# IBM Content Intelligence

Automate data digitization, extraction and language understanding with AI

## AI and automation redefine how work gets done

Most data, such as emails, text documents, web documents, voice recordings and video, is considered unstructured. Yet this content has much more structure than many would think. Metadata, page layouts, language semantics, information architecture and even grammar are structured. There's a wealth of data to mine for insights and it's often untapped given that extracting information and insights from unstructured data is a challenge. Businesses can now take advantage of complex and unstructured data with an approach that combines data-driven machine learning (ML) methods with knowledge-driven AI algorithms.

While data and document processing tools and approaches for unstructured data have recently received attention, there have been challenges to adoption and implementation. These challenges include:

– Robust digital capabilities are required to successfully execute digital strategy and scale AI and other exponential technologies.

– Issues with process simplification, including the inability to radically redesign inefficient processes and workflows due to manual handling of complex documents.

– Knowledge and skills gap in digital technologies deters the ability to capture institutional knowledge from process subject matter experts (SMEs) into a digital workforce.

– Risks of regulation and compliance violations due to manual information oversight issues in embedding expert knowledge into AI and ML procedures that use digital knowledge assets, such as domain-specific ontologies and knowledge graphs.

– Challenges in implementing AI hybrid platforms at scale that can exploit and integrate both data and knowledge assets.

## Unlock the value of unstructured data with AI

AI revolutionizes unstructured data processing and allows enterprises to make faster and informed business decisions, helps ensure an enhanced customer experience and achieves data accuracy and compliance.

According to the recent IBM Global AI Adoption Index 2021 survey,[1] businesses are rapidly intensifying and expanding the use of AI-based automation technologies. About 25% of surveyed companies are allocating AI investments toward automation of processes. Key reasons cited by companies to use automation tools are:

– Driving greater efficiencies
– Saving costs
– Giving back valuable time to employees[1]

**The IBM® Content Intelligence platform manages end-to-end complex document processing at scale. It allows for digitization, understanding and reasoning, and positions itself naturally at the core of intelligent workflows.** The platform enables and accelerates intelligent automation and captures efficiency gains to support SMEs and human agents in their daily knowledge work. IBM Content Intelligence offers an efficient way to work, with quick and useful insights, and helps enterprises provide enriched experiences to customers, partners and employees.

## What is IBM Content Intelligence?

IBM Content Intelligence is an enterprise-grade and scalable platform that eliminates the manual effort required to extract and apply context-driven data within your business functions. The IBM Content Intelligence platform establishes new intelligent workflows that enable humans to eliminate manual tasks and drive efficiency gains.

The platform allows the extraction of insights at scale from both internal and external unstructured data, including capabilities such as natural language processing, retrieving and monitoring information, and extracting automatically relevant metadata from documents. It also provides intelligent, end-to-end processing of large-scale, document-driven processes on an integrated platform—including robotic process automation (RPA), process orchestration, advanced document understanding and cognitive processing.

This cloud-based, containerized platform scales horizontally and vertically to accommodate document volumes, address difficult document processing, and meet technical and business objectives. It's designed to integrate with existing business systems and corporate processes.

IBM Content Intelligence provides value to customers with four major capabilities:

– **Digitization** combines advanced AI and ML methods with state-of-the-art optical character recognition engines to create a fully digitized copy of the original document. The platform can ingest most formats and digitize content in basically any form, including textual, tabular and geometrical. All relevant metadata is extracted from the documents with connectors to store, index or use downstream.

– **The language understanding toolkit** uses a large knowledge graph, plus a catalogue of industry-specific taxonomies that combine with modern statistical and ML-based natural language processing (NLP) techniques. It can extract business entities and their relationships and classify full documents and document pages for further downstream processing, extract semantic constructs, and provide advanced semantic search capabilities at scale.

– **The reasoning pillar** allows us to use the structure overlayed by the digitization and understanding pillars. This process simplifies the application of reasoning methods, from simple descriptive statistics to more complex predictive analytics models and graph-based inference engines. Models in the reasoning pillar support SME decision-making and pair with RPA engines to increase the process automation rate.

– **Operationalization** refers to the usage of engineering best practices in the designing and building of the platform. The IBM Content Intelligence platform is natively built on modern container technologies, which makes it portable across cloud and on-premises environments. It's designed with a microservices-based architecture to be modular and enable best-of-breed solutions. It's deeply integrated with modern open-source technologies.

Using precise AI and ML eliminates the manual effort required to extract and apply context-driven data within your business functions. IBM Content Intelligence helps clients radically transform manual, document-heavy processes. Our clients have significantly reduced the number of process steps, lowered costs by double digits and improved accuracy.

The most important aspect of our approach is the people and processes we wrap around the technology solution. Success is achieved through methodical design, testing, iteration and optimization of platform components, as well as a process to evaluate, correct and train AI components to generate continuous improvement in automation, speed and accuracy.

## Automation architecture

This holistic automation platform meets a wide range of technical objectives. The cloud-agnostic architecture allows enterprises to scale the IBM Content Intelligence platform to meet daily and long-term performance goals. In each implementation, IBM configures a minimum viable product (MVP) solution and uses the processing outcomes to inform modification or adjustment to components and processes to optimize performance and accuracy. The overall architecture creates composite and interconnected capabilities that adapt to meet the requirements of a large-scale and complex document processing solution.
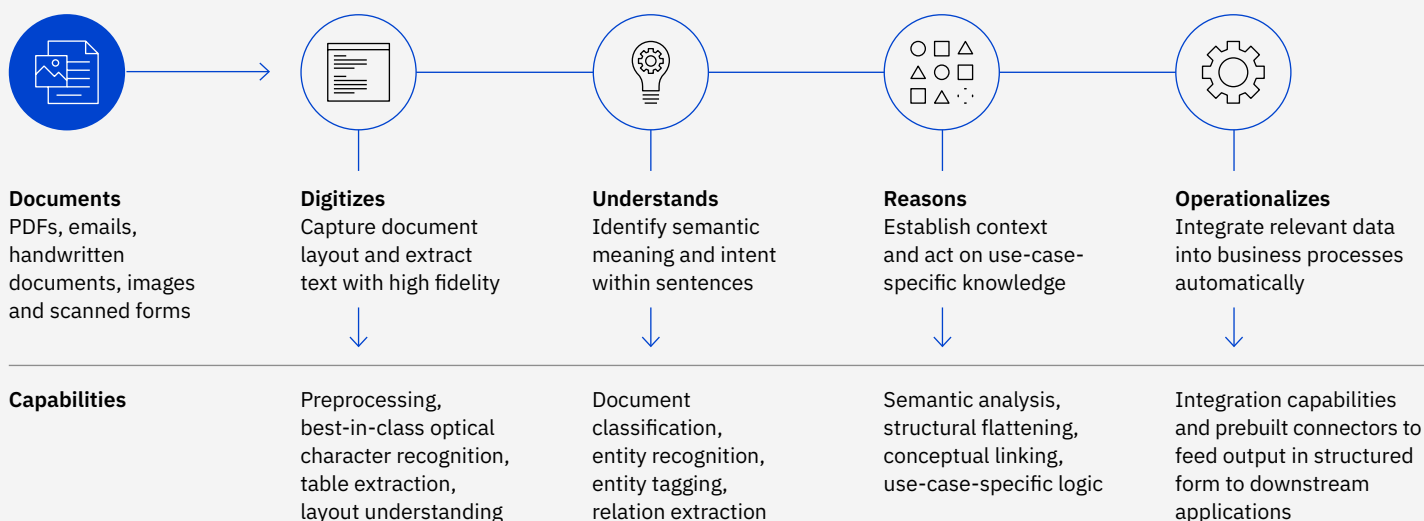
## Integrated platform architecture

Speed, accuracy and cost-effectiveness are top priorities for organizations. The IBM Content Intelligence platform is designed to quickly scan, classify, recognize, validate and export data while maintaining data integrity. The platform accommodates additional components to meet unique needs.
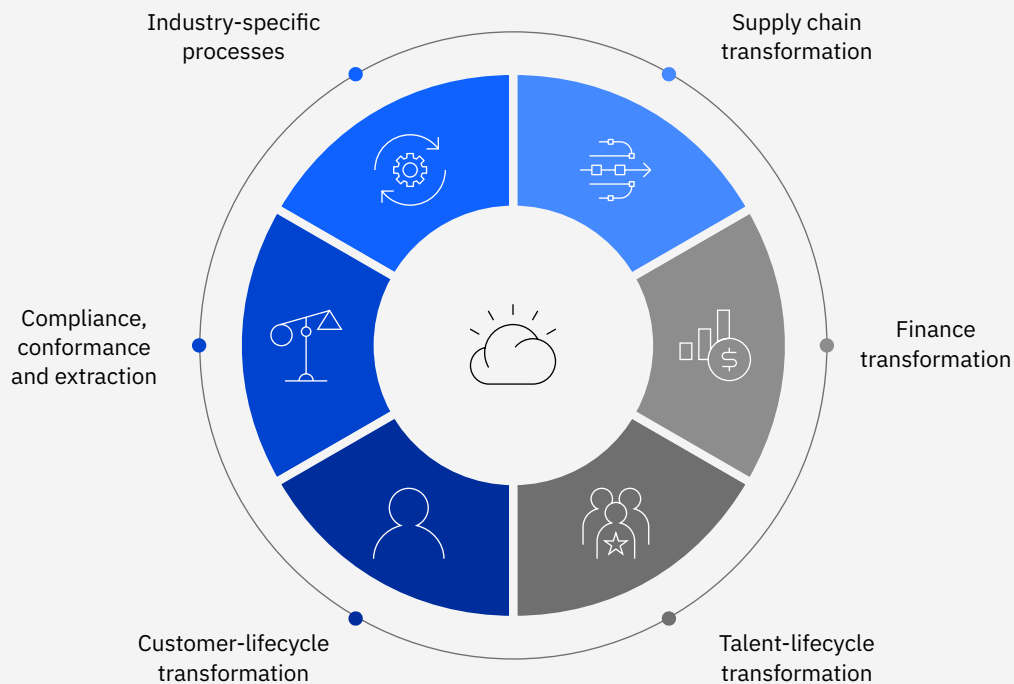
**IBM Content Intelligence key uses**

The integrated platform can:

– Perform image corrections and enhance document quality for better results.

– Set up and configure pipeline stages for processing documents in different supported domains.

– Automatically recognize and separate various supported document types within a domain.

– Extract key and extended metadata from documents based on business requirements.

– Annotate, extract and transform based on business rules.

– Use ML to extract specific keywords, phrases and concepts.

– Clean up and validate extracted data against reference values and business rules.

– Index the totality of original data, all enrichments and extracted metadata to make insights and documents available through a modern semantic search platform.

– Display extracted key values in the user interface for review and correction, which can be exported in different formats, such JSON and CSV.

## IBM Content Intelligence



| Documents | Digitizes | Understands | Reasons | Operationalizes |
|---|---|---|---|---|
| PDFs, emails, handwritten documents, images and scanned forms | Capture document layout and extract text with high fidelity | Identify semantic meaning and intent within sentences | Establish context and act on use-case-specific knowledge | Integrate relevant data into business processes automatically |

| Capabilities | | | | |
|---|---|---|---|---|
| | Preprocessing, best-in-class optical character recognition, table extraction, layout understanding | Document classification, entity recognition, entity tagging, relation extraction | Semantic analysis, structural flattening, conceptual linking, use-case-specific logic | Integration capabilities and prebuilt connectors to feed output in structured form to downstream applications |

# IBM Content Intelligence drives benefits across all business functions

Industry-specific processes

Supply chain transformation

Compliance, conformance and extraction

Finance transformation

Customer-lifecycle transformation

Talent-lifecycle transformation

## Supply chain transformation

– Vendor management
– Procurement contracts review for conformance with interval standards
– Evaluation of contract risk

## Finance transformation

– Procure to pay
  • Cognitive invoice digitization and data extraction
  • AI-enabled travel and expense (T&E) receipt validation and touchless invoice processing
– Lead to cash
  • AI-enabled order data validation
– Record to analyze
  • Risk-based reconciliations

## Talent-lifecycle transformation

– Talent acquisition
– Skill identification
– Employee reskilling and upskilling
– Attrition and churn prevention

## Customer-lifecycle transformation

– Account opening
– Customer onboarding
– Order management
– Know your customer
– Voice of the customer analytics

## Compliance, conformance and extraction

– Fulfill contractual obligations
– Compliance with regulatory requirements (GDPR, LIBOR, CCPA)
– Conformance with internal best practices, including royalty, procurement, leases, insurance and loans

## Industry-specific processes

– Claims digitization and processing
– Payment processing
– Mortgage process
– Lending process
– Trade finance
– LIBOR
– Defense Intelligence (threat and crisis prediction)
– Reinsurance life and health (L&H) underwriting modernization
– Cognitive manuals for manufacturing

# Use cases

## 01

**Smart-data room for tier-1 professional services firm**

Problem: The client needed to improve the effectiveness and efficiency of its projects through automation and AI. The existing process required users to download and review thousands of files in a nested folder structure with limited ability to search and an ad hoc approach to collaboration within engagement teams.

Solution: Upload and ingest documents overnight to IBM Cloud® with text extracted and enriched through ML models. Teams now use an intuitive application to explore documents through business concepts and can rapidly identify key issues and derive high-value insights.

Business benefit: The client reduced the time required to process and review documents, which allowed experienced professionals to focus more time on producing higher-quality insights. The smart data room enabled a first-pass analysis in far fewer days than the previous approach.

## 02

**Efficient medical claims handling for a global insurer**

Problem: The client's claims department for physician and hospital liability insurance handles thousands of complex claims each year, which require in-depth medical expertise. Most of the previous claims aren't fully digitized and not searchable.

Solution: IBM deployed a content intelligence processing pipeline to digitize and classify incoming medical claim documents and extract the most relevant concepts for claim triage, a case summarizer and a document smart reader. Legal claim experts and internal medical support use the platform to assess medical claims, search semantically through the entire claim archive and quickly navigate to the relevant content.

Business benefit: The insurer reduced the average cost for processing a claim by 15%. IBM delivered the first MVP pipeline, case summarizer and document reader within weeks and extended MVP using IBM Garage™ methodology with releases every 12 weeks.

## 03

**Claims processing for a public-sector company**

Problem: This client receives millions of mail packets each year. Mail processing times were more than 15 days, and mandatory overtime was required by claims' assistant staff. With a higher volume of claims expected, this client wanted to automate mail intake with AI and automation.

Solution: Since December 2019, IBM helped the client's intake conversion and communication services improve customer service by building an intelligent automation platform that allows associates to login in a secured environment, interact with systems and process claims.

Business benefit: The agency deployed and scaled an end-to-end intelligent automation platform that's available 24 hours a day, 7 days a week. The platform automated processes capable of ingesting unique form types and a portion of all mail intake. Claims intake and processing time was significantly reduced. The majority of all mail intake was automated within 6 months and the time savings freed up claims assistants to focus on higher-value tasks.

## Why IBM?

IBM helps transform businesses by integrating business platforms and intelligent workflows in hybrid multicloud environments. Our approach adds intelligence to your document digitization processes and helps identify additional opportunities to optimize your business processes.

IBM Content Intelligence uses state-of-the-art AI, including proprietary IBM Watson technology, open-source packages and solutions from a growing alliance and ecosystem partners including hyperscalers, such as Microsoft Azure and Amazon Web Services (AWS), and independent software vendors (ISVs), such as Planet AI, Expert.ai, Kira and Hyperscience.

To learn more about the IBM Content Intelligence platform, speak to an IBM Content Intelligence expert.

**Learn more →**

1 Global AI Adoption Index 2021, commissioned by IBM in partnership with Morning Consult, 2021