

# Amazon Textract

# Amazon Textract - Introduction

- Amazon Textract is a fully managed machine learning service that automatically extracts text and data from scanned documents that goes beyond simple optical character recognition (OCR) to identify, understand, and extract data from forms and tables.
- Textract uses machine learning to instantly read and process any type of document, accurately extracting text, forms, tables and, other data without the need for any manual effort or custom code.
- With Textract you can quickly automate manual document activities, enabling you to process millions of document pages in hours.

# Amazon Textract - Features

- **Optical Character Recognition (OCR):** Amazon Textract uses Optical Character Recognition (OCR) technology to automatically detect printed text and numbers in a scan or rendering of a document, such as a legal document or a scan of a book.
- **Form Extraction:** Amazon Textract enables you to detect key-value pairs in document images automatically so that you can retain the inherent context of the document without any manual intervention. A key-value pair is a set of linked data items.
- **Table Extraction:** Amazon Textract preserves the composition of data stored in tables during extraction. You can use this feature to automatically load the extracted data into a database using a predefined schema.

# Amazon Textract - Benefits

- **Easily implement human reviews:** Amazon Textract is directly integrated with Amazon Augmented AI (Amazon A2I) so you can easily implement human review of text extracted from documents.
- **Security & Compliance:** Textract can be used for workloads that are subject to Service Organization Control (SOC) compliance, and International Organization for Standardization (ISO) compliance as well as PCI, HIPAA, and GDPR.
- **Extract structured & unstructured data quickly and accurately:** Amazon Textract uses artificial intelligence to “read” documents, to extract not only text but also tables, forms, and other structured data without configuration, training, or custom code.

# Amazon Textract - Pricing

- With Amazon Textract, you pay only for what you use.
- There are no minimum fees and no upfront commitments.
- Amazon Textract charges you for each page you process and whether you extract only text from documents or text with tables and/or form data.
- A single page may contain between 0 and 3,000 words.
- New customers can analyze up to 1,000 pages per month using the Detecting Document Text API and up to 100 pages per month using the Analyze Document API, for the first three months.