aws marketplace

Unstructured

Reviews, tips, and advice from real users








Powered by PeerSpot

Contents

Product Recap . . . . . 3 - 5

Valuable Features . . . . . 6 - 12

Other Solutions Considered . . . . . 13 - 14

ROI 15

Use Case . . . . . 16 - 19

Setup 20

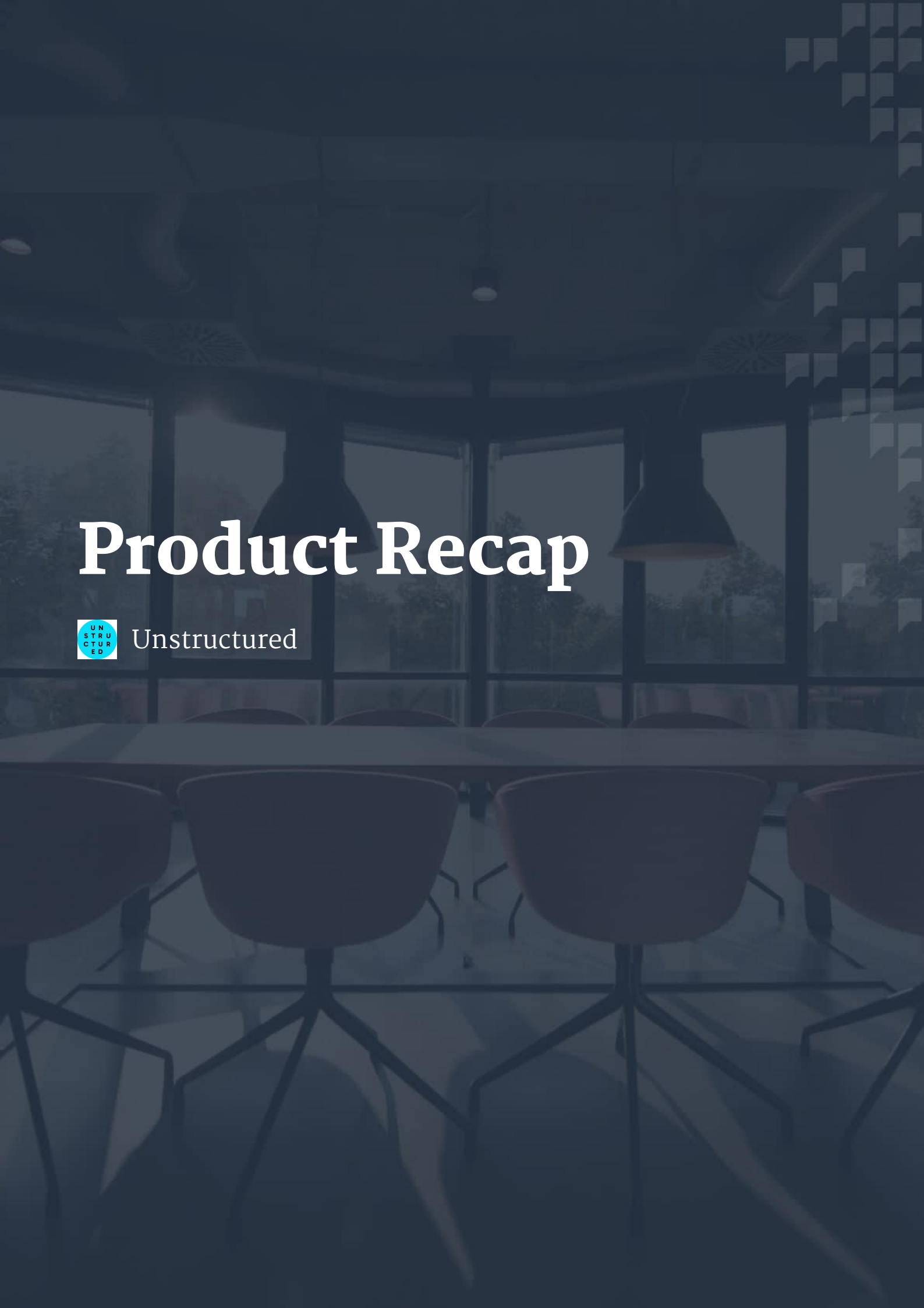
Customer Service and Support 21

Other Advice . . . . . 22 - 26

Trends . . . . . 27 - 28

About PeerSpot . . . . . 29 - 30

©2026 PeerSpot, All Rights Reserved



Product Recap

Unstructured

Unstructured Recap

Unstructured prepares documents and other unstructured files for search, analytics, and retrieval-augmented generation by extracting content, adding metadata, and producing pipeline-ready outputs.

Unstructured helps engineering and data teams connect content repositories to downstream AI and analytics workflows. It ingests files from sources such as cloud storage and collaboration platforms, then partitions formats including PDF, HTML, DOCX, PPTX, email, and images into usable text and elements. Processing options can include OCR, metadata extraction, document cleaning, and chunking, with results sent to destinations such as vector databases and data platforms. Teams can use APIs, SDKs, and deployment options to build ingestion pipelines that fit existing architectures, reducing the custom parsing work often required before documents can support search or language-model applications.

Which Unstructured features matter most?

- **Broad file parsing:** Extracts text and document elements from common office, web, email, PDF, and image formats.
- **Source connectors:** Ingests content from supported repositories and storage systems to feed repeatable workflows.
- **Document enrichment:** Applies options such as OCR, metadata extraction, and content cleaning to make extracted data more useful.
- **Chunking for retrieval:** Breaks content into segments suited to indexing and retrieval-augmented generation.
- **Pipeline integrations:** Sends processed content to supported vector stores and other downstream destinations through APIs and developer tools.
- **Deployment flexibility:** Offers implementation paths suited to different development and infrastructure requirements.

What benefits and ROI should reviews help you evaluate?

- **Parsing quality:** Look for feedback on extraction accuracy across your actual file types, including complex layouts and scanned documents.
- **Integration effort:** Assess how easily connectors, APIs, and output destinations fit existing data and AI pipelines.
- **Time to production:** Check whether teams can configure, test, and maintain ingestion workflows without extensive custom code.
- **Throughput and reliability:** Review performance at expected processing volumes, along with handling of failures and changing source content.
- **Total cost:** Compare processing, infrastructure, and ongoing maintenance costs with the effort required by alternative approaches.

©2026 PeerSpot, All Rights Reserved

- **Governance and control:** Evaluate deployment, access, and data-handling options against security and compliance requirements.

In financial services, teams can process reports, forms, and policy documents for search, compliance workflows, or internal knowledge retrieval. Healthcare organizations may prepare clinical and administrative documents for search or AI applications, subject to applicable privacy and security controls. Legal teams can index contracts and case materials, while manufacturers can make manuals, maintenance records, and technical documentation easier to retrieve. Customer support operations can also transform knowledge articles and service records into content for search and AI assistants.

©2026 PeerSpot, All Rights Reserved

Valuable Features

Excerpts from real customer reviews on PeerSpot:

- "Smart chunking is probably the biggest practical win because it breaks documents by structure, not just raw length, so the chunks are more meaningful for retrieval and RAG, which usually means better answers and less cleanup later."



TejaswiniAleti

Member Technical at ADP

- "Unstructured outshines other options for us because it provided OCR out of the box, as well as PDF parsers such as pdfminer.six, and helped us with semantic chunking by extracting paragraph titles, subtitles, different heading levels, and page structure information."



Deepak Unknown

Data Scientist at a educational organization with 201-500 employees

- "With Unstructured, I got the result in just one model call."



Yagnesh Unknown

Senior AI/ML Engineer at a consultancy with 11-50 employees

©2026 PeerSpot, All Rights Reserved

"The faster document ingestion has resulted in customer satisfaction, leading to higher quality answers using AI that improved customer satisfaction and NPS score."



Validated user

Product Manager at a tech vendor with 11-50 employees

"The best part of Unstructured is the integrations and connectors, which I really liked. They were seamless, and that is what I particularly liked about it."



Hemanth S

Applied AI at a tech services company with 11-50 employees

"Unstructured helped my organization by saving me hours, since I was able to load data fast compared to other solutions."



Validated user

AI Engineer at a university with 1,001-5,000 employees

©2026 PeerSpot, All Rights Reserved

What users had to say about valuable features:

"Unstructured can process many types of files, such as PDF, HTML, and CSV. It also has very good documentation, which makes it easy to use, and it is a very popular library, so it gets updated quickly.

"The documentation was quite easy to understand, and that is how it helped me. Frequent updates have not helped me in a particular use case per se, but in a vibrant library that is updated very regularly, they are always helpful.."

Yagnesh Unknown
Senior AI/ML Engineer at a consultancy with 11-50 employees

Read full review

"In my opinion, the best features of Unstructured are that it is easy and friendly to use and has a quick start time.

"What makes the best feature stand out is the user experience with Unstructured. It is easy to use.

"Among additional features, I would mention integrations, and I appreciate that Unstructured is easy to use, especially in Python.."

Validated user
AI Engineer at a university with 1,001-5,000 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"We wanted to do a quick test before actually paying, so we wanted something out of the box, such as the open-source Docker deployment that Unstructured provides. At the same time, we wanted other extractors out of the box. We wanted an OCR extractor to be part of the library that we would use along with structured parsers. That is where Unstructured outshined other options for us.

"For example, we were parsing PDF documents. Unstructured provided OCR out of the box, as well as PDF parsers such as pdfminer.six.

"We specifically focused on pdfminer.six because we wanted to do semantic chunking. For semantic chunking, we needed to know the paragraph titles, the subtitles, the different heading levels, and the page structure information. That is where Unstructured had the biggest impact for us, because it helped us with semantic chunking.."

Deepak Unknown
Data Scientist at a educational organization with 201-500 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"What stood out to me most about Unstructured is the scheduling and getting the data, along with MCP, which I really appreciate and which made my work faster. Everything I used was in a free trial, and I was impressed with what it does for a free trial. I can also integrate it with MongoDB and other tools. Altogether, I was satisfied.

"In my opinion, the best features Unstructured offers are the ETL (extract, transform, load) capabilities, MCP, and the schedulers I worked with on this pipeline. That is what I explored, and I took the best of it.

"The most helpful part of the ETL process was the initial setup. When I was starting a project, I wanted fast prototyping, and the prototyping was quick enough to get the work done. I really liked the usability from a customer point of view. The best part of Unstructured is the integrations and connectors, which I really liked. They were seamless, and that is what I particularly liked about it.."

Hemanth S
Applied AI at a tech services company with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"We are now using Unstructured every day, and it is useful when we want answers and AI to be used on a PDF or something similar. We use Unstructured to convert it into a vector database to make retrieval augmentation or any kind of AI processes easy.

"The document parsing stands out, as the document ingestion is very fast in Unstructured, 20 to 40% faster than the industry products available. If HR wants to upload a PDF on our platform, we use Unstructured to digest the data, and it is 20 to 40% faster than other solutions.

"The faster document ingestion has resulted in customer satisfaction, leading to higher quality answers using AI that improved customer satisfaction and NPS score. NPS has improved by at least 10 to 15 points since we started using Unstructured, not only for data digestion but also for retrieving data when we have to use AI or RAG.."

Validated user

Product Manager at a tech vendor with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"The best features are smart chunking, broad file type support, and the ability to pull from multiple sources and send data to multiple destinations. For a team like ours, those matter because they save time on document cleanup and make the output more usable for AI workflows.

Smart chunking stands out because it uses document structure and metadata to break content into more useful pieces for retrieval. Support for many file types, including PDFs, images, and other messy enterprise files, helps when your source data is inconsistent. Connectors and orchestration allow ingestion from many sources and loading to many destinations without building everything from scratch. Structured output for AI makes it easier to feed data into RAG, search, or agentic systems.

Smart chunking is probably the biggest practical win. It breaks documents by structure, not just raw length, so the chunks are more meaningful for retrieval and RAG, which usually means better answers and less cleanup later. Instead of splitting text blindly, Unstructured uses document elements and metadata, so the output is easier to search and reuse. That made the workflow smoother because we spent less time fixing bad chunks and more time on the actual AI logic.

The biggest productivity gain was fewer manual steps in preprocessing. Once the chunks were cleaner, it became easier to build and test downstream search or chatbot flows. That usually improves project speed and consistency.

Unstructured had a positive impact mainly by cutting manual preprocessing work and making the output more reliable for downstream AI tasks. The most measurable benefit I would mention is time saved because smart chunking and cleaner ingestion reduce the amount of document cleanup and rework the team has to do. A rough estimate I can provide is around 40 to 60% savings, especially for document cleanup and chunk review.."

TejaswiniAleti
Member Technical at ADP

Read full review

©2026 PeerSpot, All Rights Reserved

Other Solutions Considered

"We did not use a direct alternative before. We were handling it manually, and Unstructured was a much better fit for automation, speed, and cleaner, AI-ready output.."

TejaswiniAleti
Member Technical at ADP

Read full review

"We did evaluate a few alternatives, mainly Docling and LlamaParse, and some older OCR or IDP style approaches, depending on the workflow. We have used some of them before using Unstructured.."

TejaswiniAleti
Member Technical at ADP

Read full review

©2026 PeerSpot, All Rights Reserved

"We previously used PyPDF, PyMuPDF, and pdfminer.six, and we also used Unstructured. We did not switch away from those solutions, as we were using them alongside Unstructured, but they were very limited in their set of capabilities.

"They could only parse structured or readable PDFs. They could not parse unstructured content or images in the PDFs. That is where Unstructured outshines them, because it reuses those libraries but also provides OCR and image reading. The mix of image reading plus structured information is the reason for switching.."

Deepak Unknown
Data Scientist at a educational organization with 201-500 employees

Read full review

"Unstructured was the first tool I used, so I have not explored much else. I wanted to explore, and there were two to three other options, including Firecrawl, Bright Data, and Google Cloud supporting BigQuery. At that time I was exploring options, so I came across Unstructured, and my initial impression was quite good. It does the work. The MCP documentation is good, and the community is good altogether.."

Hemanth S
Applied AI at a tech services company with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

ROI

Real user quotes about their ROI:

"There was a clear ROI, mainly through faster processing and lower manual effort. A reasonable way to describe it is that we saved roughly 40 to 60% of the manual work in document preprocessing, which translated into faster turnaround and less time spent on cleanup.."

TejaswiniAleti
Member Technical at ADP

Read full review

©2026 PeerSpot, All Rights Reserved

Use Case

"I use Unstructured to process a bunch of messy PDFs and internal documents so they could be used in an AI search or RAG workflow. Instead of manually cleaning the files, it helped extract and chunk the content into a format that was easier to work with.."

TejaswiniAleti
Member Technical at ADP

Read full review

"My main use case for Unstructured is loading my AI data.

"A specific example of how I use Unstructured to load my AI data is parsing unstructured data from a bunch of different paths that I have.

"I appreciate that Unstructured is an open-source Python library, so it is easy to use and implement in my workflow.."

Validated user
AI Engineer at a university with 1,001-5,000 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"I was using Unstructured to parse documents and extract their text verbatim for a RAG application.

"As a specific example, we parsed the Confluence Wikia for our RAG. The Confluence Wikia consists of HTML-like pages containing information about our products, their releases, and the changes. We wanted to parse those documents, extract their information, and use it for our purpose. Those pages also contained PDF release documents, which we wanted to parse and use for the RAG application.."

Deepak Unknown
Data Scientist at a educational organization with 201-500 employees

Read full review

"We are a culture operating system that analyzes organizational culture, and we have an AI bot that joins calls to create structured culture intelligence reports. When we talk about HR, there are HR policies, PDFs, and performance documents generated by the HR or human resource department in the company. If we need to digest that data, we use Unstructured to create a vector database of these unstructured data.

"If an HR manager wants to use HR policies, HR documents, and performance data in Instill, they can upload their document, and we use Unstructured to convert those PDFs into a vector-based database.."

Validated user
Product Manager at a tech vendor with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"I do not use Unstructured in my day-to-day work. My main use case for Unstructured is to parse complex PDFs and get the various components of the PDFs readily, so that I do not have to break the PDF into images, tables, and other elements myself. Unstructured does that for me with its models and logic.

"A specific example is a complex PDF that contained various images, tables, and text. I gave the PDF to Unstructured, and it returned the images, tables, and other elements as different objects in Python. I can now use the tables and images separately, rather than having them embedded in the PDF as one page, and I can process them as much as I want and however I want.."

Yagnesh Unknown
Senior AI/ML Engineer at a consultancy with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"My main use case for Unstructured is to see the benchmarking stories and then determine how to get into the data. I have gone through the basic AI layer, and I did so through MCP, so the data was delivered to my agent.

"For a recent project, I used Unstructured mostly through MCP, which is easily accessible. I connected through it first, and it was so easy to use. I then got the data I required, especially for ETL pipelines, which cover extract, transform, and load. I got this data pipeline setup in place.

"With MCP, you can get all the documentation and structure all the data together. For example, on a particular day, we can get the data from some point using a scheduler, so I got scheduled data that way.."

Hemanth S
Applied AI at a tech services company with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

Setup

The setup process involves configuring and preparing the product or service for use, which may include tasks such as installation, account creation, initial configuration, and troubleshooting any issues that may arise. Below you can find real user quotes about the setup process.

"It was reasonable. The pricing was usage-based, setup cost was fairly low, and licensing was straightforward since everything was handled through our AWS billing and subscription flow.."

TejaswiniAleti

Member Technical at ADP

Read full review

"Unstructured is deployed in our organization on a private cloud, which is SAP Cloud, specifically SAP BTP. In SAP BTP, we were using Cloud Foundry. It was our own custom deployment, so it was easier for us to evaluate. The only part we had to take care of from Unstructured was the Docker part. The rest of the pricing numbers were easy for us to set up because we were very transparent about it, and we could turn it.."

Deepak Unknown

Data Scientist at a educational organization with 201-500 employees

Read full review

©2026 PeerSpot, All Rights Reserved

Customer Service and Support

"We did not opt for any customer support in our implementation. However, on GitHub and the forums, the team is very responsive, a full 10 out of 10. I would rate the customer support nine out of 10.."

Deepak Unknown
Data Scientist at a educational organization with 201-500 employees

Read full review

"Customer support appears to be good and fairly structured. Unstructured offers support through various channels. I would give customer support an 8 out of 10. It looks well-organized with solid documentation and clear support channels, but there is still room to improve responsiveness and hands-on guidance for tricky cases.."

TejaswiniAleti
Member Technical at ADP

Read full review

©2026 PeerSpot, All Rights Reserved

Other Advice

“My reason for choosing Unstructured is that Unstructured is comfortable to use and I enjoy it, but there are always things that could be made better.

“The accuracy and reliability of Unstructured's AI output is pretty good, but I still prefer using other models such as Claude and OpenAI..”

Validated user
AI Engineer at a university with 1,001-5,000 employees

Read full review

“Features and usability are fine, and it is one of the best products available.

“I chose a rating of 10 out of 10 because they are very focused on doing what they do at the best quality and speed, and what they are not doing is outside their scope. They claim faster processing and converting into a vector database faster, building a vector database from unstructured data, which they provide at a very fast speed and quality.

“The governance and security regarding Unstructured's AI capabilities are good, as we have SOC 2 and other compliance certificates from Unstructured. I give this product a rating of 10 out of 10..”

Validated user
Product Manager at a tech vendor with 11-50 employees

Read full review

“Unstructured looks strong on security and governance for this kind of AI data

©2026 PeerSpot, All Rights Reserved

platform. Its public documentation mentions encryption in transit and at rest, secrets management, RBAC, audit-friendly logging, and compliance alignment with standards like SOC 2 Type 2, ISO 27001, GDPR, FedRAMP, and CMMC 2.0 Level 2.

The setup is cloud-based and typically in a SaaS or dedicated instance model.

I would tell anyone to start with one clear use case and test it on your messiest real documents first. That gives the fastest read on whether it will actually save time in your workflow.

I would rate Unstructured an 8 out of 10 overall. It is strong enough to make a real difference in document ingestion and chunking, but there is still room for improvement in edge case handling and visibility..”

TejaswiniAleti
Member Technical at ADP

Read full review

©2026 PeerSpot, All Rights Reserved

"I rate Unstructured 8 out of 10.

"I chose 8 out of 10 because of its overall usability, documentation, good support, frequent changes, and very good community.

"The AI capabilities of Unstructured, including the accuracy and reliability of its output, are quite good, and I would give it 8 out of 10 for accuracy and reliability. Anyone can use Unstructured for production-grade applications.

"My recommendation to others is to read the documentation first or take a short course on Unstructured, and only start using it once you have done that. If you just use AI tools to write code that uses Unstructured, you will lose most of its functionalities.."

Yagnesh Unknown
Senior AI/ML Engineer at a consultancy with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"I rate it eight out of 10, considering it is open source. Because it is open source out of the box, you can quickly test it before going for a paid enterprise edition.

"With the enterprise edition, they maintain it, while with the open-source edition, you maintain it. It gives you both options and does not restrict you to a single business model of theirs.

"Unstructured's AI governance and security are pretty open, because you are aware of which libraries it uses. For example, even for OCR, it uses Tesseract and PyTorch, so you are familiar with which models it is using. That transparency helps us evaluate security and governance out of the box.

"Had it been a black box or something proprietary of their own, that would have been a turndown. If we know which models it is using and what the limitations of those models are, it becomes very easy to evaluate it when we have to submit the governance report for our project.."

Deepak Unknown
Data Scientist at a educational organization with 201-500 employees

Read full review

©2026 PeerSpot, All Rights Reserved

"On a scale of 1 to 10, my rating for Unstructured is 8.5, and my final overall rating is 8.

"I did not check the details of Unstructured's governance and security much. However, I assume that it follows a set of standards for Europe and the US, based on what they mentioned in their initial terms and conditions, and I believe in that.

"I would put the accuracy of Unstructured's output at 80% and give its reliability a 9 out of 10. For a prototype project, the accuracy met my expectations and is more than enough. It is perfect.

"We have not deployed Unstructured in my organization. We have just made a basic flow out of whatever cloud you are providing.

"My advice to others is that Unstructured is the best modern way of handling data and data pipelines. It offers MCP, easy and seamless integrations, and well-documented material, as well as the pipeline itself. Altogether, I would suggest that everyone use it to build an MVP as quickly as possible, for the use case I worked on.."

Hemanth S
Applied AI at a tech services company with 11-50 employees

Read full review

©2026 PeerSpot, All Rights Reserved

Top Industries

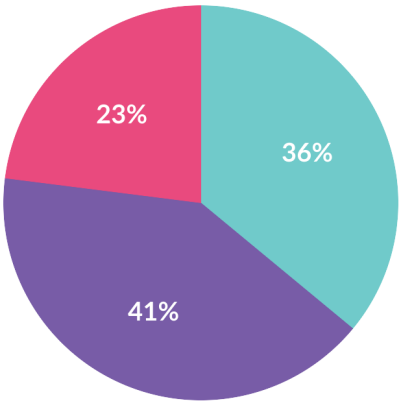
by visitors reading reviews

Construction Company	28%
Financial Services Firm	13%
Comms Service Provider	11%
Manufacturing Company	11%

Company Size

by reviewers

by visitors reading reviews



Large Enterprise Midsize Enterprise Small Business

©2026 PeerSpot, All Rights Reserved

About this buyer's guide

Thanks for downloading this PeerSpot report.

The summaries, overviews and recaps in this report are all based on real user feedback and reviews collected by PeerSpot's team. Every reviewer on PeerSpot has been authenticated with our triple authentication process. This is done to ensure that every review provided is an unbiased review from a real user.

Get a custom version of this report... Personalized for you!

Please note that this is a generic report based on reviews and opinions from the collective PeerSpot community. We offer a customized report of solutions recommended for you based on:

- Your industry
- Company size
- Which solutions you're already considering

The customized report will include recommendations for you based on what other people like you are using and researching.

Answer a few questions in our short wizard to get your customized report.
Get your personalized report here

©2026 PeerSpot, All Rights Reserved

About PeerSpot

PeerSpot is the leading review site for cloud, AI, and business software. We created PeerSpot to provide a trusted platform to share information about software, applications, and services. Since 2012, over 22 million people have used PeerSpot to choose the right software for their business.

PeerSpot helps tech professionals by providing:

- A list of products recommended by real users
- In-depth reviews, including pros and cons
- Specific information to help you choose the best vendor for your needs

Use PeerSpot to:

- Read and post reviews of products
- Access over 30,000 buyer's guides and comparison reports
- Request or share information about functionality, quality, and pricing

Join PeerSpot to connect with peers to help you:

- Get immediate answers to questions
- Validate vendor claims
- Exchange tips for getting the best deals with vendor

Visit PeerSpot: www.peerspot.com

PeerSpot

244 5th Avenue, Suite R-230 • New York, NY 10001

reports@peerspot.com

+1 646.328.1944

©2026 PeerSpot, All Rights Reserved