

Capturing data intelligently

**AN EASY WAY OF ENSURING
COMPLIANCE WHEN COMMUNICATING
WITH YOUR CUSTOMER**

White paper

TABLE OF CONTENT

1. THE UNSTRUCTURED DATA AND PAPER CONUNDRUM	3
1.1. Unfortunately, paper often reigns supreme...	3
1.2. Why are companies still using paper?	4
1.3. The other side of the problem: unstructured information	4
2. CONQUERING THE PAPER MOUNTAIN AND DIGITALLY UNSTRUCTURED WITH CAPTURE TECHNOLOGY	5
2.1. What to look for in a capture product	5
3. INTELLIGENTLY CAPTURING YOUR CONTENT	6
3.1. What is intelligent capture?	7
3.2. Why you should capture content intelligently	8
3.3. Compliance without the headaches	10
3.4. Two types of capture	11
3.5. The power of mobile intelligent capture	12
4. INCREASING YOUR INTELLIGENT CAPTURE SOFTWARE'S RELIABILITY	16
4.1. Natural language processing	16
4.2. Liveness detection and face recognition	16
4.3. Pattern recognition through machine learning	17
5. STRUCTURING THE UNSTRUCTURED	18
5.1. Why don't we just structure all incoming information?	18
5.2. Getting the help of customers and partners thanks to upload portals	18
6. ROBOTS VERSUS ALGORITHMS: IS RPA ENOUGH?	20
6.1. Reaching the limits	20
6.2. Machine learning to the rescue	21
7. ONCE YOU GO INTELLIGENT CAPTURE, YOU NEVER GO BACK	22



1. THE UNSTRUCTURED DATA AND PAPER CONUNDRUM

Capturing customer data from incoming mail or messaging channels and for account creation is crucial in many businesses. Yet surprisingly, quite a lot of data is still captured on paper or is being sent by and to companies in an unorganized manner. This makes getting those data into the right systems much more challenging, and often leads to longer times for document processing, maintenance, classification, and information retrieval, as well as higher handling and material costs. In short, your business processes work less than optimally, impacting the speed at which you can help customers and your bottom line.



73%
of owners
and decision
makers



print out information at least

4X/DAY

1.1. Unfortunately, paper often reigns supreme...

Gartner Research VP Ken Weilerstein said employees print **about 400 pages per month** on average. Bringing that number down could save an organization quite a bit, even when we just consider paper and toner costs. Wakefield Research discovered that 73 percent of owners and decision makers at companies print out information at least four times each day. **InfoTrends** and **Wakefield** looked further into the costs, stating that companies might spend close to 25,000 euros each year on document management systems and maintenance to manage an average of 5,000 pages per month.

1.2. Why are companies still using paper?

Even with advances in intelligent document management platforms, it seems hard to ban paper from the office. This begs the question as to why companies are still clinging to paper despite its use being slow and costly compared to digital tools. We deciphered a few reasons:

- 1. Employees are used to a paper workflow** and are not always eager to make the switch.
- 2. Paper is a tangible object which gives a sense of security** compared to something more abstract than a hard drive or even the cloud.
- 3. In rare cases, regulations prohibit the digital storage of certain documents, but concerns about not being able to comply with the law are a much bigger inhibitor.**
- 4. There's a lack of understanding** about the possibilities of digital technologies. Many businesses have no clear understanding of the potential that intelligent capture technology holds for their company.



1.3. The other side of the problem: unstructured information

It's not just paperwork that hinders capturing customer data. Even in companies that have drastically diminished their paper usage and made the digital switch, obstacles still appear. Often, those obstacles take the form of unstructured content. Client requests, accident forms and insurance claims, support tickets, and more may enter your organization digitally via mail or other channels such as WhatsApp. However, information often needs to be read, sorted and registered manually in the right systems by an employee. Once again, this slows down the handling time of requests and the general speed of your processes, especially at peak times. Moreover, incoming unstructured data requires your workforce to dedicate their time to repetitive administrative work, when they could be advising and helping clients.

2. CONQUERING THE PAPER MOUNTAIN AND DIGITALLY UNSTRUCTURED WITH CAPTURE TECHNOLOGY

The first step in most business processes is to gather customer data from different inbound channels and through mediums such as paper and email. It is then sent and organized in the designated systems. So the best way to reduce paper use and add structure to your digital information is by tackling this entry point. In almost all cases, software to capture content digitally can be instrumental in achieving this.

2.1. What to look for in a capture product

There are many types of content – think customer emails, account creation forms, invoices, insurance claims, and more – flowing into your organization from a plethora of digital and paper channels. It's essential that you find capture software that can handle all of them. In general, you'll need to capture structured and unstructured content. Structured content can be invoices that use the same template every time or website forms with clearly labeled fields. Unstructured content may consist of emails that don't follow a set format, as well as images and video that are integral to insurance claims or provide customer details such as IDs.

Capture structured and unstructured documents

Generally speaking, information is extracted from documents with a recognizable layout or template, meaning there will only be a slight variation in the way files are structured. Think questionnaires or invoices. Your capture

products can read and interpret the information on the document, usually on the basis of manually coded rules.

For unstructured content, many programs utilize keyword recognition to match documents to the available document classes. But determining the keywords manually and writing rules to account for all the variables in a layout requires considerable effort and is time-consuming. To simplify and speed up this work, organizations are adopting more and more intelligent technologies, such as machine learning, to capture unstructured content. Artificial intelligence and machine learning let you capture many different unstructured formats. With pattern recognition – a machine learning technique – you would never have to use a template ever again. You could also extract information from images and even video with intelligent capture. More on that later in this white paper!





3. INTELLIGENTLY CAPTURING YOUR CONTENT

It is clear that capture applications infused with intelligence are becoming the norm when battling the increasing inflow of content. Machine learning algorithms can significantly scale back the number of manually executed tasks, such as determining the myriad types of layout. Soon, we won't just have capture. We'll have intelligent capture.

3.1. What is intelligent capture?

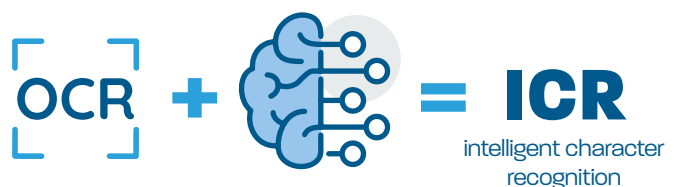
Intelligent capture (IC) is not a singular technology, but rather an approach that combines multiple artificial intelligence (AI) technologies, optical character recognition (OCR) technology for recognizing text on images and data extraction with machine learning (ML) to properly understand and sort that text. This combination of technologies means that intelligent capture can now convert any human-readable content – not only text but also images and video – into digital and actionable information.

The 'capture' part of the term IC might suggest that this is where it ends. However, that would miss the most important aspect of IC: efficiently and effectively extracting information and organizing the extracted data so that it can serve as the basis or trigger for other automated processes in your business applications. This process of organizing the data usually occurs in a few steps. First, OCR software searches for the necessary data on the document. Then classification algorithms – often driven by ML – extract this data to match it to the necessary records, files, and fields and determine the document type. Next, the extracted information is validated through techniques that can range from simple to complex ones. More on that later.

Optical and intelligent character recognition: an integral part of intelligent capture

An intelligent capture product would be incomplete without optical character recognition. OCR is the technique of recognizing and digitizing text in images. It's so essential to IC that accuracy needs to be through the roof.

To increase that accuracy, more AI can be added. For instance, to improve recognition of handwriting or actually interpret the extracted text. As soon as AI is involved, OCR becomes intelligent character recognition (ICR). Docbyte uses different algorithms to bump up accuracy rates to unseen heights and get robust, high-fidelity intelligent character recognition. We also harness other visual algorithms to help us with face or object recognition. In an insurance scenario, for instance, AI can detect whether there's a car, a house or a bike in the picture that you're attaching to your claim.



3.2. Why you should capture content intelligently

AI is a buzzword. It gets slapped on just about everything, with varying results. But when it comes to capturing your information, intelligent technologies really do provide many benefits and better results compared to 'regular' non-AI-driven capture. That's why Docbyte has infused its capture product with machine learning, among other things. Here are some of the most important advantages, most of which result from IC's ability to extract data from unstructured content:

1

IMPROVED CUSTOMER RETENTION

Service level expectations are always increasing. Whether it's filling out a claim request, opening a new account or applying for a loan, people hate waiting and they expect to be helped straightaway. IC enables you to meet those high expectations by creating a far more pleasant and efficient experience for customers interacting with your products and services. For instance, in an intelligent digital mailroom, requests can automatically be forwarded to the right department instead of waiting on manual triage. That way your customer is helped by the right people much quicker.

2

REDUCED ADMINISTRATION

IC software takes care of all data recognition, extraction, classification and transformation, even for unstructured content. That means fewer repetitive administrative tasks for your employees, enabling them to focus on what matters most: your customer.

3

MORE AUTOMATION

IC ensures you obtain high-quality Know Your Customer (KYC) data and adds the necessary document metadata. This provides an excellent basis for further automation of your case management, unlocking more digital transformation possibilities for your organization and reducing the cost of that innovation.

4

REDUCED COSTS

IC's ability to automate a large part of your processes results in significant cost reductions. At the most basic level, costs for paper, toner, and physical storage are minimized. But the impact of accelerating your critical business processes and the resulting increase in efficiency is most obvious in your bottom line. Moreover, employees are no longer bogged down by repetitive administrative tasks. So they have more time to focus on their core jobs, which results in better service, more sales, and an efficient organization.

5

IMPROVED COMPLIANCE

With less room for human error, the quality and accuracy of your KYC or client data will improve. Moreover, the digitalization that IC brings enhances the tracking and security of your documents. Being compliant with all the necessary regulations becomes easier than ever.

6

INCREASED INSIGHT INTO YOUR ORGANIZATION AND CUSTOMERS

IC ensures that you have more digitized and correctly classified information on just about everything that happens in your company. By feeding this prime data to analytics tools, you increase the possibility of gaining new insights that can help optimize how your organization works.



3.3. Compliance without the headaches

In this era of GDPR, companies are required to pay close attention to privacy and other regulations. This is a constant concern, especially in the banking and insurance industry, where a lot of effort is being put into compliance. However, intelligent capture can alleviate or at least ease some of those worries:



- A digital mailroom or onboarding experience with IC requires less human interaction. This reduces the chance of human error, which in turn increases the quality and accuracy of KYC data.



- Intelligently captured documents have the correct metadata and labels automatically attached to them. This facilitates file retrieval and improves file traceability.



- Documents or data ingested through intelligent capture can be time-stamped effortlessly. As a consequence, it is easy to implement automatic rules for file deletion or document expiration date reminders based on legal requirements.

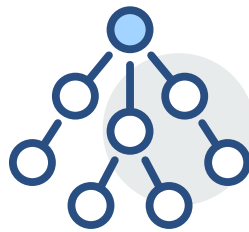
3.4. Two types of capture

There are many different types of capture, but the two most common ones are batch and distributed capture.



Batch capture

Batch capture is all about the streamlined and automated scanning of large volumes of documents. It's usually adopted to capture incoming information from traditional paper sources or to digitize your existing paper archives. It can be used in a variety of scenarios, such as scanning, importing and processing images, automated indexing, field-based data capture, and more.



Distributed capture

In this day and age of globalization, with companies having multiple branch offices and people sending information about accidents, loans, and more on the go, it's impossible to perform all the capturing processes at one location. Distributed capture – with mobile capture being the most obvious kind – makes the core functionalities of your capture software available for various systems and applications. It becomes more straightforward to use predefined process transactions and this method also guarantees compliance requirements. When optimized for smartphones and tablets, distributed capture is ideal for on-the-go teams.

3.5. The power of mobile intelligent capture

Smartphones are omnipresent. They provide powerful computing, a wide array of sensors and a quick connection to the internet. So the devices are highly suitable for bringing intelligent capture solutions to where they need to be, with mobile capture. This type of distributed capture provides all the right ingredients to efficiently and intelligently connect your processes to people and products or services.



Amplifying the benefits of intelligent capture

For Docbyte, mobile capture isn't a nice-to-have, it's a must-have. Its ability to further enhance the benefits that intelligent capture brings leads us to believe that mobile capture is an integral part of any IC solution.

According to an AIIM study, the speed at which mobile IC makes data available and its ability to turn processes paperless is unparalleled. A mobile app digitizes your business process – such as onboarding a new client – from the get-go, so that the client can fill out forms digitally, take a picture of their ID to add details to their file, and more. There's no need for paper forms to be scanned and processed at your HQ or copy contact details from the client's driver's license.

As a consequence, mobile IC reduces the end-to-end processing time even more, creating a faster and better experience for your customers. Moreover, it reduces manual handling and human error, resulting in high-quality data that can be easily tracked. Mobile IC also makes it easier to enforce compliance with regulations such as GDPR, as there's only one way of handling sensitive and personal information. With very high expectations when it comes to service, providing a user-friendly, enjoyable and safe experience is key to retaining customers.



High tech in the palm of your hand

Smartphones are incredibly potent pieces of hardware. To put it into context: Apollo 11's guidance computer put a man on the moon with 100,000 times less processing power than an iPhone. Smartphone cameras are good enough for even professional photographers to use them in shoots. And, over the years, we've packed our devices with as many sensors as possible. Gyroscope, accelerometer, light sensors, you name it. In short, smartphones have the potential to pack quite a punch when it comes to intelligent capture:

1

HIGH-QUALITY IMAGES FOR MORE ACCURATE DATA

Smartphone cameras outperform just about any scanner. This means that mobile device images of insurance or registration forms, IDs, driver's licenses, and more are of an incomparable high quality. As a consequence, techniques such as optical character recognition and other intelligent technologies have much less of a hard time recognizing and extracting text and data from pictures. Besides that, based on the information a gyroscope or light sensor loads into the IC application, for instance, clients receive clear instructions on how to take a pixel-perfect photograph. Ultimately, this will all have a positive impact on the accuracy of extracted data, getting even closer to 100 percent for recognition and extraction.

2

PROCESSING DATA ON THE EDGE

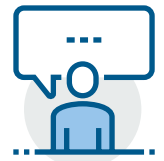
Your phone has orders of magnitude more computing power than what it took to land on the moon. So it's no stretch to think that data extraction and data transformation lie within the capabilities of these mobile devices. By bringing part of IC to the smartphone, it's possible to provide real-time feedback on the extracted data, enabling clients to validate or correct information on the spot. Moreover, with on-device processing you don't need an active internet connection.

On-the-spot onboarding with intelligent mobile capture

We'd like to introduce you to Jack. He's in the market for a new TV. His local electronics store has a wide range of available options, but unfortunately, they're rather expensive. A consumer loan with a monthly fee would help Jack get the TV of his dreams. The store uses an IC product to process and send those loan requests to a partnering bank. Jack goes straight to the sales clerk in the store. The clerk scans Jack's ID with his phone camera fitted with a designated mobile capture application. Now, AI-driven magic happens. Optical Character Recognition (OCR) reads the text on the image and a classification algorithm attaches the right labels to the information. This data is then sent to the bank or store's systems and added to Jack's file or a new file is created. To finalize the deal, a digital contract is generated based on that information, which Jack can easily sign digitally.

By intelligently capturing Jack's information, the first, tedious steps of a loan request happened in just a matter of minutes – seconds even. Risk assessment and approval by the bank now take place much more smoothly. Moreover, it's much easier for the bank to assess Jack's financial situation as the application offers the bank immediate access to his credit history.

Even the approval can be given automatically, if need be, granted that Jack meets certain criteria. Before you know it, Jack walks out of the store, with the loan and TV in hand. How's that for efficiency?



Jack
wants to buy a new TV with a consumer loan.



Sales clerk
scans Jack's ID with a mobile and intelligent capture app.



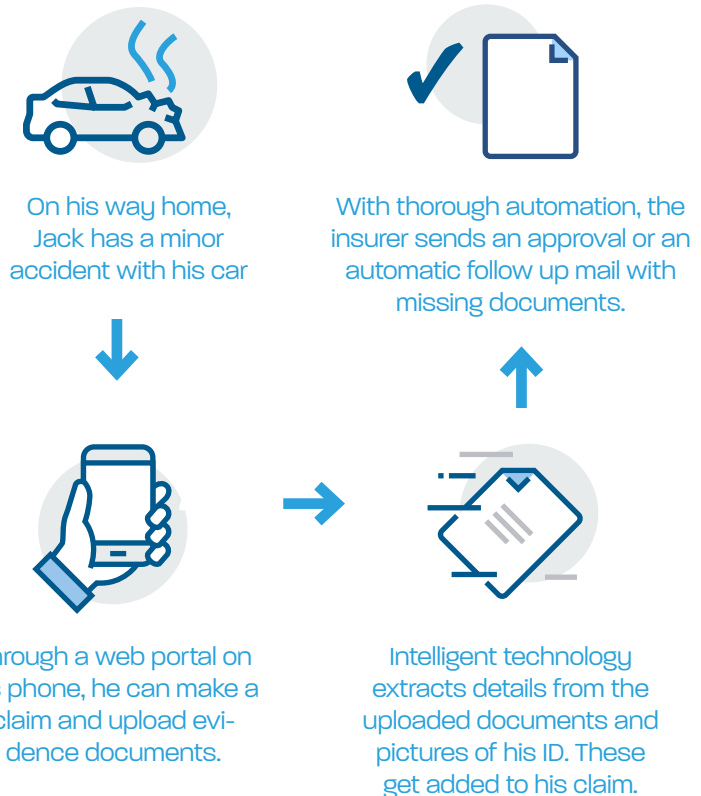
Data is extracted through intelligent technology, sent to the bank and a loan contract is drawn up.



Jack digitally signs the contract in a matter of minutes and walks out with his TV.

Easy to file damage claims through mobile capture

Disaster strikes. On his way home with his brand-new TV, Jack gets into an accident. Nothing serious, but his car bumper is damaged and one of the headlights is broken. He now needs to file a claim with his car insurance company. Usually, Jack sends a mail asking which documents he needs to submit to support his claim. It will take his contact a day or two to respond. Only after providing the documents is his claim handled. However, Jack recently switched to a new organization that uses a portal – available on the go via a web app – with intelligent capture. Jack now only needs to open the portal, click on 'upload evidence documents' and the software takes him through the step-by-step process.



First, Jack needs to enter his details and those of the person whose car he crashed into. The portal instructs him to take a quick snap of both driver's licenses. OCR knows where to find the necessary information on the document. Machine learning then extracts the right data to use for automated form filling or other steps in the process. Jack just needs to press the 'confirm' button to validate the data. Next, he is prompted to describe the damage to both cars. This is usually done using a paper European accident statement and Jack can take a picture of this. The data on the form can then be extracted and added to his claim. He then uses the application to take pictures of the damaged areas. Object recognition can even indicate dents, which Jack can confirm, or he can adjust the position and size of the indicator boxes.

Jack can do all this on the side of the road in a matter of minutes and his damage claim is immediately sent to his insurance company. His claim is logged and forwarded to the right people and departments, which have access to all the necessary information. If the insurance organization happens to have automated the next steps as well, Jack might even have an approval of his claim by the time he gets home. And if the shock of the accident made him miss a few documents? Then the automated systems will detect this and send an automatic message with a list of the missing files.



4. INCREASING YOUR INTELLIGENT CAPTURE SOFTWARE'S RELIABILITY

Not all intelligent capture applications are created equal. Varying levels of confidence are possible, based on the techniques applied. Docbyte believes we need to get that level as high as possible, because IC's advantages only come to fruition if you can get accurate data. Your processes will require less human intervention – reducing error rates, which in turn simplifies automating the next steps and increases your ability to comply with regulations. The following technologies are only a handful of what we at Docbyte consider crucial for good results.



4.1. Natural language processing

It's not much use having text digitized if your software doesn't understand what is written. That's why natural language processing (NLP) is crucial. NLP is an umbrella term that covers all techniques concerning the interactions between computers and human/natural languages. The goal is to teach machines to read, decipher, understand, and make sense of human language.



4.2. Liveness detection and face recognition

Do you want to make sure that the person submitting the information is who he or she claims to be? Face recognition in your IC application can verify this type of information through a simple selfie. Going one step further is liveness detection through a video clip. Fingerprints and selfies can be spoofed, but liveness detection can review more parameters on moving images, such as eye movements, skin texture, blink analysis, and more.

4.3. Pattern recognition through machine learning

A lot of intelligent capture software already uses machine learning to determine what is where on templates, such as invoices. The only downside is that it needs to be fed a sizeable stack of completed documents first. The software can deal with small variations, but if the layout changes drastically you need to start over. To avoid this, Docbyte goes one step further and utilizes pattern recognition so that it's possible to recognize and extract information template-independently.

With pattern recognition, algorithms don't evaluate field location. Instead, they analyze how the information within those fields is constructed. For instance, a bank account number in Belgium always starts with BE and two numbers. Or an address is always a street name, a house number, postal code, and city.



Become template-independent

Pattern recognition is an incredibly powerful tool. In many companies and tools, changes to templates result in many hours of programming or retraining models to adjust business rules and applications. That's a considerable time and cost investment just to keep your tools working the way they should. With pattern recognition, your template can change overnight and our algorithms will still be able to extract the right information.

Consequently, becoming template-independent has many benefits. The effort required to take into account a new layout can be invested in innovation and improving the customer experience and service. That's cutting costs while increasing your ability to make successful sales. Moreover, it speeds up the process of getting your IC solution online, once again reducing the work involved. At Docbyte, we can get a project online with multiple document types spanning several hundred layouts in record time and achieve over 95 percent accuracy.

5. STRUCTURING THE UNSTRUCTURED

A lot of the advantages that intelligent capture brings – if not all of them – stem from its ability to make sense of unstructured content such as videos, images, and mails. With traditional information transfer methods, such as post or email, where there is little enforced structure, this capability is key. Organizations with conventional capture technology have to dedicate a large percentage of their workforce to dealing with this huge amount of unstructured content, both old and new, and the various communication channels they are sent through. The result is that the information intake is not linked to business processes. This in turn leads to loss of time, human error, and long response times.

Even with intelligent capture, it can be challenging to structure incoming data as different techniques are required – depending on how data enters your company. Consequently, there is no one-technique-fits-all solution.

5.1. Why don't we just structure all incoming information?

In an ideal world, we would solve the challenge of unstructured content by making sure we only receive structured information. While that may sound plausible in theory, realistically speaking this is almost impossible. While paper volumes are diminishing and email relevance is waning, both won't be going anywhere in the foreseeable future. Moreover, new communication channels aren't necessarily tackling this problem either. Collaboration spaces such as Microsoft Teams and Slack are also not very well-organized in the way they transfer information. So we'll also have to rely on intelligent capture by using, for example, bots to make sense of the unstructured data.

5.2. Getting the help of customers and partners thanks to upload portals

We won't get rid of unstructured content entirely, but we can reduce the volume. The best way to do so is to encourage assistance from your information providers, i.e. your customers, partners, and other stakeholders. By asking them and educating them on how to send data to you in a structured way, you can at least make a dent in the heap of unstructured content flowing into your organization.

The best way to do this is by providing an upload portal, either through an app or web browser, where customers can upload the appropriate files or even fill out forms. This is often called capture at the edge, as you get your data by capturing it at the 'edge' of your network through the customer's own devices. You can also call this distributed capture.

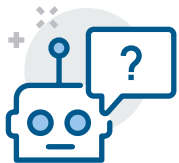
Even upload portals need IC

You are probably thinking: "If we can have customers deliver mostly structured information, do we still need to invest in intelligent capture?" The answer is yes. First, there will inevitably still be a varying amount of unstructured content flowing into your systems. Secondly, your clients won't take the brunt of the work. Adding tags about the information type, detailed values, and so on would simply take up too much of their time. So leveraging the functionalities of intelligent capture for a web portal would still be essential. It would minimize the workload for your customers. With a real-time IC product or mobile capture, documents are immediately uploaded digitally. Values are extracted on the spot and your clients just need to validate them. Moreover, this immediate feedback loop creates a sense of involvement, stimulating customers to use your upload portal in the future instead of paper or email.

6. ROBOTS VERSUS ALGORITHMS: IS RPA ENOUGH?

Any talk about intelligent capture should also mention robotic process automation (RPA), as there are quite a few similarities between the two. In a nutshell, RPA software follows all the manual steps a person would take to complete a certain task. To take over and automate the task, it then creates a list of those steps. Next, the software performs all the steps in the list just like a human would do in the user interface of the appropriate applications.

Based on manually coded rules, RPA can also be utilized to power a capture application. A rule can be used, for instance, to match the front and back of an ID. If the name on both sides match, they belong together and can be moved to that client's file.



6.1. Reaching the limits

We wouldn't exactly call RPA intelligent automation. For simple tasks, however, RPA is better than intelligent technologies with complex algorithms. It requires very little effort to set up compared to training machine learning models, making it more suitable for very structured, simple steps in your process.

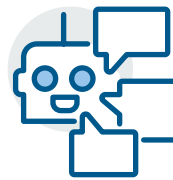
However, when push comes to shove, there's a limit to what RPA can automate. Firstly, robotic process automation requires your processes to follow the exact same course every time, but often there are deviations. Secondly, it's impossible to create rules for every plausible scenario. You can take into account ten different deviations, but not a hundred, let alone a thousand.



6.2. Machine learning to the rescue

When there are too many exceptions to implement, machine learning provides solace. Adding intelligent technologies to your process automation transforms RPA into intelligent process automation (IPA), of which intelligent capture is a part. A possible scenario where IPA, or at least machine learning, could prove to be more valuable than RPA is the approval or denial of loans.

A person's eligibility for a loan is based on their credit history, salary, and many other factors. Some ground rules can be laid down, but in the end it is a complex process where many parameters need to be considered. Moreover, company guidelines and rules shift constantly. Manually coding rules that adhere to all these parameters and exceptions is a Herculean task. Machine learning, however, constantly analyses all available data and comes up with those guidelines itself. After training for a while, algorithms understand without human intervention why a 950 euros loan will be approved and a 1000 euros loan won't be. ML can also discern the reason for denying the same loan at a later date. The result is software that can automatically judge without human intervention and with much less development effort.

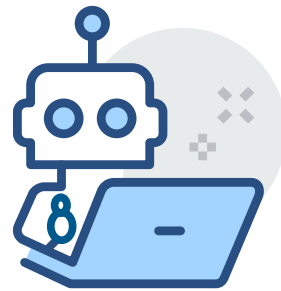


7. ONCE YOU GO INTELLIGENT CAPTURE, YOU NEVER GO BACK

The advantages of capturing your content intelligently, preferably through mobile capture, are abundant. IC paves the way for further automation, and ultimately, a more cost-efficient and compliant organization. And it's not just the internal workings of your company that'll fare well with it. Clients will get their requests answered much quicker, experience a better and smoother interaction with you, and many more benefits. Improving interactions will especially help you gain an advantage in an increasingly competitive environment where consumers aren't afraid of switching providers.

Artificial intelligence, machine learning, OCR – it's all here to stay and rightfully so. We guarantee that once you've added IC to your business processes, you'll never want to go back. So why not get a head start on the competition and unlock its advantages now?

Our experts are more than happy to help you figure out how best to enhance your processes with intelligent capture. Give us a call, write us an email, or pay us a visit.



PHONE: +32 (0)9 242 87 30

MAIL: info@docbyte.com

Kortrijksesteenweg 1144B
9051 Gent (St-Denijs-Westrem)
België

docbyte.com

