

AutoDoc

Intelligent Platform for Text Data
Extraction, Analysis, and Generation
(OCR, IDP, LLM)

We design, deploy, and support IT products

AutoDoc Capabilities

Reads, checks, and analyzes any text



Text Recognition



Text analysis, data validation, generating insights



Send Structured Results to Target System



Extract valuable data



Compare Different Document Versions

Implementation Results

Speed

Automatic data extraction and transfer to the accounting system – **5-15 seconds**

Cost

Average processing cost reduction of **80%** through automation.

Optimization

Average payroll cost reduction of **70%**

Accuracy

For high-quality images or digital data: extracted data accuracy of **98-100%.**

AutoDoc Advantages

Intelligent

We use cutting-edge text data extraction technologies:

- Recognition (OCR)
- Analysis and extraction (NLP)
- Analysis and generation (LLM)

Vendor-independent — users can:

- Configure document classification rules (type detection)
- Create new document recognition templates
- Launch training of data extraction models
- Design and customize document processing workflows

Fast & Easily Scalable

- Average page processing speed: 1–2 seconds
- Linear performance growth with horizontal scaling
- Orchestration and on-demand resource provisioning based on Kubernetes, enabled by microservice architecture

AutoDoc Applications

Lending

Tender documentation

Registration certificates, licenses, patents

Accounting

Account Opening

HR records

Document comparison

Contracts, statements, purchase orders

Insurance

Insurance contracts, premium payment receipts

Contract Management

Invoices, statements, contracts, UPDs, promissory notes

Inquiry Processing

Emails, website and messenger inquiries, contact center

Government Agency Document Processing

Court rulings, orders

AutoDoc Integration



**Open API for easy integration
with any system**



**Invoke AutoDoc from an
external system**



**Integration with corporate
Active Directory (AD)**

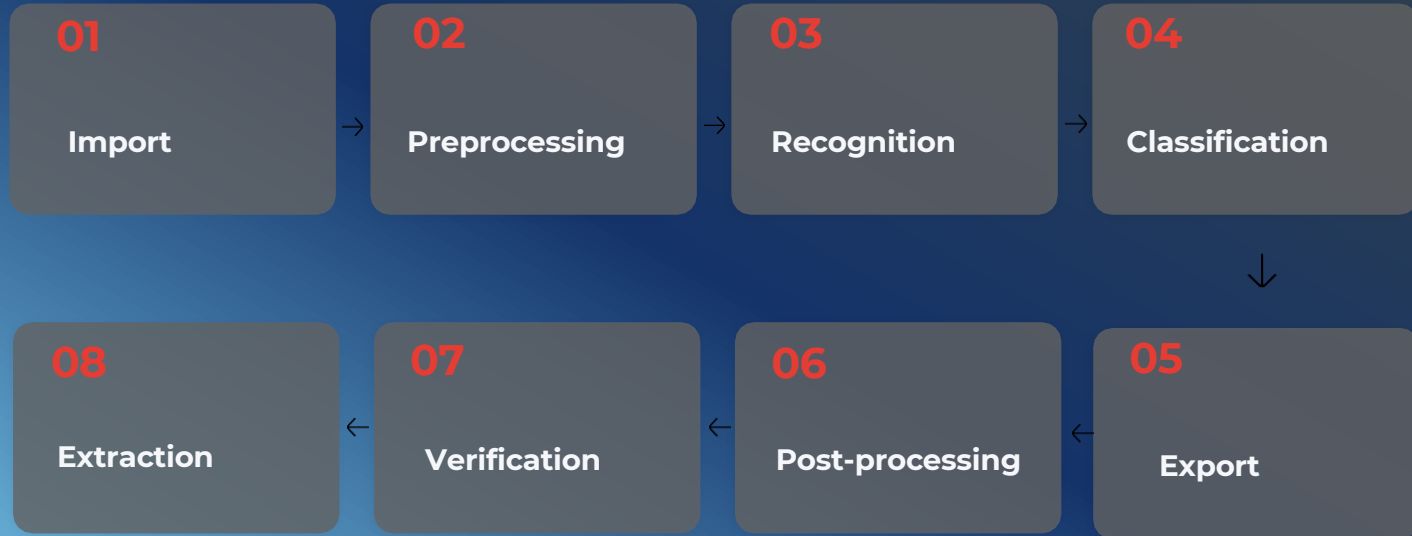


**Ready-made connectors to
ERP, BPM, ECM, EDM
systems, and others**



**Embed AutoDoc into
the interface of an
external system**

Main Document Processing Stages



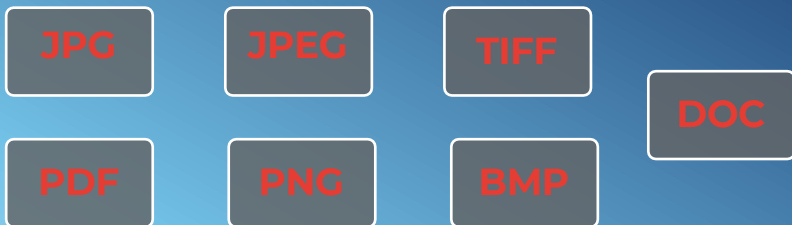
Import

Data Sources:

- Paper (scanner)
- Photos from a mobile phone
- Network folder
- Corporate systems (CRM, ECM, EDM, ERP)
- Email

AutoDoc captures files from any source or accepts a file sent from any system.

Supported Formats



Preprocessing

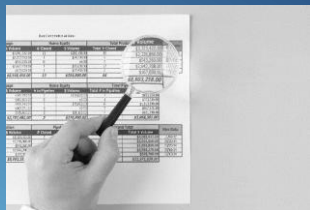
- **Image rotation.** 90° clockwise/counterclockwise or 180°
- **Deskewing.** The system automatically rotates the resulting image to the correct angle relative to the horizontal (0 degrees)
- **Noise removal,** Eliminates artifacts caused by paper quality, scanning, or other factors
- **Image enhancement.** The system automatically crops document margins, leaving only the relevant data in the document image

Recognition

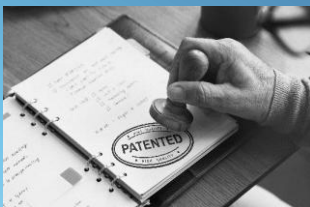
Recognition of all characters on the image:



Text



Complex tables



Seals, stamps, signatures



Handwritten Text



Graphic Objects in Drawings



GOST-compliant fonts

Classification

System Capabilities:

- 01 Process a batch of documents within a single file
– automatic separation into individual documents
- 02 Verify document integrity
- 03 Determine document type (class) based on image and text content
- 04 Verify that all required documents are present in the batch

Data Extraction

Based on cutting-edge technologies:

OCR

Structured documents:

- rigid structure where each attribute is always in its place

NLP

Unstructured documents:

attributes may be located:

- in different places
- in different formats
- with errors

LLM

Free-form documents:

- attributes are not explicitly stated in the text; their meaning must be inferred from context and generated

Post-processing

Validation Checks:



Cross-checks – verify common attribute values across all documents in a batch



Auto-checks



Reconciliation with external sources (ERP, ECM, BPM, CRM, and others)



Calculations (e.g., totals in balance sheets)

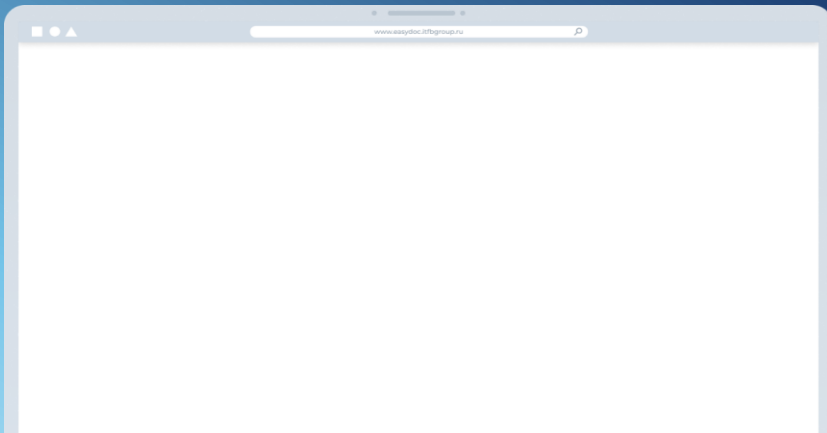


Template compliance checks – verify that specific parts of the text match a template (e.g., key clauses in contracts)

Verification

User-Friendly Verification Workstation:

- The scan highlights where an attribute was recognized from
- Image scale is adjustable for user convenience
- Auto-check errors are highlighted
- Unrecognized attributes can be quickly re-recognized by selecting an area with the mouse
- Document composition can be changed by dragging pages
- Ability to add documents or pages
- Verification statistics are available



Export

What Data Can Be Exported:



Structured recognized / extracted data



System confidence percentage for recognized data



Data coordinates



Images



Document version comparison results

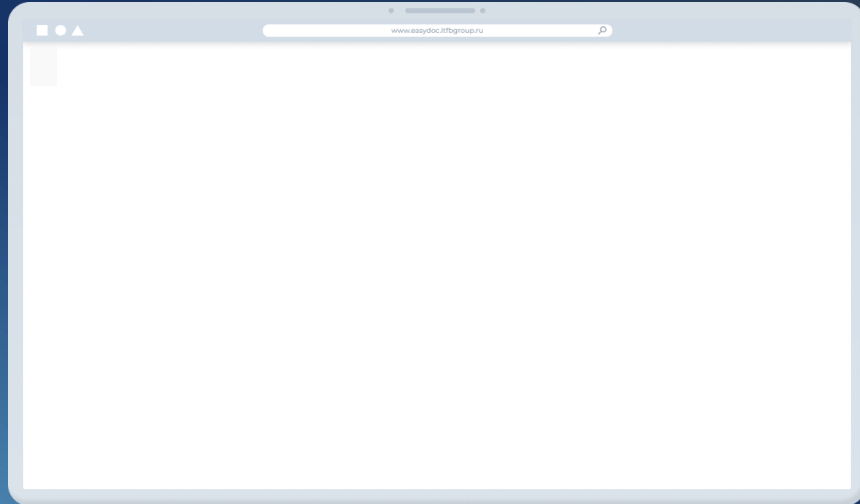
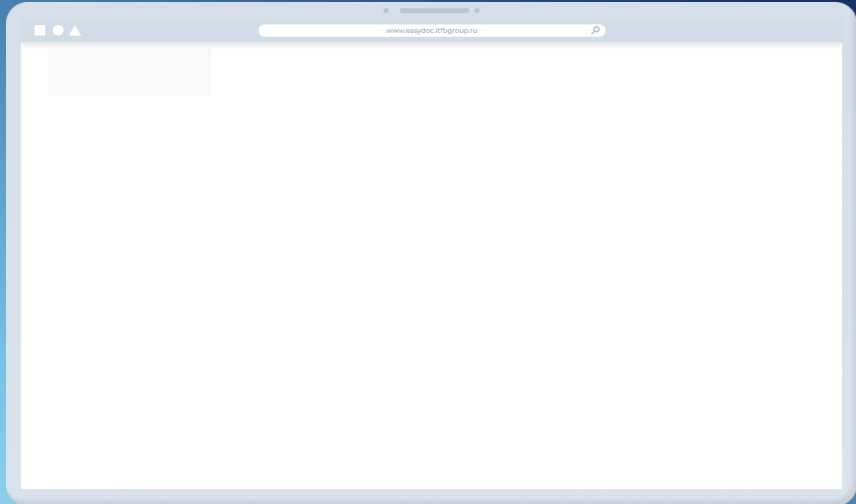
Document Version Comparison

- Compare two versions in any format: PDF, PDF/A, DOCX
- Highlight removed, added, and changed blocks
- Preserve original formatting
- Compare text and tables
- Detect graphic artifacts (signatures, stamps, images)
- Export results as a report

Administration

System Administrator Dashboard:

- Display statistics for core system components on the main page
- Quick drill-down from any segment or widget to detailed data

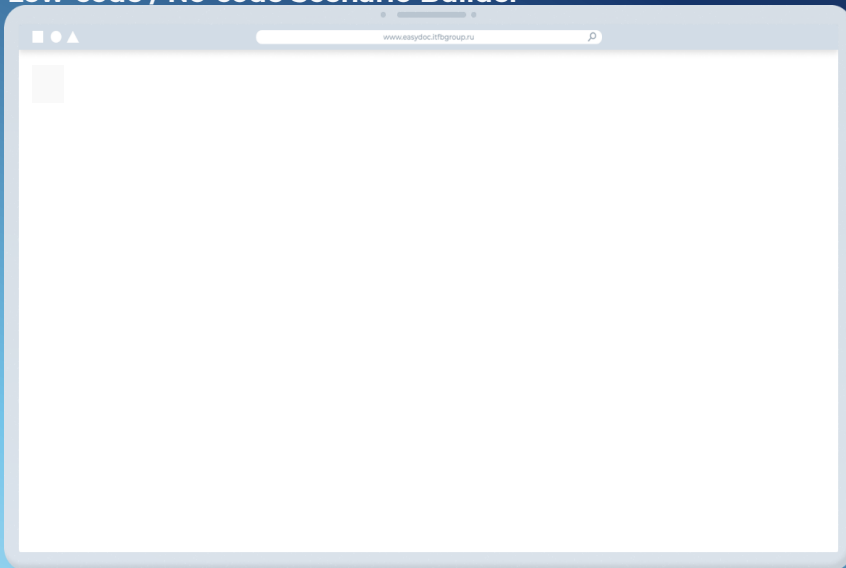


Workflow Designer

Manual Configuration Capabilities:

- Create document processing workflows
- Configure parameters for each stage
- Configure batch processing and validation
- Add classification models
- Add recognition and data extraction models for specific document types

Low-code / No-code Scenario Builder



Capture Studio

Create and Debug Classification and Data Extraction Models

To create a classification model, upload an archive with document samples – the system will train automatically.

Data Extraction Model (from 10 samples):

- Mark up the image to identify target attributes
- Regular expressions – rules for data lookup

