts, incoverification docs, healthcare financial docs, debt collection insurance docs, free-text emplaboratory requisition forms, Commercial Code (UCC) filin financial reports

Automati on Anywhere	IQ Bot	Banking, financial services and insurance (BFSI), healthcare and pharma, professional services, consumer packaged goods (CPG), and state and local government and education (SLED)	Invoices, purchase orders (P orders, bank statements, bill statements, mortgage docur insurance claims, healthcare application forms
Blue Prism	Blue Prism Enterprise, Blue Prism Cloud	Banking and financial services, insurance, telecom, IT hardware, utilities and retail	Invoices, purchase orders, a statements, sales orders, import/export documents
Celaton	inSTREAM	Retail and e-commerce, transport and logistics, (warehousing, couriers and support services), general insurance, manufacturing (heavy industry and automotive), BPO and BPS, travel and leisure	Sales orders, employee reco customer correspondence, o invoices and accounts payal
Cinnamo n Al	IDP, NLP	Insurance (general and health), banking, manufacturing (consumer nondurable products and heavy industry)	Account opening and closur insurance claims
Cortical.i	Contract Intelligence and Message Intelligence	Insurance	Insurance policies, contracts submission emails, social m invoices
Datamati cs	Datamatics TruCap+ IDP	Banking and financial services, health insurance, manufacturing, transportation (motor passenger and freight), healthcare	Invoice, identity documents, order, trade finance, bill of la waybill, packing list, bank sta tax return documents,

DocDigiti zer	"PowerCapture": DocDigitizer Cognitive Data Extraction "WorldObjects": DocDigitizer Semantic Data Platform	Banking, telecom, utilities, manufacturing, government,	Invoices/receipts, tax returns passport/ID cards, employee proof of address, bank stater
EdgeVerv e Systems	XtractEdge Platform 2.8	Banking, insurance (general and health), manufacturing (consumer nondurable products), IT service and software	Contracts, claim forms, invoi application forms, artwork ar financial statements, payslip purchase orders, governmen ID cards, customer onboardin documents
Eigen Technolo gies	Eigen	Banking, financial services, general insurance, manufacturing (supply chain), real estate	Vendor contracts, invoices, mortgages, prospectuses, ba statements, unsecured lendin tax forms, brokerage statema annual reports, derivatives co
Ephesoft	Transact, Semantik	Banking, government, healthcare, manufacturing, insurance	Invoices, government forms, applications, utility bills, clair patient records, application f medical billing, financial stat employee records
Evolution Al	Evolution Transcribe and Evolution NLP	Banking, general insurance, communications, media and services including information technology services and software	KYC, financial reports, bank statements, trade documents mortgage applications, insur claims
Exponent ial Al	Intelligent Document Processing	Health insurance, manufacturing (life sciences and healthcare products), healthcare, banking, retail	Medical charts, lab reports, c adverse event literature, heal claims, preauthorization forn regulatory updates/reports, r

Golem.ai	DocuChecker	Banking, government, retail, insurance, electric and gas utilities	Tax notice, proof of address, voter identity cards, driver's l salary slips, registration cert government-emitted certificadiplomas
Hypatos	Hypatos Studio	Manufacturing (automotive, heavy industry, natural resources and materials), general insurance, banking, retail, airlines	Sales invoice, payment remine confirmation, commercial ta statements, receipts, delivery annual report, bank statements damage reports, titles and depayslips
Hypersci ence	Hyperscience Platform	Government, financial services, insurance, BPO and professional services, healthcare	Mortgage and loan application invoices, insurance claims, concepts, bank statements, fin account enrollment, bills of HCFAs, prescription forms, Wother tax forms
IBM	IBM Cloud Pak for Business Automation; includes Automation Document Process and Datacap	Banking, financial services, government insurance, healthcare	Sales orders, bank statemen insurance claims, KYC, morte applications, government se shipping documents, accour opening/closure, claims rela financial accounting services
Indico Data	Indico Intelligent Process Automation ("Indico IPA")	Insurance, banking, securities, manufacturing (heavy industry), real estate	Claims, contracts, leases, mo invoices, Association for Coo Operations Research and Development (ACORD) forms syndicated loan documents, images

Infinia ML	Infinia Document Processing and Infinia Document Insights (IDP,	Healthcare, legal/compliance/technology and operations, finance	Contracts, legal documents, records or forms, lab tests, on trial documents, requests for (RFPs), call transcripts, final documents, scanned marked documents with media
Infrrd	Titan IDP Platform	Health insurance, banking, manufacturing (heavy industry)	Invoices, receipts, bill of lad reports, W-2, payslips, bank statements, mortgage docu insurance documents, claim documents, engineering dra product catalogs
inovoo	NOVO CxP	Healthcare, insurance, government, retail, transportation	General mailroom documen emails including. attachmer images, objects), account o documents, claims docume invoice, order documents, professional teeth cleaning, Euro Payments Area (SEPA) specific health forms, insura claims
IntraFind	Document Analyzer	Manufacturing (automotive, consumer nondurable products, and natural resources and materials), government, retail	More than 600 file formats, OCR, can be processed, incl contracts
Kanverse	Kanverse Intelligent Document Processing	Manufacturing (IT hardware), retail, general insurance, aviation	Accounts payable invoices, ACORD forms, insurance supplemental forms

Kofax	Kofax TotalAgility	Banking, general insurance, government, healthcare, manufacturing (natural resources and materials)	Structured, semistructured a unstructured documents wit without tabular data
Larsen & Toubro Infotech (LTI)	Aspect by Fosfor	General insurance, energy resources and processing, banking, life sciences and healthcare products, government	Sales invoices, KYC docume contract, mortgage applicat insurance policy quotations, binder documents, mixed bacheques, trade finance docu
Metamaz e	Metamaze	Banking, insurance, logistics and supply chain, government, manufacturing	Mortgage applications, insuclaims, supply chain docum invoices, prospectus
OpenText	OpenText Intelligent Capture and OpenText Magellan	Banking, insurance, government, healthcare, utilities	Invoices, medical claims, m loan applications, tax forms and contracts
Perfect Memory	Semantic platform, As-a- Brain Technology	Media and entertainment, telecom, retail, sports	Textual, images and videos
qBotica	DoqumentAl	Financial banking, renewable energy resources, transportation, insurance, retail	Invoices, purchase orders, b statements, purchase requis forms, settlement summary utility bills, emails (unstruct content), contractual agreer forms (passports, driver's lin national IDs)

Retarus	Retarus Intelligent Capture Services	Manufacturing, retail, wholesale trade, banking and securities, healthcare	Sales orders, invoice proces response		
Rossum	Rossum	IT services and software, banking, retail, transportation, consumer nondurable products	Invoices, receipts, PO, credi bills, delivery orders, remitta advices, bonds, bills of ladir quotations		
smartlay ers	smartlayers	Banking, insurance, telecom, utilities, transportation	Structured and unstructured documents, without charts diagrams, unstructured and documents, handwritten textabular data, invoices (more labels processed), reports		
UiPath	UiPath Document Understanding	Financial services, healthcare, insurance, government, manufacturing	Invoices, contracts, purchas insurance claims, patient fo statements		
WorkFusi on	Intelligent Automation Cloud	Banking, securities, insurance, healthcare, transportation	Account opening/closing, K mortgage application, bank statements, insurance clain money-laundering (AML), co life cycle management, Lon Interbank Offered Rate (LIB transition, physician creden		
Xtracta	Data Extraction Software	Transportation, wholesale trade, restaurants and hotels, government	Invoices, expense receipts, orders, contracts, property documents, insurance form certificates of goods and b lading, manifests and so or customs purposes		

Source: Gartner (February 2022)

Table 3: Ingestion Formats Supported by the IDP Product

Vendors $_{\downarrow}$	Email ↓	SMS ↓	XLS ↓	Images $_{\psi}$	Audio $_{\downarrow}$	Video
ABBYY	Υ	Υ	Υ	Υ	Р	Р
Alkymi	Υ	Υ	Υ	Υ	N	N
Altilia	Υ	R	Υ	Υ	R	R
Appian	Υ	Р	Υ	Υ	Υ	Υ
Applica	Υ	Υ	Υ	Υ	N	N
Automation Anywhere	Υ	N	N	Υ	N	N
Blue Prism	Υ	R	R	R	R	R
Celaton	Υ	R	Υ	Υ	Р	Υ
Cinnamon Al	Υ	Υ	Υ	Υ	Υ	R
Cortical.io	Υ	Υ	Υ	N	N	N
Datamatics	Υ	N	Υ	Y	R	R

/endors $_{\downarrow}$	Email ↓	SMS ↓	XLS $_{\downarrow}$	Images $_{\downarrow}$	Audio ↓	Video
DocDigitizer	Υ	N	N	Υ	N	N
EdgeVerve Systems	R	N	R	Υ	N	R
Eigen Technologies	Y	Υ	Υ	Υ	N	N
Ephesoft	Υ	N	Υ	Υ	N	N
Evolution Al	Υ	N	Υ	Υ	N	N
Exponential Al	Υ	Υ	Υ	Υ	Р	Р
Golem.ai	Υ	Υ	Υ	Р	Р	N
Hypatos	Υ	N	R	N	N	N
Hyperscience	Υ	N	Υ	Υ	N	N
IBM	Υ	N	Υ	Υ	Υ	N
Indico Data	R	N	Υ	Υ	Р	Υ

Vendors $_{\downarrow}$	Email ↓	SMS ↓	XLS $_{\downarrow}$	Images ↓	Audio $_{\downarrow}$	Video
Infinia ML	Υ	N	Υ	Υ	N	N
Infrrd	R	N	R	Υ	N	N
inovoo	Υ	Υ	Υ	Υ	Р	Р
IntraFind	Υ	N	Υ	Р	Р	Р
Kanverse	Υ	N	R	Υ	N	N
Kofax	Υ	Υ	Υ	Υ	Υ	N
LTI	R	N	Υ	Υ	R	R
Metamaze	Υ	N	R	Υ	N	N
OpenText	Υ	N	Υ	Υ	Υ	Υ
Perfect Memory	Υ	Р	Υ	Y	Y	Υ
qBotica	Υ	Υ	Υ	Υ	Р	N
Retarus	Υ	Υ	Υ	Υ	N	N

Vendors ↓	Email ↓	SMS ↓	XLS $_{\downarrow}$	Images $_{\downarrow}$	Audio $_{\downarrow}$	Video 、
Rossum	Υ	N	R	Υ	N	N
smartlayers	Y	R	R	Υ	N	N
UiPath	Υ	Υ	Υ	Υ	N	N
WorkFusion	Υ	Υ	Υ	Υ	N	N
Xtracta	Υ	N	Υ	Υ	N	N

Legend: Y = Available; N = Not available; R = In the roadmap; P = Currently using partner solution for this cap

Source: Gartner (February 2022)

Market Recommendations

- Develop a list of use cases that require data extraction and specify the data types (structured or unstructured) from which data needs to be extracted. Also include other requirements such as language support, workflow capabilities, integrations and vertical-specific needs.
- Factor in any future growth or specialized requirements to help select between platforms and specialized off-the-shelf applications.
- Shortlist vendors based on the above requirements. Utilize this Market Guide and the Competitive Landscape research (see Competitive Landscape: Intelligent Document Processing Platform Providers) for a long list of vendors to choose from.
- Conduct POCs and evaluate solutions for accuracy, efficiency, ease of use, security and overall
 cost.

Note 1: Representative Vendor Selection

Vendors were selected based on the following:

- Those whose core product met the definition of "IDP"
- Those most mentioned in client interactions
- Broad awareness in the market, as identified through external review of publications and other public research

Note 2: Gartner's Initial Market Coverage

This Market Guide provides Gartner's initial coverage of the market and focuses on the market definition, rationale for the market and market dynamics.



Become a Client

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."

About Careers Newsroom Policies Site Index IT Glossary Gartner Blog Network Contact Send Feedback

Gartner

© 2023 Gartner, Inc. and/or its Affiliates. All Rights Reserved.