CORSEN AI Academy

Document AI: OCR and Document Intelligence

Course syllabus

Version 2026-11

Overview

Intelligent document extraction with OCR 3: text recognition, annotations, tables, document QnA, and batch processing.

Learning objectives

- Call the Mistral OCR API in Python to convert a document into structured text
- Extract tables, headers, and pagination from a document without losing its original structure
- Configure the Annotations API to extract structured metadata, including in batch processing
- Query one or more documents with Document QnA and obtain sourced answers
- Integrate OCR into a RAG pipeline and industrialize the document pipeline in production

Audience and prerequisites

Developers, engineers, and integrators looking to leverage Mistral AI via API in their applications.

Ability to read Python and JSON, command line basics. A Mistral AI Studio API key to try the examples.

Duration and format

Estimated duration : 1 h 30 min. 100% online self-paced course, accessible 24/7 from a browser. Progress saved, resume anytime. Estimated duration: 1 h 30 min.

Content

- Document AI — Overview (3 lessons)
- Advanced OCR (5 lessons)
- Document Annotations (4 lessons)
- Document QnA (4 lessons)

Assessment

Practice exercises proposed throughout the lessons. Each lesson ends with its Key takeaways, supporting self-assessment. A final quiz of 5 corrected questions, at the end of the last lesson, validates acquired skills.

Accessibility

Self-paced course accessible to people with disabilities: text content compatible with screen readers, keyboard navigation, and compliant contrasts. For any adaptation needs, email hello@corsen.ai.

Price and recognition

Free

Certificate of completion issued by Corsen AI.

Access time

Immediate after account creation and email confirmation.

Corsen AI — Paris, France — SIRET 10297040700015
academy@corsen.ai — https://academy.corsen.ai