What is OCR and how does it work?

OCR (optical character recognition) converts virtually any kind of image containing written text (typed, handwritten or printed) into machine-readable text data.

**Unstructured Documents**
- Layouts that change over time or between sources.
- Data that is not always in the same location.
- Long PDFs made of multiple document types.

**Noisy Documents**
- Noisy, has colors, low quality
- Inconsistent orientation
- Data with contextual relationships

**Complex Documents**
- Complex tables
- Stamps, logos, symbols
- Graphs, charts, images, handwriting

**Use Cases**
- Data Entry Automation
- Bar Code Scanning
- Number Plate Recognition
- Personal Verification
- E-commerce Payments
- Social Media Monitoring
- Verifying Legal Documents
- Multilingual Document
- Medical Billing

Industry we serve
- Retail
- BFSI
- Government
- Education
- Healthcare
- Manufacturing
- Technology
- Transport & Logistics

**What is OCR?**
OCR (optical character recognition) converts virtually any kind of image containing written text (typed, handwritten or printed) into machine-readable text data.

**OCR Market Size**
- $10.6 bn Market size value in 2022
- $33.4 bn Revenue forecast in 2030
- 15.4% Growth Rate (CAGR) from 2022 to 2030

**The Process of OCR**
- Input Image
  - Preprocessing
  - Text Detection
  - Text Recognition
  - Verifying the Accuracy
- Output
  - Text

**Benefits of Automated OCR Workflows**
- Improved Accuracy
- Reduced Vulnerability & Risk Exposure
- Streamlined Processes
- Improved Productivity
- Lower Operational Costs
- Increased Speed-up
- More Storage Space
- Better Accessibility
- Improved Speed-up
- Elimination of Manual Data Entry
- Improved Accuracy
- Improved Productivity
- Lower Operational Costs
- Increased Speed-up
- More Storage Space
- Better Accessibility
- Improved Speed-up
- Elimination of Manual Data Entry

**Challenges**
- Unstructured Documents
  - Complex tables
  - Stamps, logos, symbols
  - Graphs, charts, images, handwriting
- Noisy Documents
  - Noisy, has colors, low quality
  - Inconsistent orientation
  - Data with contextual relationships
- Complex Documents
  - Complex tables
  - Stamps, logos, symbols
  - Graphs, charts, images, handwriting