



# NOTES FROM THE FUTURE OF DOCUMENT AUTOMATION





BY THE TIME YOU'VE READ THIS GUIDE, YOU WILL UNDERSTAND THE BENEFITS THAT DEEP LEARNING CAN BRING TO YOUR DOCUMENT PROCESSING WORKFLOWS AND ULTIMATELY TO YOUR ORGANIZATION'S BOTTOM LINE.

YOU WILL KNOW WHAT'S NEW IN THE INTELLIGENT AUTOMATION SPACE AND WHAT'S LIKELY TO COME. YOU WILL ALSO DISCOVER WHY APPLICA IS THE MOST ADVANCED DOCUMENT AUTOMATION SOLUTION ON THE MARKET. AND YOU WILL GAIN A SENSE OF HOW COMPANIES THAT ADOPT A SOLUTION AS INNOVATIVE AS OURS ARE BOUND TO HAVE THE ADVANTAGE FOR YEARS TO COME.



# HOW MUCH ARE YOU ACTUALLY AUTOMATING?

While automation technology is likely familiar to your organization, legacy tooling has inherent limitations. In reality, the average enterprise today is only automating templated documents – which account for 20% of the document types that come through the door. Even if you are getting 99% precision on those documents, you are still struggling with relatively low straight-through processing rates due to the variability problem.

A simple fact is the majority of business documents don't share common templates or sometimes even the same language. Variable documents have long inhibited higher rates of automation and straight-through processing because documents are created by humans for humans, and thus, their variability is limited only by human imagination. As long as such a paradigm exists, variability remains a pivotal challenge.

Despite universal demand, the optimal solution to document automation simply did not exist until we arrived on the scene and designed it ourselves. Applica's singular approach to automation stems from our desire to solve problems that no one else has in order to make text-based work less repetitive and mundane, more error-proof, and as efficient as possible.

Nearly every automation tool on the market struggles with document variability as a result of using handcrafted rules and templates, which are far too simplistic to reflect the variability of real-world business documents. Crucially, other tools lack the all-important feature of scalability, which inoculates a business against the costs of fluctuation in the volume of incoming documents.

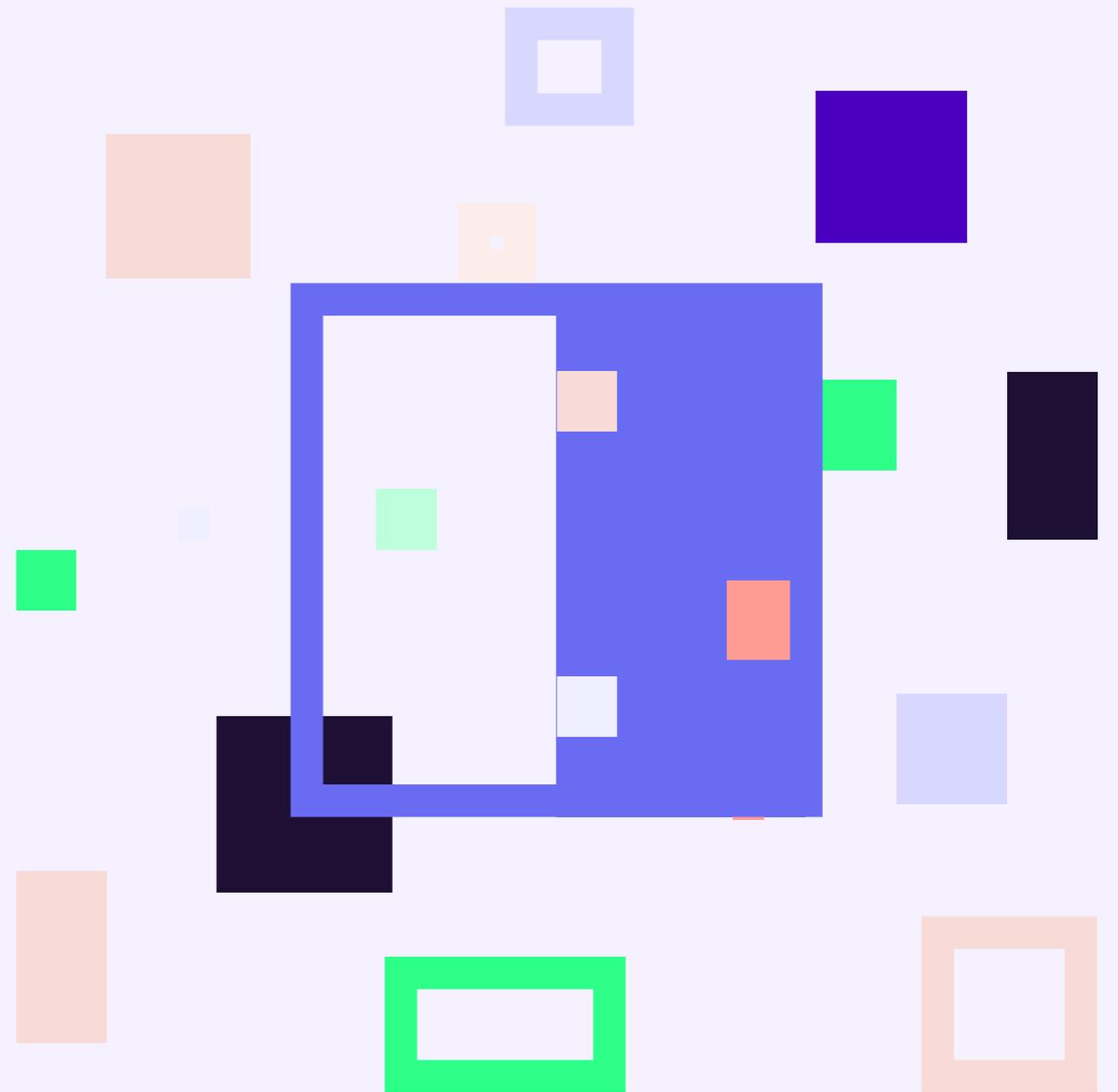


Applica's unique document automation solution reflects our understanding that deep learning is the only viable approach to solving for variability. Without an ability for the technology to "learn on its own" - known as "zero-shot" learning - there is no way to provide ubiquitous document automation. Moreover, our solution is easy to deploy by non-coders and it scales up and down to keep step with real-time document processing needs - for every type of document a business might handle.

TAKE A MOMENT TO ASK  
YOURSELF ONE SIMPLE  
QUESTION: WHY ARE YOU STILL  
ACCEPTING AUTOMATION  
TECHNOLOGY THAT ONLY MEETS  
A SMALL FRACTION OF YOUR  
BUSINESS NEEDS?

# TYPES OF DOCUMENT VARIABILITY OFTEN SEEN IN ORGANIZATIONS

1. Documents (forms) follow the same general format, but there still are minor differences (e.g.: a 2021 form is a bit different from the one used in 2020)
2. Documents (e.g.: invoices) could be in many formats with a “heavy tail” of various items
3. An organization can have a number of different documents sets





# WHAT KIND OF AUTOMATION IS RIGHT FOR YOU?

Choosing the right automation solution concerns the goals of automation and the technical aspects of the data requiring extraction. The question of goals requires analysis of the business process at stake. Is automation possible? Will automation speed, refine, or enhance the results? The question of technology concerns the optimal type of automation. Here, the choices concern document structure, software requirements, data storage constraints, engineer involvement, and – most importantly – the complexity of the data in question.

While some problems can currently be solved adequately using legacy rule-based systems (think pre-programmed decision tree computing rather than “true” AI), the industry as a whole is moving in the direction of deep learning.

Deep learning is new enough that it can easily confuse or instil fear. But considered from a holistic perspective, it’s continually proving to turn what was impossible into an everyday reality. For example, SpaceX can precisely maneuver a 12-story tall rocket hurtling through space and land it on an ocean barge at a designated time, over and over again. Deep learning taught that rocket how and where to land, all without any pre-programmed rules. If deep learning can provide the intelligence needed to land rocket ships, imagine what can happen when it is applied to document automation.



# THE EVOLUTION OF AUTOMATION TECHNOLOGY





# THE DEEP LEARNING ADVANTAGE

Whether the documents you need to automate are tightly structured, like invoices, or based on highly variable text with low predictability, like contracts, what works best is AI that adapts to any structure and actually comprehends the meaning of what is being ingested, classified, and extracted. What is needed is technology thus far associated with different domains – including self-driving cars and mRNA vaccines for COVID-19 – but not yet unleashed in the document processing sphere. This tech has to be dynamic, mutable, and able to integrate new knowledge even mid-task. It cannot be limited to a set of specific decisions, as is the case with traditional machine learning.

The system must be smart and agile enough to learn on its own, which can happen when the underlying technology is able to process text-based content in a similar way that humans do. And that is the very definition of deep learning.

What gives deep learning its unique edge with regard to the automation of text is the non-specific and extensive way it is trained – on a dataset orders of magnitude larger than that used with machine learning and, notably, without a specific procedural goal. And that's before the client is even on board. Then, when it is time to give a task to a deep learning-based model, the specific set of relevant examples and instructions can be negligible, plus you don't have to know every possible inquiry in advance. The system has, in a way, "seen it all," so now it is easy to teach it new tricks. Traditional machine learning, in turn, relies on modeling a hefty stack of training data – often thousands of documents – to a machine that is essentially a blank slate. You have to know what you are going after, and you better get it right the first time.



Deep learning can be thought of as the equivalent of a really fast and precise human knowledge worker who can think for themselves (and who never gets anything wrong), while machine learning is more like a fast, precise parrot. The latter is good at answering the questions you knew to ask. The former is free to address any question that may wind up relevant in the future.

In fact, it might even answer some you didn't know you had. Consider, for example, the chess engine AlphaZero, which learns chess by playing against itself. It tests out countless strategies and learns as it goes, extrapolating experience-based wisdom in that non-analytical way children acquire skills and knowledge through play.

Applica trains itself in a similar way, testing multiple hypotheses as it goes in order to learn how to "win" at extracting information from an infinite variety of documents.

According to Edward Benson, expert on intelligent text automation and founder of the natural language database startup NLUDB, deep learning has become an important component of any company's document automation strategy. First, because there is a valuable category of business problems for which deep learning is uniquely capable of automation. And, second, because deep learning can be combined with traditional IT techniques to make the whole system better than its parts.



# THE FUTURE OF DOCUMENT AUTOMATION

Deterministic approaches to automation, which rely on templates, rules, and coding, may be sufficient for straightforward tasks and fixed document types right now, but they lack the flexibility to adapt as needs and situations evolve. Like other rigid technologies, they have limited utility. Thus, investments in these technologies are bound to wind up obsolete when methods and workflows undoubtedly advance. Many products based on these legacy methods claim to be AI-based, but are easily revealed to be subject to so-called “AI washing.” According to Gartner, over one thousand software vendors describe themselves as specializing in or tied to artificial intelligence, though in many cases this is unfounded. In fact, Gartner analysts consider AI mischaracterization to be one of the top three problems currently impeding real development and adoption of artificial intelligence.

Most of the products that actually do involve artificial intelligence technologies are examples of what is known as weak AI. Also known as narrow AI, this category of solutions employs artificial intelligence technologies in a high-functioning system that replicates – and perhaps surpasses – human intelligence for a dedicated purpose. In contrast, strong AI (also known as artificial general intelligence or AGI) represents generalized human cognitive abilities in software, which allow it to self-train its way to a cogent solution also when facing an unfamiliar task.



In other words, weak AI entails well-designed algorithms, while strong AI designs algorithms well all by itself.

Automation expert Edward Benson lists several crucial advantage companies will gain if they are early adopters of deep learning-based automation technology. One is variance-robust document understanding, which results from the broad linguistic base that large-scale language models bring to the table. This means that the tech can learn about concepts rather than specific words. The second is multilingual processing as a default. This will allow companies to have a unified engineering approach across all of their geographic regions. The third consists in dramatically reduced training data requirements. After all, fine-tuning a pre-trained deep learning model takes a fraction of the time, effort, and expense of training it from scratch.

If an organization isn't currently utilizing a deep learning-based approach to automation, it is only a matter of time until the legacy products that are "working fine for now" will wind up unable to keep up with ever-evolving business needs. Fortunately, there is Applica's strong AI, which can be incorporated to handle all of an organization's automation needs. Unsurpassed in the marketplace, Applica continuously wins third-party benchmarks and technology awards against some of the largest tech companies in the world.



## WHY CHOOSE APPLICA?

Applica's unique deep learning approach is the only platform on the market that can handle document variability because it can learn on its own.

When our company was formed, we set out to harness the potential of strong AI and deep learning to automate unstructured and semi-structured documents, which had long resisted the capabilities of machine-based solutions. For years this had been the area no one wanted to touch, because the returns seemed far off and the technology seemed impossible. But then we happened.

Thanks to a tireless, insightful, and driven R&D team – and owing much to the way we swore allegiance to one specialized area – we achieved our goal of giving clients a robotic advantage where there hadn't been one before.

Today Applica makes it possible to extract information from documents that feature unpredictable content and running text. More importantly, we make this process fast, reliable, lucrative, scalable, and operable by end users who don't need to be engineers.

Applica's proprietary technology can be defined as a super-specialized type of intelligent document processing – ubiquitous, structure-agnostic, layout aware in a dynamic way, quick to train, and self-correcting. It is of-the-moment deep learning applied to document-based workflows, and it is able to process documents previously impossible to automate.



Every other platform must be told what to learn by highly skilled humans masquerading as AI, busy creating endless rules and/or templates on the backend. But no solution based on machine learning can match Applica's adaptability, versatility or strategic edge.

	Rules/template based-solutions	Natural language based-solutions	Applica's deep learning-based solution
Automates plain text	✗	✓	✓
Automates documents with tables	✓	✗	✓
Automates templated documents	✓	✓	✓
Automates highly variable documents	✗	✗	✓
Self-learns from a small number of examples	✗	Possibly	✓
Easy deployment and maintenance by business users	✗	✗	✓



## HOW APPLICA WORKS

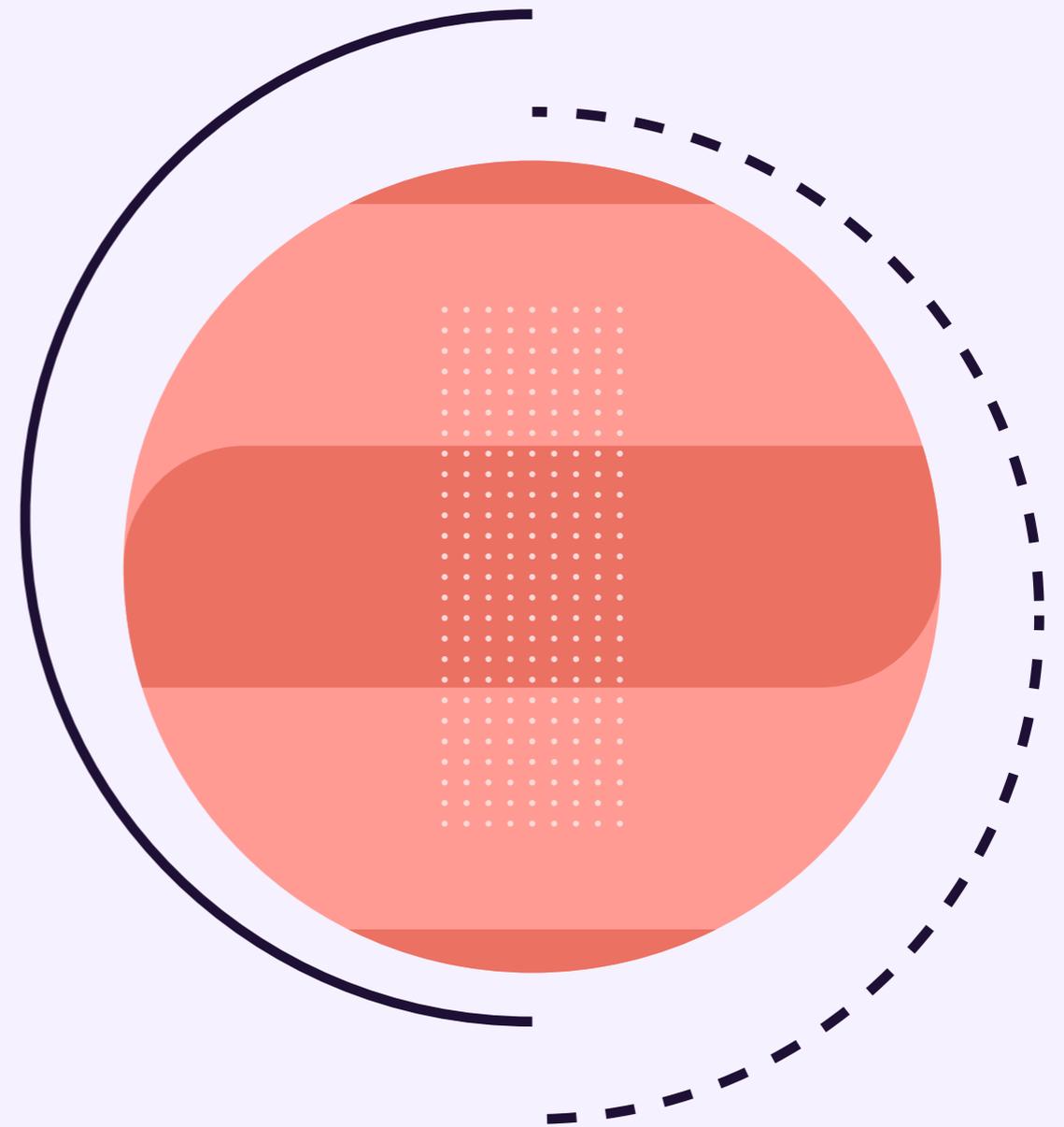
By combining technologies that previously handled only structured or only unstructured documents, Applica created an innovative hybrid model that can handle automation of a vast range of document types, whether scanned from paper or born-digital. Thus, by merging our exclusive layout-aware Language Model (LAMBERT) with 2D Contextual Awareness, Applica can handle not just plain text or forms in an “either-or” capacity, but crucially also documents in which information exists as both pure text and as parts in which the layout plays a greater role. Additionally, Applica’s innovative new language model TILT simultaneously learns layout, visual aspects, and textual semantics, along with exhibiting an unprecedented ability to “read” graphical embeddings. Having long eluded other AI solutions to text scrutiny, graphical embeddings, such as signatures, have been the holy grail along the path to full-spectrum document automation.

Another distinctive advantage of TILT is the model’s generativity, which maximizes the utility of each piece of training data. This is what allows the machine to learn on a minimal number of samples. And the less data the solution needs to see, the less time, effort, and cost is required for setting up every operation.

By considering both textual and graphical aspects before finalizing results, Applica can swiftly automate a form accompanied by instructions or disclaimers, or a report with both tables and long paragraphs. This mimics the way a human works with documents containing varying layouts and combinations of plain text, forms, tables, and charts. Coverage of all channels of information enables precise semantic analysis and information extraction, reducing the workload for some business processes by 90% and error rates by 85%.

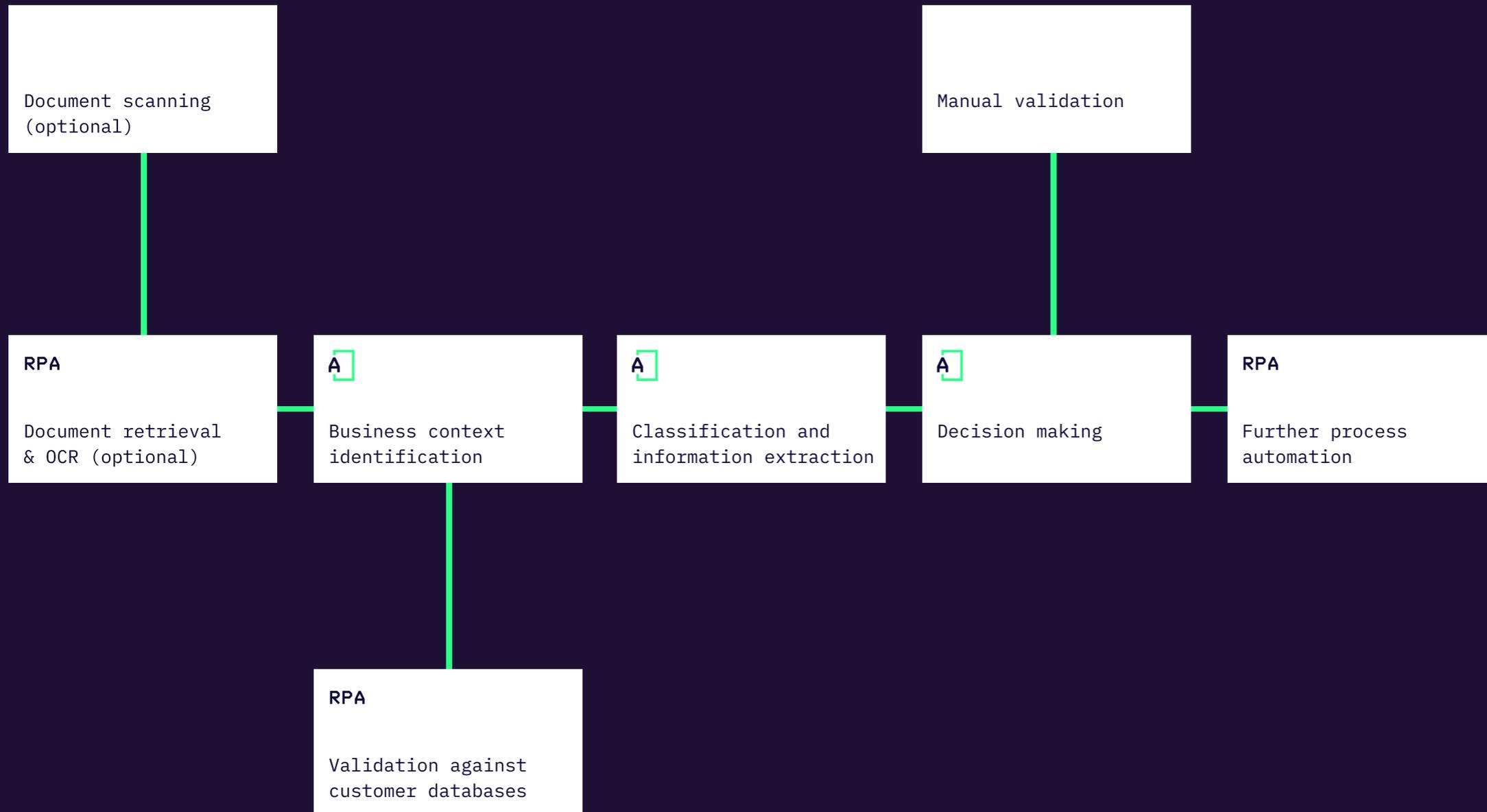


Because our unique method is not based on handcrafted templates or keywords, Applica can be applied to a wide range of document types. In fact, any time an organization is faced with unstructured or semi-structured documents containing business-actionable information requiring extraction, Applica can bring speed, accuracy, and scalability to the process. Once you define what is relevant and what you are looking for, Applica will seamlessly handle information extraction, classification, and decision making tasks.





# EXAMPLE WORKFLOW WITH APPLICA





CURIOUS?  
CONTACT US!

[sales@applica.ai](mailto:sales@applica.ai)