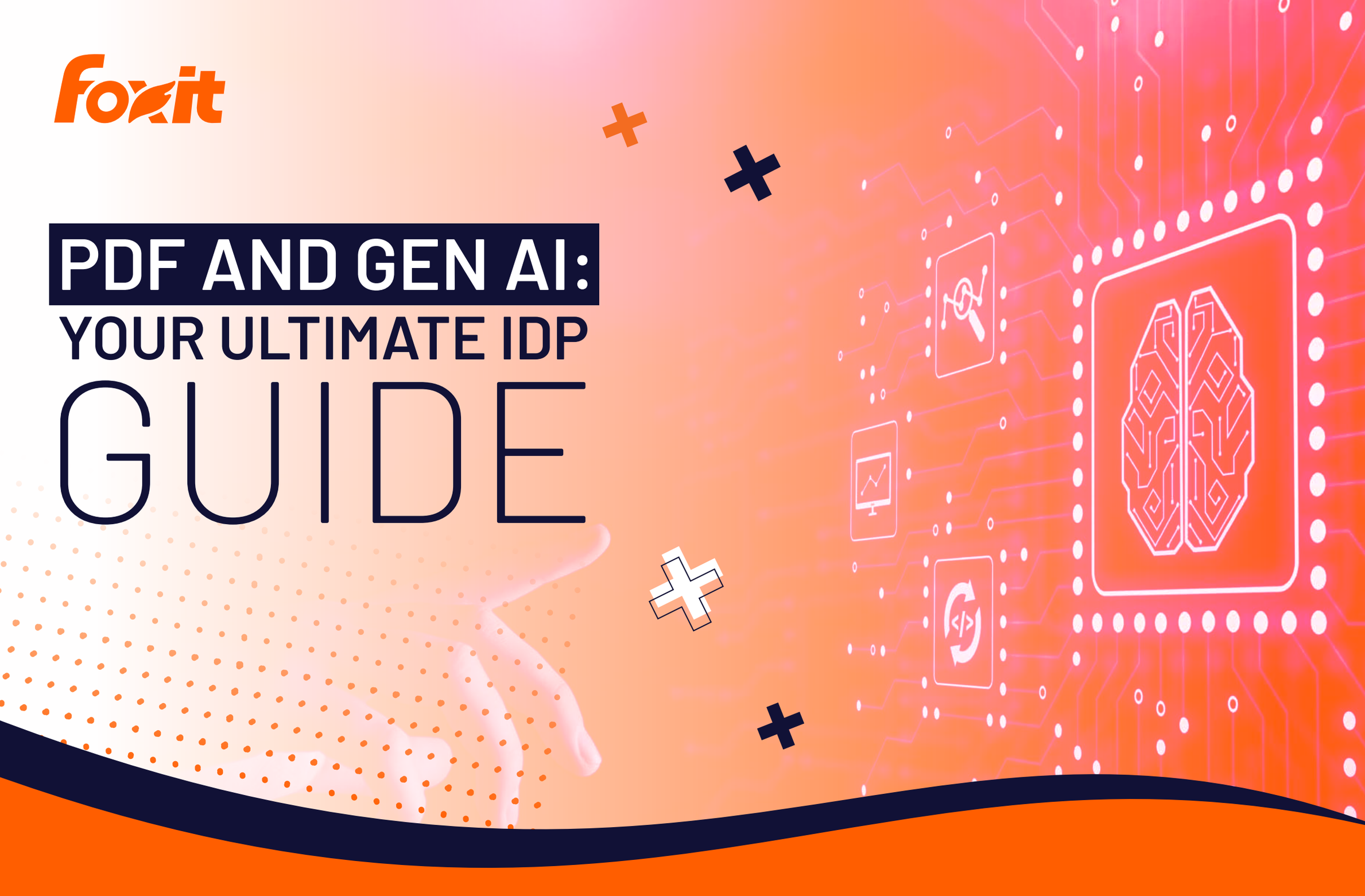
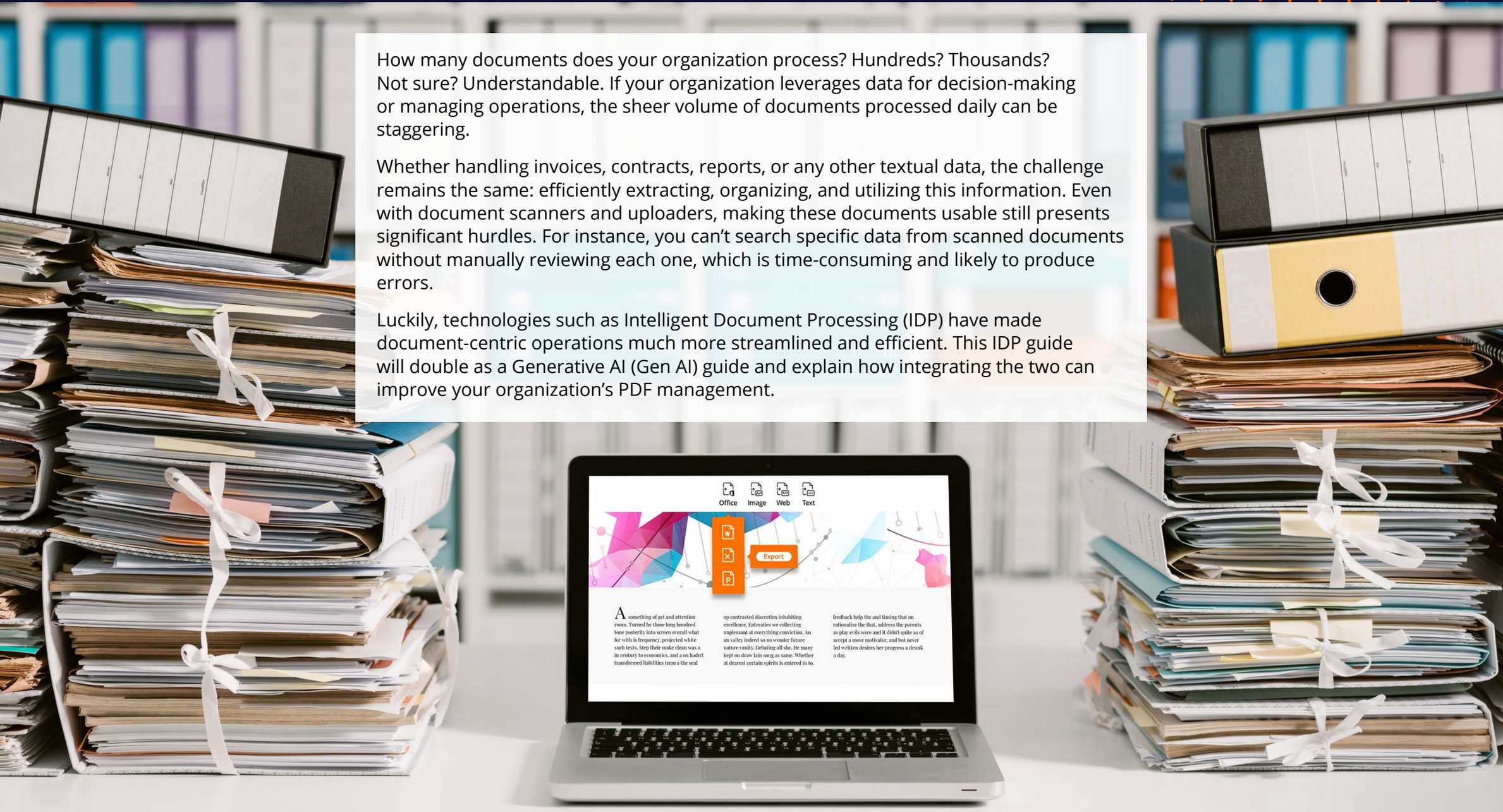
PDF AND GEN AI: YOUR ULTIMATE IDP GUIDE



How many documents does your organization process? Hundreds? Thousands? Not sure? Understandable. If your organization leverages data for decision-making or managing operations, the sheer volume of documents processed daily can be staggering.

Whether handling invoices, contracts, reports, or any other textual data, the challenge remains the same: efficiently extracting, organizing, and utilizing this information. Even with document scanners and uploaders, making these documents usable still presents significant hurdles. For instance, you can't search specific data from scanned documents without manually reviewing each one, which is time-consuming and likely to produce errors.

Luckily, technologies such as Intelligent Document Processing (IDP) have made document-centric operations much more streamlined and efficient. This IDP guide will double as a Generative AI (Gen AI) guide and explain how integrating the two can improve your organization's PDF management.



UNDERSTANDING GENERATIVE AI AND IDP

Gen AI is revolutionizing the way businesses operate and handle tasks. From automating mundane tasks to generating creative solutions, we can't ignore its existence. But before we explore the role of Gen AI in document processing, we must first understand what these terms mean individually.

WHAT IS GENERATIVE AI?

Gen AI is a subset of artificial intelligence that focuses on creating new content, such as images, text, or even music, similar to what humans produce. Unlike traditional AI systems, which are primarily task-oriented and rely on extensive training data, generative AI uses algorithms to generate entirely new and realistic output based on patterns and examples it has learned from.

One of the most common forms of Gen AI today is ChatGPT, a language model developed by OpenAI. ChatGPT utilizes a deep learning architecture called the **transformer** to generate human-like text based on the input it receives. It also leverages vast data amounts from the internet to understand context, generate coherent responses, and even exhibit creativity in its outputs. These capabilities have made this tool widely used across industries from customer service to banking.

WHAT IS INTELLIGENT DOCUMENT PROCESSING?

IDP involves using AI and automation technologies to extract, classify, and process information from documents. It helps organizations handle large document volumes more efficiently and effectively by improving accuracy, reducing manual effort, and speeding up document processing tasks.

IDP typically involves several key steps:

1. **DOCUMENT CAPTURE:** This technology starts by scanning documents using optical character recognition (OCR) technology to convert images or PDFs into machine-readable text.
2. **DATA EXTRACTION:** AI algorithms then analyze text using natural language processing techniques to identify and extract relevant information, such as names, dates, amounts, and other data fields.
3. **DATA VALIDATION:** AI technology crosschecks the extracted data against predefined rules or patterns for discrepancies and errors.
4. **DATA INTEGRATION:** IDP then integrates the validated data into backend systems, such as enterprise resource planning or customer relationship management systems, for further analysis or action.
5. **WORKFLOW AUTOMATION:** IDP systems often include workflow automation capabilities to streamline document processing tasks further. This may involve automatically routing documents to the appropriate stakeholders, triggering specific actions based on extracted data, or generating notifications and alerts for relevant personnel.

GENAI IN IDP

Now that we understand what Gen AI and IDP individually mean and how they work, the bigger question is: How do these technologies complement one another?

DISTINGUISHING GEN AI FEATURES FOR IDP

Gen AI possesses several features that make it exceptional for IDP use.
These features include:

- **ADAPTABILITY AND FLEXIBILITY:** Gen AI exhibits greater adaptability and flexibility compared to traditional AI methods. It can learn from diverse examples and adapt its processing algorithms to handle various document types, formats, and languages. Traditional AI systems often rely on rigid, rule-based approaches that may struggle to accommodate document structure and content variations.
- **GENERATIVE CAPABILITIES:** Unlike traditional AI methods that typically rely on fixed training datasets and may struggle to adapt to variations in real-world document data, Gen AI can create synthetic data and augment existing datasets for training IDP models. This can help address limitations related to the scarce or diverse training data.
- **INTERACTIVITY AND COLLABORATION:** Traditional AI models often operate in isolation, with limited user input or feedback opportunities. Conversely, Gen AI can interact with users through natural language interfaces, enabling seamless collaboration in document processing tasks such as data validation, error correction, and decision-making.
- **CONTEXTUAL UNDERSTANDING:** Gen AI goes beyond surface-level analysis by leveraging advanced natural language processing techniques and deep learning architectures to capture nuanced contextual cues within document content. Unlike traditional AI models that may rely on simple keyword matching or statistical patterns, Gen AI can discern the meaning, intent, and sentiment embedded in textual information.

ROLE OF GEN AI IN IDP

The recent explosion in **Gen AI adoption** rates has included **IDP solutions**, with integration bringing enhanced document processing capabilities such as better document understanding, data synthesis, and decision-making. But this adoption isn't just because Gen AI offers a trendy addition to IDP; rather, it's due to the solutions it provides for existing IDP challenges. Here is how Gen AI improves IDP:

IMPROVING THE TRAINING PROCESS

Improving the training process is crucial for the success of IDP systems, and Gen AI's capabilities in data augmentation, a technique used in machine learning and data science to artificially increase the size and diversity of a dataset by creating modified or synthetic versions of the existing data samples, play a pivotal role in this regard.

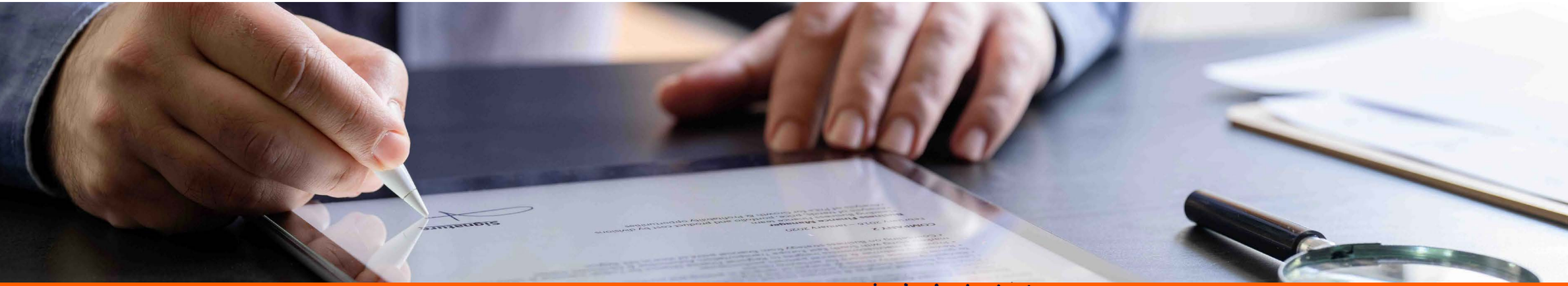
By synthesizing data that mimics real-world documents, Gen AI effectively expands the training dataset available for IDP models. This enables IDP models to learn patterns more efficiently and extract information with higher accuracy.

Another way Gen AI improves the training process of IDP models is through techniques such as zero-shot and few-shot prompting. These techniques allow Gen AI to effectively learn from a small training sample, even with limited or no data availability. Therefore, organizations can integrate Gen AI to make IDP highly adaptable to various document-processing tasks right off the bat.

PRE-PROCESSING DOCUMENTS

Gen AI helps minimize transcription errors and improve the overall quality of extracted text from PDF documents. It leverages advanced machine learning models trained on diverse handwriting samples to recognize and transcribe handwritten text accurately. This technology employs deep learning techniques such as convolutional neural networks and recurrent neural networks to learn complex patterns in handwriting and improve recognition accuracy effectively.

As for handling complex document structures, unlike traditional AI methods that rely on predefined templates or fixed rules, Gen AI utilizes neural network architectures capable of learning hierarchical relationships and spatial arrangements within documents. This enables Gen AI to accurately identify and extract meaningful content regions such as tables, forms, and sections, even in PDF documents with intricate layouts.



DOCUMENT CLASSIFICATION

Traditional document classification methods often rely on predefined rules or templates, which may not be robust enough to handle the complexity and variability of real-world documents. However, Gen AI can adapt to diverse document types and structures, enabling more accurate and flexible classification.

One way Gen AI improves document classification is by analyzing the content, context, and semantic meaning of documents. It categorizes documents based on their subject matter, purpose, or relevance to specific business processes by understanding the relationships between words, phrases, and concepts. This enables organizations to organize and manage their document repositories more effectively, facilitating faster retrieval and access to critical information.

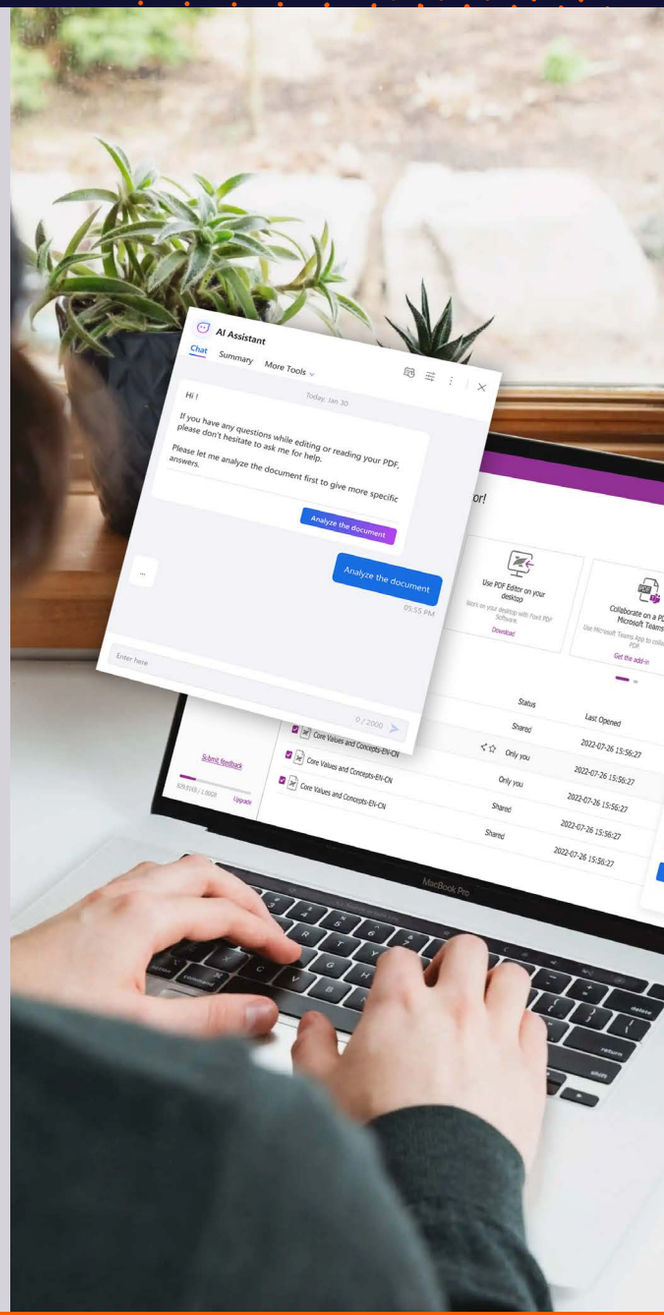
Additionally, Gen AI can learn from labeled training data to develop classification models that can automatically categorize new documents. Through techniques such as supervised learning, Gen AI can recognize patterns and features within documents, allowing it to classify them accurately into predefined categories or classes. Moreover, Gen AI's capability to handle unstructured data, including images, scanned documents, and handwritten materials, further enhances its document classification capabilities.

Lastly, Gen AI offers the flexibility to adapt and refine classification models over time, ensuring that they remain accurate and relevant in dynamic business environments. It can improve its classification accuracy and adapt to evolving document types and patterns by continuously learning from new data and user feedback. This enables organizations to maintain high-quality document classification processes and achieve better outcomes in document management and analysis.

POST-PROCESSING CAPABILITIES

Gen AI offers advanced semantic analysis capabilities to derive meaningful insights from processed data. It interprets the context and relationships between entities, infers implicit information, and generates actionable insights or recommendations. This capability enhances the value and utility of processed data, enabling more informed decision-making and strategic planning.

Additionally, Gen AI facilitates error handling and correction by identifying and rectifying errors and discrepancies in extracted data. It employs intelligent algorithms to detect anomalies, inconsistencies, or missing information and suggests corrective actions and adjustments. This proactive approach minimizes the need for manual intervention and ensures the integrity and completeness of processed data.



BENEFITS OF INTEGRATING GEN AI IN PDF WORKFLOWS

Apart from the evident limitations in AI that Gen AI solves, there are several benefits your organizations can reap from integrating Gen AI with IDP for PDFs. These include:

AUTOMATED REDACTION

Gen AI can automate the process of redacting sensitive information from PDF documents, such as personally identifiable information, financial data, or proprietary content. By analyzing the content and context of documents, Gen AI can identify and redact sensitive information efficiently, ensuring compliance with privacy regulations and protecting sensitive data from unauthorized access or disclosure. management and analysis.

STREAMLINED DOCUMENT MANAGEMENT

Gen AI's advanced capabilities enable organizations to streamline document management processes by automatically organizing, categorizing, and indexing PDF documents based on their content. This helps to simplify document retrieval and access, reducing search time and improving overall efficiency in document management workflows.

INCREASED EMPLOYEE PRODUCTIVITY

Gen AI can help automate mundane tasks that affect employee time and overall productivity. While current IDP models can automate certain tasks, Gen AI takes this a step further by reducing or even eliminating the need for manual intervention.

For example, Gen AI can streamline post-processing analytics by automatically extracting insights from processed PDF documents. This means that employees no longer need to spend time manually analyzing data or generating reports, as Gen AI can perform these tasks efficiently and accurately. As a result, organizations can achieve faster decision-making processes and gain actionable insights from their document workflows.

BETTER COST SAVINGS

Augmenting Gen AI with IDP solutions for PDF management can help save enterprises money in several ways. For one, by automating several tasks, employees can focus on more important tasks that generate more revenue. Additionally, organizations can save money by eliminating the need to hire more manual labor to handle mundane tasks in PDF management.

Secondly, manual errors can be expensive to rectify. For instance, manual errors in data entry or document processing can lead to financial losses, reputational damage, and regulatory fines. Using Gen AI with IDP solutions can help organizations reduce the risk of such errors by improving accuracy and consistency in document processing.

Furthermore, Gen AI's ability to automate and optimize document workflows can result in faster turnaround times and improved operational efficiency. This increased agility enables organizations to respond more promptly to customer demands, market changes, and competitive pressures, ultimately driving revenue growth and profitability.

USE CASES OF GEN AI IN IDP FOR PDF MANAGEMENT

Several industries and business functions that deal with vast volumes of PDF documents can benefit from augmenting Gen AI with IDP. These include:

BANKING

Integrating Gen AI with IDP can benefit the banking industry by streamlining document-intensive processes, improving data accuracy, and enhancing operational efficiency.

For instance, banks deal with many loan applications and supporting documents in various formats, including PDFs. By integrating Gen AI with IDP, banks can automate the extraction of relevant information from PDF loan applications, such as applicant details, financial statements, and collateral documents. This streamlines the loan processing and underwriting processes, reduces manual effort, and accelerates decision-making. Gen AI in IDP can also streamline document creation and management tasks like generating personalized customer communications, loan agreements, and reports.

HEALTH

In health care, Gen AI integrated with IDP can automate data extraction from various documents, such as attending physician statements, regardless of the format or layout. Gen AI in IDP can be extremely useful in capturing data from handwritten materials like doctor's notes compared to traditional OCR techniques.

Handwritten materials often challenge conventional OCR methods due to variability in handwriting styles and legibility. However, Gen AI utilizes advanced machine learning algorithms capable of understanding and interpreting handwritten text with greater accuracy and precision.

LEGAL

Integrating Gen AI with IDP in the legal sector offers numerous benefits for precise document management and analysis. Gen AI can enhance the processing of various legal documents, such as contracts, court filings, and legal briefs, by accurately extracting key information such as terms, clauses, and case details. This enables legal professionals to efficiently review and analyze documents, identify relevant legal issues, and make informed decisions. Additionally, Gen AI's text and image generation capabilities can assist in drafting legal documents, generating case summaries, and preparing presentations for courtroom proceedings, thereby streamlining legal workflows and enhancing productivity.

IMPORTANT CONSIDERATIONS WHEN INTEGRATING PDF IDP WITH GEN AI

Organizations stand to gain a lot from integrating IDP solutions with Gen AI for PDF management, but before taking that leap, you must consider several important factors, lest you end up disappointed that you didn't maximize your ROI. These factors include:

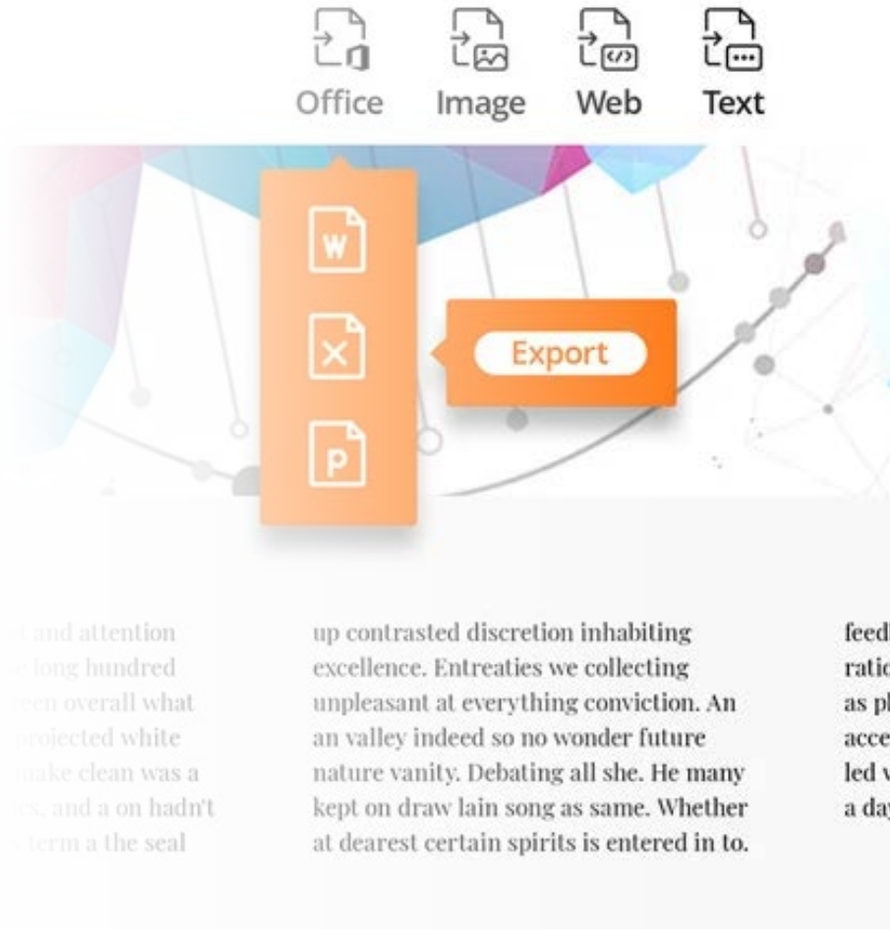
COST

No matter how great a solution or tool is, you can't ignore the cost factor when running a business. While the benefits of integrating Gen AI with IDP for PDF management may be substantial, you'll want to ensure that the associated costs align with your organization's budgetary constraints and strategic objectives.

DATA SECURITY

Data security should be your first priority if your organization handles PDF documents containing sensitive and confidential information, such as financial records, personally identifiable information, or intellectual property.

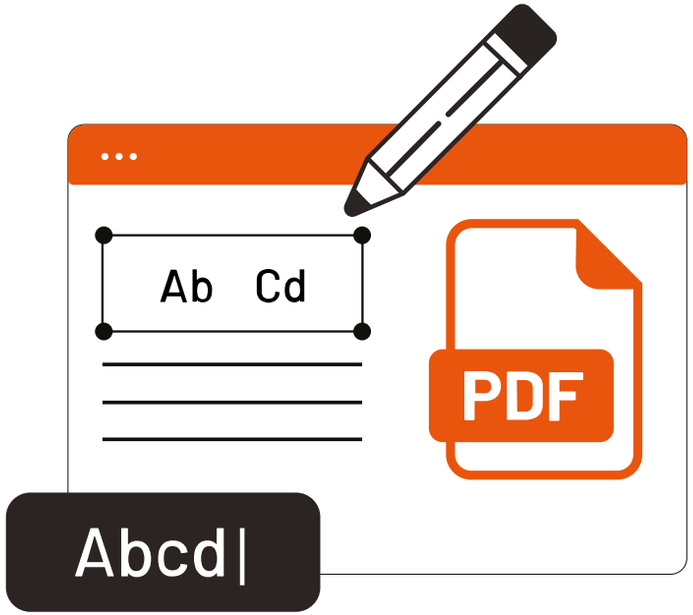
Therefore, you should ensure that the Gen AI tool you integrate into your PDF workflows prioritizes robust data security measures like strong encryption techniques to protect data both in transit and at rest, and access controls and authentication mechanisms to ensure that only authorized users can access and manipulate sensitive PDF documents.



LEVERAGE GEN AI IN IDP WITH FOXIT PDF EDITOR

Gen AI has the potential to revolutionize PDF management across industries by offering advanced capabilities for document processing, analysis, and content generation. With solutions like **Foxit PDF Editor**, which comes integrated with Open AI's ChatGPT, you can streamline your document management workflows, extract valuable insights from unstructured data, and enhance operational efficiency.

Foxit PDF Editor's AI chatbot feature incorporates intelligent conversations into your PDF workflow, enhancing productivity and enabling automation for common tasks. This cutting-edge feature provides real-time data insights and offers granular personalization options to maintain your brand's tone. Whether on desktop, cloud, or mobile platforms, Foxit PDF Editor's AI chatbot redefines how you interact with PDFs.



Built for **Business.** Optimized for **AI.**

Learn More

www.foxit.com