



## Executive Summary

In einem Satz: Intelligente Dokumentenanalyse mittels Natural Language Processing (NLP) auf Basis von Machine Learning (ML) ermöglicht **die Automatisierung wissensbasierter Prozesse** durch die Erfassung strukturierter Daten in Dokumenten.

Mit der digitalen Transformation verhält es sich wie mit vielen anderen Dingen: Wo Chancen entstehen, gibt es auch Risiken und damit verbundene Herausforderungen: Mehr Daten, mehr Informationen, mehr Dokumente bedeuten auch mehr Arbeit, diese auszuwerten, zu verarbeiten und sie im Unternehmenssinn zu nutzen. Sprich: **unstrukturierte Daten in strukturierte Vorteile zu verwandeln** und so die Brücke zwischen bereits existierenden und im Einsatz befindlichen Systemen und neuen Erkenntnissen zu schlagen, die in einer Vielzahl von Dokumenten schlummern (beispielsweise Verträge).

Herkömmliche Workflow- und Content Management-Systeme (zum Beispiel DMS) sind an und für sich hervorragende Lösungen. Ein besonderer Mehrwert entsteht durch die Verknüpfung der Systeme. Während viele triviale Geschäftsprozesse mittels regelbasierter Algorithmen gut optimierbar sind, liegen viele Dokumente mit prozessrelevanten Informationen ungenutzt in Content Management-Systemen und können nicht als strukturierte Daten herangezogen werden.

**NLP ist eine Lösung, um genau diese Herausforderungen in einen wertvollen Vorteil zu verwandeln.**

Im ersten Schritt werden dabei mittels Optical Character Recognition (OCR) Dokumente in ein maschinenlesbares Format übertragen. Im zweiten Schritt werden mit NLP auf Basis von Machine Learning aus diesen unstrukturierten Texten strukturierte Assets – in Sekundenschnelle und mit einer Genauigkeit, die mit einer Bearbeitung durch einen Menschen vergleichbar ist..

Es hängt vom individuellen Geschäftsprozess ab, wie und wofür NLP und Dokumentenanalyse eingesetzt werden können. Dies kann sein, um Prozesse komplett zu automatisieren, unerkannte Geschäftsrisiken zu identifizieren oder auch Fachkräfte durch automatisierte vorbereitende Tätigkeiten zu entlasten.

Auf jeden Fall aber kann intelligente Dokumentenanalyse die so immanent wichtige Brücke schlagen, die sich als Teil der digitalen Transformation zwischen Alt-Systemen und neuen Anwendungen ergeben hat. **Für nahtlose Prozessautomatisierung ohne große angelegte Architekturprojekte.**

## Herausforderung: Prozessrelevante Informationen sind oft in Fließtexten verborgen

Es gibt viele Branchen, die geschäftsbedingt ein hohes Dokumentenaufkommen haben – die Versicherungsbranche beispielsweise. Während ein Teil der Dokumente strukturierte Daten zur Weiterverarbeitung direkt beinhalten, zum Beispiel Formulare, ist dies bei freier Kommunikation oder offen gestalteten Texten nicht möglich. Verträge, Briefe und E-Mails folgen keinem vorgegebenen Muster, sie können nicht einfach automatisiert verarbeitet werden. Doch gerade in diesen Quellen verstecken sich viele wertvolle Informationen.

Hinzu kommt der zunehmende Wechsel von synchroner zu asynchroner Kommunikation sowie die schwindende Bereitschaft von Kunden, sich als "Dateneingabe-Helfer" im Laufe der Customer Journey zu betätigen. Convenience ist das Schlagwort des kundenorientierten Service-Design. Mit so wenigen Angaben wie möglich in einem idealerweise völlig freien Format zum Erfolg. Die Zukunft bringt es mit sich, dass die Komplexität von Kommunikation vom Anbieter selbst gelöst und nicht mehr wie gehabt zum Kunden verlagert werden kann.

Selbiges gilt für die interne Fähigkeit, lange Dokumente und große Mengen von Informationen in Textform schnell verfügbar und verwertbar zu machen. Qualifizierte Mitarbeiter sind heute schon Mangelware, ein Trend der sich in den kommenden Jahren noch weiter verschärfen wird. Kunden erwarten Resonanzen in Sekunden. Dies ist mit herkömmlichen Methoden nicht zu bewältigen.

Es geht also darum, bestehende Systeme schnell zu erweitern und zukunftsfähig zu machen, um den Prozesserfordernissen der kommenden Jahre zu begegnen. Möglichst ohne große IT-Projekte. Dabei erweist es sich als Vorteil, auf Bestehendem aufzusetzen und dieses sinnvoll zu erweitern. OCR als Technik ist in vielen Unternehmen Standard, ebenso Document Management Systeme (DMS), ERP, CRM und Robotic Process Automation (RPA).

Die Aufgabe besteht also darin, diese Systeme miteinander zu verbinden und neben einer effektiveren Kundenkommunikation Kostensenkungen und Zeitersparnisse zu realisieren. Ziel ist dabei immer eine Vollautomatisierung. Dies kann in Zeiten der Wissensarbeit nur dann gelingen, wenn Systeme mit tatsächlichem semantischem Verständnis aufwarten.

Kurz gesagt: Maschinen, die Dokumente wie Verträge lesen können, heben die Prozessautomatisierung auf ein neues Level.

## Case Study: Einkaufsoptimierung durch automatisierte Erkennung von Vertragsfristen

### Herausforderung:

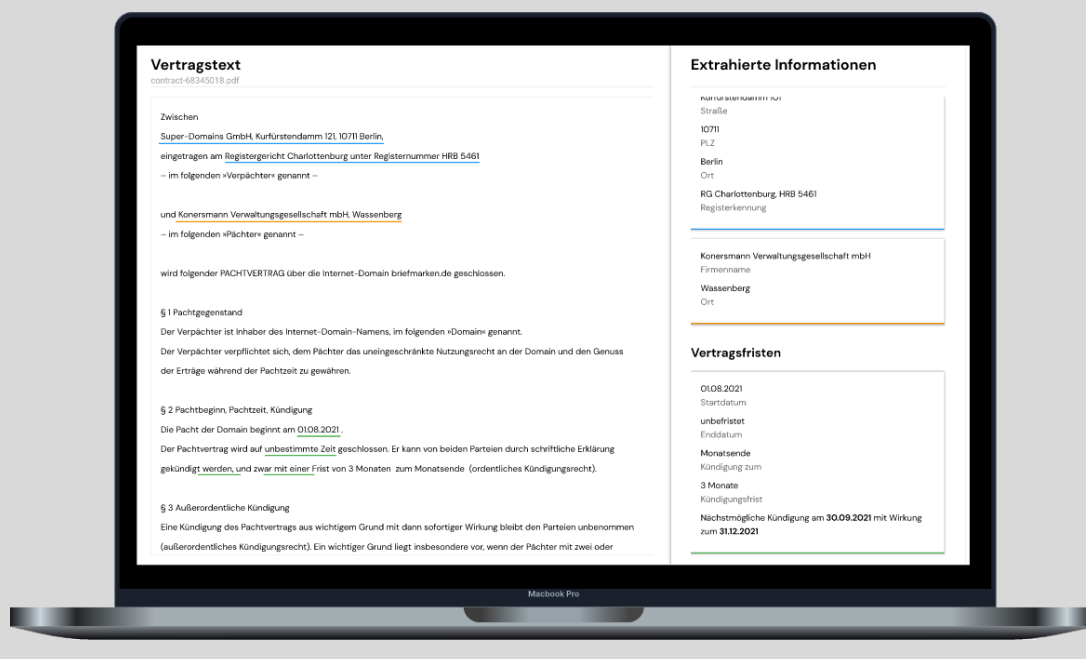
- Ein Hidden Champion der Automobilbranche hatte mehrere tausend aktive Verträge mit Kunden und Lieferanten.
- Kündigungszeitpunkte und Preisgleitklauseln konnten nur sporadisch nachgehalten werden. Für alte Verträge lagen die Information überhaupt nicht vor, für neue wurden sie mit schwankender Qualität erfasst.

### Lösung:

- Mit dem doculect Standardmodul "Vertragsstammdaten" wurden relevante Informationen aus Verträgen automatisch extrahiert.
- Kritische Fälle wurden Sachbearbeitern zum Review vorgelegt – Human-in-the-Loop.

### Ergebnis:

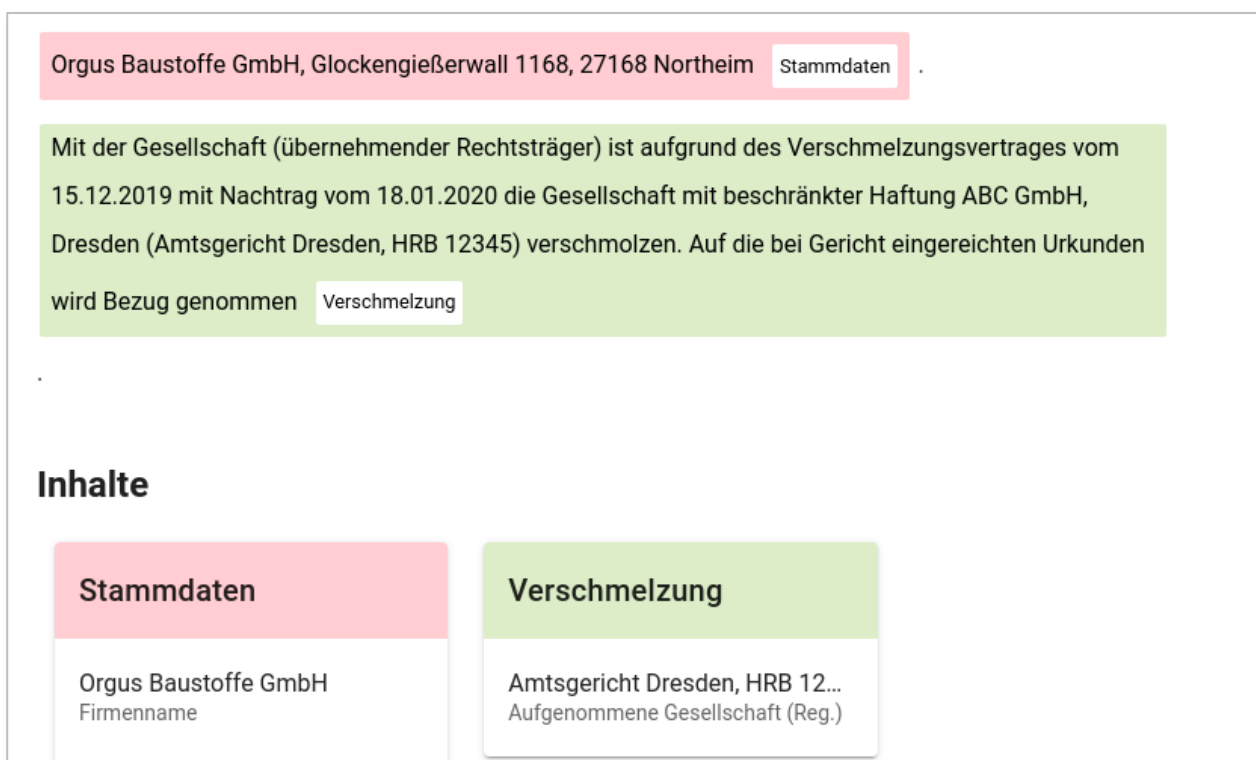
- Wichtige Vertragsfristen wurden sauber im DMS erfasst, verantwortliche Personen wurden rechtzeitig automatisch erinnert.
- In den folgenden 12 Monaten wurden Einsparungen in siebenstelliger Höhe erzielt durch rechtzeitig gekündigte beziehungsweise nachverhandelte Verträge und in Anspruch genommene Preisgleitklauseln.



## Lösung: automatische Extraktion strukturierter Informationen mit eingebauter Qualitätsschleife

Dokumentenanalyse ist ein sehr großer und diversifizierter Bereich mit einer Vielzahl von Use Cases. Die inhärente Komplexität speist sich dabei aus vor allem drei Quellen:

1. Unterschiedlichen Dokumententypen wie Verträge, E-Mails, Briefe, Formulare, Notizen, historische Dokumente aus Unternehmensarchiven, aber auch Transkriptionen von Audio-Dateien haben jeweils eigene Charakteristika und liegen in verschiedenen Formaten vor.
2. Relevante Informationen sind in Dokumenten nicht einheitlich formuliert. Eine effektive Dokumentenanalyse muss Informationen aus unterschiedlichen Begriffen und abhängig vom Kontext erkennen.



The diagram illustrates the process of document analysis. It shows a document with two highlighted sections: a pink section for 'Stammdaten' (Company Data) and a green section for 'Verschmelzung' (Merger). Below these sections, the extracted information is organized into two columns. The 'Stammdaten' column lists 'Orgus Baustoffe GmbH' as the company name. The 'Verschmelzung' column lists 'Amtsgericht Dresden, HRB 12...' as the court and 'Aufgenommene Gesellschaft (Reg.)' as the acquired company.

Orgus Baustoffe GmbH, Glockengießerwall 1168, 27168 Northeim **Stammdaten**

Mit der Gesellschaft (übernehmender Rechtsträger) ist aufgrund des Verschmelzungsvertrages vom 15.12.2019 mit Nachtrag vom 18.01.2020 die Gesellschaft mit beschränkter Haftung ABC GmbH, Dresden (Amtsgericht Dresden, HRB 12345) verschmolzen. Auf die bei Gericht eingereichten Urkunden wird Bezug genommen **Verschmelzung**

**Inhalte**

Stammdaten	Verschmelzung
Orgus Baustoffe GmbH Firmenname	Amtsgericht Dresden, HRB 12... Aufgenommene Gesellschaft (Reg.)

Abbildung: Je nach Zielvorgabe werden relevante Informationen extrahiert und klassifiziert.

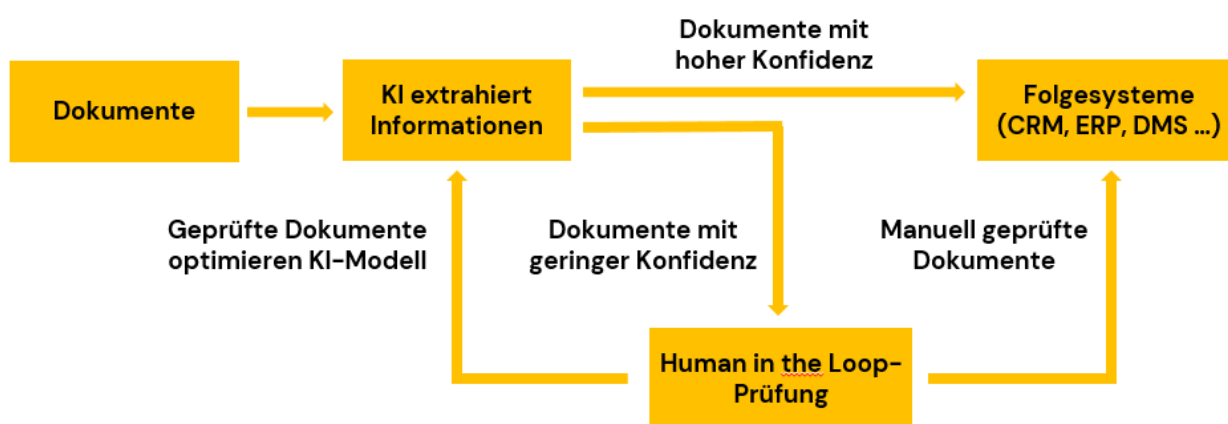
3. Informationen müssen entlang unterschiedlichster Prozesse aufgenommen und strukturiert verarbeitet wieder zurück übergeben werden, es bedarf also einer flexiblen Integration mit diversen Systemen.

Technisch muss außerdem zwischen einer Umsetzung als Echtzeit- und einer Massenverarbeitung der Dokumente („Dokumente“ hier und im Folgenden als Oberbegriff aller Daten, die digital vorliegen) unterschieden werden.

Wirtschaftlich ist hierbei zu beachten, dass es für wiederkehrende Echtzeitverarbeitung oft vorhandene KI-Modelle gibt, die individualisiert und auf entsprechende Ziele hin angepasst werden können. Bei der einmaligen Massenverarbeitung von Dokumenten hingegen ist es regelmäßig notwendig, Modelle dezidiert zu erstellen. Es gilt also zu überprüfen, ob der Einzelfall den Aufwand rechtfertigt und ob sich ein regelmäßiger Nutzen für das einmal trainierte Modell identifizieren lässt.

Massenverarbeitung:

In einem Archiv, das kontinuierlich mit allen digitalisierten Dokumenten sowie Zwischenergebnissen aus den Workflows angereichert wird, werden zentrale Ergebnisse gesammelt. Diese Erkenntnisse dienen dazu, KI-Modelle zu trainieren. Essenziell ist dabei der Human-in-the-Loop:



Die Überwachung und Evaluierung der Analyseergebnisse erfolgt durch Fachkräfte, um gegebenenfalls einzugreifen und zu steuern. Diese durch die Massenverarbeitung entwickelten Modelle kommen dann später in der Echtzeitverarbeitung unterstützend zum Einsatz.

Unsupervised Learning schafft hier besondere Mehrwerte, beispielsweise durch die Möglichkeit von Themenanalysen, die Aufschlüsse über die Inhalte eines Dokumentenarchivs semantisch zusammenfassen. So können Veränderungen im Kundenverhalten oder eine Häufung bestimmter Vorgänge identifiziert werden. Dabei sorgt eine spezifische Mustererkennung dafür, dass Auffälligkeiten erkannt werden und der Datenbestand qualitativ hochwertig bleibt. Fragliche Dokumente werden auf diesem Wege indiziert und für eine manuelle Betrachtung gesondert ausgewiesen und an den Human-in-the-Loop übergeben.

### Echtzeitverarbeitung:

Eine Optimierung von Prozessen/Workflows durch die Schaffung von strukturierten Daten ist das Ziel der Echtzeitverarbeitung in der Dokumentenanalyse. Hierfür werden fortlaufend lernende KI-Module verwendet. Vereinfacht formuliert erfolgt die Wertschöpfung in vier Schritten: Pre-Processing und Laden der Dokumente, Extraktion der gewünschten Informationen, gegebenenfalls Übergabe an den Human-in-the-loop oder an weitere definierte Redundanz-Prozesse und zuletzt Rückübergabe der gewünschten Informationen im definierten Datenformat zur weiteren Verarbeitung.

Entlang dieser Wertschöpfung kommen – je nach Anforderung und Datenquelle und Datenqualität – unterschiedliche KI-Module zum Einsatz.

So können zum Beispiel mit einem Klassifikationsmodell Eingangsdokumente automatisch dem richtigen Bearbeiter, Folgesystem oder Archivordner zugeordnet werden. Mit einem Extraktionsmodell können relevante Informationen wie Kunden- oder Vorgangsnummer, Vertragsparteien oder Details zum Vorgang (zum Beispiel Schadendetails in einer Versicherungsmeldung) automatisch erkannt und verarbeitet werden. Ein Clustering-Modell kann Unregelmäßigkeiten in der Kundenkommunikation oder auf Social Media Kanälen in Echtzeit erkennen und so ein rechtzeitiges Eingreifen ermöglichen.

### Zwischenfazit:

Natural Language Processing auf Basis von Machine Learning ist der Booster für intelligente Dokumentenanalyse und ermöglicht Next Level-Prozessautomatisierung.

## **Ergebnis: stärkere Prozessautomatisierung und ein strategischer Wissensvorsprung**

Intelligente Dokumentenanalyse schlägt die Brücke zwischen etablierten Anwendungssystemen und ermöglicht die Automatisierung der Wissensarbeit.

### Muster ermöglichen Potenziale – unleash the potential

In strukturierten Daten lassen sich Muster suchen und erkennen. Diese bergen regelmäßig ein enormes Potenzial für die Optimierung von Produkten, Dienstleistungen und Prozessen – welches manuell in keinem unternehmerisch vertretbaren Kosten-Nutzen-Verhältnis gehoben werden könnte.

Intelligente Dokumentenanalyse, die Bestandssysteme verbindet und unstrukturierte Daten auswertet, ermöglicht es erstmals, diese Potenziale skalierbar zu heben und die damit verbundenen Geschäftserfolge systematisch zu realisieren.

#### Geschäftsprozesse automatisieren und optimieren

Prozesse können durch Dokumentenanalyse beschleunigt werden – verkürzte Reaktionszeiten steigern Conversion und Kundenzufriedenheit gleichermaßen. Intern werden Prozesse vereinfacht, beschleunigt und kontinuierlich an sich eventuell ändernde Situationen angepasst. Gewisse Prozessschritte entfallen ganz, sodass sich Fachkräfte einzeln und Unternehmen im Gesamten mehr auf ihre Kernkompetenzen konzentrieren können.

#### Unerkannte Geschäftsrisiken identifizieren

In vielen Dokumenten und den damit verbundenen Prozessen schlummern nicht nur Potenziale, sondern auch Risiken. Diese effizient zu erkennen stellt einen großen unternehmerischen Vorteil dar. Mittels intelligenter Dokumentenanalyse können Unternehmen gezielt reagieren – und proaktiv agieren, um solchen Risiken vorzubeugen.

#### Fachkräfte unterstützen

Automatisierte Prozesse, die intelligente Dokumentenanalysen durchführen, entlasten wertvolle Mitarbeiter, die ansonsten in monotonen Routineprozessen gebunden sind. Durch diese Entlastung werden Kapazitäten für höherwertige Aufgaben freigesetzt. Als Nebeneffekt steigt so auch die Mitarbeiterzufriedenheit.

#### Quintessenz:

Sobald Dokumente digitalisiert und intelligent ausgewertet sind, ergeben sich vielfältige Möglichkeiten, die Wertschöpfungskette zu optimieren. Bisher unbekannte Muster und Zusammenhänge werden sichtbar, Potenziale erkannt, Risiken vermieden und Wissen über Prozesse extrahiert.



## Case Study: Cyber-Risiken in Gewerbeverträgen erkennen

Erkennen von expliziten und impliziten Einschlüssen von Cyber-Risiken in Gewerbeversicherungsverträgen im Großkunden-Individualgeschäft.

### Situation:

- Cyber ist explizit und implizit in Versicherungsverträgen eingeschlossen.
- Diese Verträge sind nicht einheitlich und können daher im Regelfall nicht strukturiert dokumentiert werden.
- Es besteht keine Klarheit über das gesamte Cyber-Risiko im Portfolio.

### Herausforderung:

- Die Formulierungen in Verträgen sind heterogen.
- Der Vertragsbestand ist aufgrund Indizierungen und unterschiedlichen Tarifgenerationen uneinheitlich.
- Underwriter sind als Ressource zu wertvoll, um sie mit einer manuellen Auswertung zu beauftragen.
- Herkömmlich werden für das Training eines NLP-Modells große Mengen an Daten benötigt.

### Lösung:

- Identifikation von typischen Formulierungen anhand einiger weniger Beispieldokumente.
- Training eines NLP-Modells mit entsprechend synthetisch generierten Trainingsdaten.
- Auswertung des Bestands mit dem trainierten Modell.

Unstrukturierte Informationen aus Verträgen und Schadenmitteilungen mittels **Natural Language Processing (NLP)** extrahieren.

### Direkte operative Benefits:

- **Mehr Umsatz** durch bessere Conversion von Offerten dank einer deutlich schnelleren Bearbeitung.
- **Geringere Kosten** in der Schadensachbearbeitung durch automatisch erstellte Zusammenfassungen der Schadensmeldungen.

### Ergebnis:

- Mehr Umsatz dank schnellerer Bewertung und Bearbeitung von Offerten.
- Strukturierte Datenbank aller historische Cyberdeckungen im Bestand.
- Neue Verträge werden automatisiert ausgewertet und hinzugefügt.
- Risiko- und Profitabilitäts-Bewertung für Cyber über das gesamte Portfolio in Echtzeit.
- Kumulanalyse für Cyber kann teilautomatisiert erfolgen.

## Roadmap: vom Proof of Concept zur Unternehmensplattform

Es empfiehlt sich, intelligente Dokumentenanalyse in einem strukturierten Prozess zu implementieren. Dabei sollte zuerst ein erfolgversprechender und klein geschnittener Proof of Concept (PoC) ins Auge gefasst werden, bevor das letztendliche Ziel einer autark nutzbaren NLP-Plattform für das gesamte Unternehmen konzipiert und umgesetzt wird.

Auf diesem Wege ist sichergestellt, dass die notwendige technische und fachliche Expertise sukzessive aufgebaut werden kann, sowie dass alle Mitarbeitenden den Wert der Technologie erkennen und schätzen lernen. Denn nur so kann eine Lösung entwickelt und eingesetzt sowie eine genaue Roadmap mit entsprechenden KPIs erstellt werden. Grob zusammengefasst:

- In einem Kick-off-Workshop wird die Ist-Situation evaluiert, das Problem eruiert und das Ziel definiert.
- In der ersten Phase, der Pilotlösung, wird das Problem durchdrungen, in Annahmen gegossen und durch Algorithmen und Modelle gelöst. Das Ergebnis ist ein Proof of Concept. Dauer etwa: 3-6 Wochen.
- In der zweiten Phase, der Implementierung und Integration, wird die Lösung für den Produktiveinsatz vorbereitet und iterativ in relevante Systeme und Prozesse integriert. Dauer etwa: 1-3 Monate.
- In der dritten Phase wird die Performance der Dokumentenanalyse evaluiert und überwacht – die Lösung wird kontinuierlich angepasst und optimiert. Diese Phase ist de facto ab der Implementierung und Integration ein fortwährender Prozess.

## Zusammenfassung

Strukturierte Daten aus Dokumenten zu extrahieren und wissensbasierte Prozesse zu automatisieren ist schon heute schnell und pragmatisch umsetzbar – ohne dass dafür langwierige Projekte mit vielen unterschiedlichen Abteilungen aufgesetzt beziehungsweise etabliert werden müssen.

Ebenso ist die technische Umsetzung weniger ein Thema, da eine intelligente Dokumentenanalyse-Plattform unabhängig von bestehenden Systemlandschaften genutzt werden kann. Dies gilt für den Fall, wenn ein neues Document Management System integriert oder ein bestehendes System angepasst werden soll, ebenso bei Robotic Process Automation.

## Empfehlung

Wenn Sie bereits Data Teams im Einsatz haben, sie bereits RPA aufgesetzt haben, Sie über ein aktuelles Echtzeit-DMS verfügen, dann lassen Sie sich beraten, wie Sie mit einfachen Mitteln und Lösungen schnell und pragmatisch aus unstrukturierten Daten wertvolle Assets generieren können. Denn eine intelligente Dokumentenanalyse kann die Brücke schlagen, die Sie auf Ihrem Weg zu unternehmerischen Herausforderungen zum Ziel führt.

## Über tetrel

Bei tetrel arbeiten autark agierende Data Science Teams für Unternehmen und Regierungsorganisationen an schlüsselfertigen Lösungen zur Dokumentenanalyse, um strukturierte Daten aus komplexen Dokumenten zu extrahieren. Diese Teams werden individuell je nach Anforderung zusammengestellt.

tetrel ist ein unabhängiges und eigenfinanziertes Unternehmen. Gemeinsam mit unseren Partnern ist es unsere Mission, strategische Anwendungsfälle für Künstliche Intelligenz zu identifizieren, zu bewerten und in produktive Lösungen umzusetzen. Dabei verstehen wir KI als Werkzeug zur Unterstützung hoch qualifizierter Mitarbeiter und Hebel zur effizienten Produktivitätssteigerung.

## Ihr Ansprechpartner



Johannes Humbert  
+49 176 83 33 51 46  
johannes.humbert@tetrel.ai