



# DER STARTPUNKT JEDER ANTWORT – **DIE FRAGE VERSTEHEN**



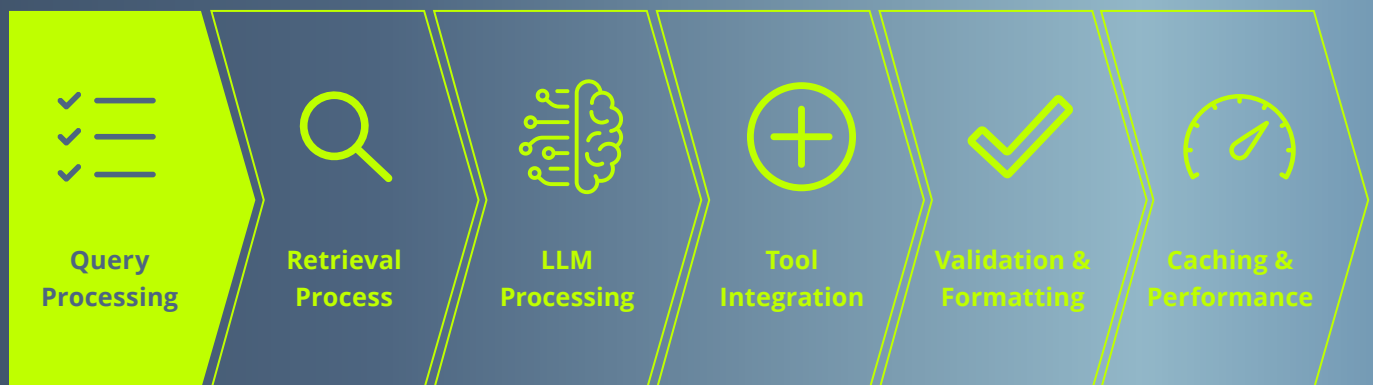
Query-Processing  
im Companion for First

# Wie Ihre Frage im Companion for First „verstanden“ wird – bevor die Antwort entsteht.



Der **Companion for First** ermöglicht Ihnen erstmals, Fragen in natürlicher Sprache an Ihre First Produktivumgebung zu richten: „Wie haben sich unsere Immobilienfonds im letzten Quartal entwickelt?“, „Wie hoch ist das Gesamtvolumen aller nachhaltigen Investitionen in diesem Jahr?“, „Wie lautet der Marktwert aller Aktien mit ESG-Rating A+ zum Stichtag 31. September?“ und unendlich vieles mehr.

Wenige Augenblicke nach der Eingabe erscheint die Antwort auf dem Bildschirm – als Zahl, Tabelle oder aussagekräftige Grafik. Zwischen der Ein- und Ausgabe durchläuft Ihre Anfrage die sogenannte RAG-Pipeline, das Herzstück des Companion for First. **In dieser Verarbeitungskette werden die Nutzeranfragen zunächst aufgeschlüsselt und interpretiert. Das ist die Aufgabe des „Query-Processings“, um das es heute gehen soll.**



**Das Query-Processing beschreibt den Weg, wie eine Nutzeranfrage verstanden, optimiert und für die Beantwortung vorbereitet wird. Hier entsteht der Plan, was als Reaktion auf die Anfrage geschehen soll, was also konkret zu tun ist, um die gewünschten Daten und Informationen zu liefern.** Ein wesentlicher Teil der Intelligenz des Companion for First ist deshalb im Query-Processing angesiedelt.

Im Anschluss geht es in den nachfolgenden Stufen der RAG-Pipeline weiter: Anhand des Plans aus dem Query-Processing wird die Nutzeranfrage aktiv in Datenzugriffe umgesetzt, mit weiteren relevanten Informationen verknüpft, das Ergebnis schließlich in prägnanter Form aufbereitet und als Antwort auf dem Bildschirm präsentiert. Das alles geschieht innerhalb von Sekunden.

**Das erste Ziel:  
Die Anfrage richtig  
verstehen**

# Einfach mal fragen

Bislang mussten Nutzer für frei gestaltete Abfragen an ihr First System eine komplexe Syntax erlernen, die Struktur der Datenbank (mehr als 1.500 Tabellen) verstehen, wissen, welche Felder wie verknüpft sind oder besser gleich die Kollegen vom IT-Support hinzuziehen. Doch das war alles gestern.

Heute fragen Sie den Companion for First, wie es Ihnen gerade in den Sinn kommt. Ganz gleich, was Sie in Bezug auf Ihre Daten und Bestände wissen möchten – auf Deutsch, Englisch, Französisch oder gerne auch in Fachchinesisch.

Der Companion zwingt Sie nicht in eine starre Ausdrucksweise oder Form. Er erkennt vielfältige Synonyme und Umschreibungen, kann komplexe Abfragen aus einem einzigen Satz herauslesen und flexibel auf neue Datenquellen (Assets) reagieren. Das System lädt dazu ein, Fragen zu stellen: für alle, die es genau wissen wollen – hier und jetzt.



## KI VERSTEHEN

In dieser Reihe bisher erschienen:

- Grundlagen der KI-Technik
- Die RAG-Pipeline im Companion for First

### Sie möchten mehr erfahren?

Alle Infos zum Companion for First auf einen Blick gibt es hier.

### Besuchen Sie auch den Fact Blog.

Dort finden Sie stets aktuelle Artikel zum Thema, etwa ein Interview mit dem technischen Projektleiter für die Entwicklung des Companion for First, Norman Janert.



# KI plus klassische Informatik

Wie schafft es der Companion for First, Nutzeranfragen in natürlicher Sprache zu verstehen? Dahinter steckt eine Mischung aus moderner KI-Technik und klassischer Informatik. Dieser Hybridansatz bedient sich des Sprachverständnisses großer externer KI-Modelle (Large-Language-Models bzw. LLMs wie ChatGPT, Claude, Gemini und Co.), erweitert durch das Wissen des Companions über Begriffe, Synonyme, Konstanten und Methoden, die im Zusammenhang mit First gebräuchlich sind.



Die KI-Technik in den LLMs sorgt bei dieser Aufgabenteilung dafür, dass der Companion nicht auf bestimmte Eingaben vortrainiert werden muss, dass er nicht bloß einen endlichen Satz von Anfragen kennt und bei allem, was auch nur einen Hauch darüber hinausgeht, ratlos zurückbleibt. Vielmehr sind nahezu unendlich viele Variations- und Kombinationsmöglichkeiten relevanter Begriffe erlaubt, wie es dem Wesen menschlicher Sprache entspricht. Nicht restlos alles wird die KI auf Anhieb verstehen können, aber das Allermeiste und Gebräuchlichste sehr wohl. Ansonsten hilft bekanntlich nachfragen – ebenfalls durch die KI.

## LLMs plus klassische Informatik in optimaler Aufgabenteilung

Zum Feld der klassischen Informatik im Companion for First gehört die semantische Suche, eine kontextorientierte Ähnlichkeitssuche auf Basis von Vektormodellen. Bevor Nutzeranfragen zum LLM gelangen, werden die zentralen Begrifflichkeiten innerhalb der **RAG-Pipeline** aufgelöst und normiert. Fragt der Nutzer nach Aktien, Wertpapieren, Equities, Actions (französisch für Aktien) oder Shares? In allen Fällen geht es um den gleichen Typus von Asset!

Mehr als 16.000 Begriffe werden auf diese Weise aufgelöst. Sie stammen aus den sogenannten „First Konstanten“. Das sind vom System vorgegebene Festwerte plus die System-einstellungen mit Geschäftslogik und Begriffsübersetzungen, die in jeder Instanz von First kundenindividuell gespeichert sind. Als **ProductType** zählen dazu beispielsweise Aktien, Anleihen, Immobilien, Fonds etc., als **KeyFigure** Marktwert, Buchwert, Rendite, Risiko oder als **Scenario**: Basis, Stress, Worst-Case.

# Wissen, was zu tun ist

Ein **Large-Language-Modell** ist darauf spezialisiert, Sprache in ihre Bestandteile zu zerlegen, diese in Beziehung zu setzen und daraus Sinnzusammenhänge zu erschließen. Aber wie kann es Depotbestände aggregieren, diese nach Datum gruppieren oder auf Basis von Filterregeln selektieren? Die klare Antwort: Es kann dies gar nicht. Und es soll diese Fähigkeit auch nicht beherrschen, weil alle sensiblen Daten etwa über konkrete Marktwerte in First unter Verschluss bleiben.

Stattdessen erstellt das LLM aus der Analyse der Nutzeranfrage und ggf. Rückfragen beim Companion einen abstrakten Ausführungsplan zur Ermittlung der gewünschten Daten. Dieser wird dann innerhalb des Companions lokal umgesetzt. Das ist die Aufgabe des **Tool-Processings**, dem nachfolgenden Schritt in der RAG-Pipeline, den wir uns in einem der kommenden Whitepaper genauer anschauen werden.

Verbleiben wir an dieser Stelle jedoch erst einmal beim **Query-Processing**. Das gleicht ein wenig einem Ping-Pong-Spiel zwischen dem externen Sprachmodell und dem Companion.

Es gliedert sich im Wesentlichen in **zwei Phasen**:

1. Im ersten Schritt geht es darum, die Anwenderfrage zu verstehen.
2. Im zweiten Schritt wird das passende Ziel-Werkzeug für deren Beantwortung ausgewählt und anschließend werden die erforderlichen Parameter aus der Nutzerabfrage abgeleitet, die in den Ausführungsplan des Companions einfließen.



## GUTE GRÜNDE, NICHT NUR AUF DIE KI ZU SETZEN

Die Vertraulichkeit der in First gespeicherten Daten und Bestände ist nur ein Grund, warum sich der Companion for First nicht exklusiv auf die KI-Funktionen externer Sprachmodelle stützt: Klassische (deterministische) Algorithmen haben zwar einen enger gefassten Wirkungsbereich, arbeiten dafür aber in der Regel schneller, präziser und verbrauchen weniger Ressourcen als vergleichbare KI-Algorithmen. Das beschleunigt die Abläufe, senkt die API-Kosten und erleichtert die Protokollierung aller Verarbeitungsschritte im Sinne einer durchgängigen Compliance.



# Phase 1

## Anwenderfrage verstehen

Zunächst gilt es, unspezifische Begriffe aus der Anfrage (z.B. „Wie ist der Wert meiner Aktien“) in First-Domänenbegriffe zu übersetzen, damit klar wird, worum es konkret gehen soll. Dies ist ein iterativer Prozess: Der Companion konfrontiert das externe LLM mit der Anwenderfrage und liefert im sogenannten „Kontext“ (Zusatzinformationen an das Sprachmodell) die Namen und Beschreibungen mehrerer Meta-Tools (übergeordnete Werkzeuge) wie **search\_constants()**, **search\_concepts()**.

Überall dort, wo das LLM Begriffe aus der Anwenderfrage nicht eindeutig auflösen kann, gibt es in seiner Antwort den Auftrag, das jeweils passende Meta-Tool aufzurufen, damit der Companion diese Begriffsaufklärung lokal innerhalb der RAG-Pipeline übernimmt. Die Ergebnisse werden anschließend an das LLM zurückgespielt.

So kann es mehrmals zwischen dem LLM und dem Companion hin- und hergehen. Sinnvollerweise wird dabei eine gewisse Obergrenze an Durchläufen vorgegeben, nach deren Erreichen klar ist, dass das LLM die Anfrage nicht vollständig auflösen kann (derzeit 10 Durchläufe). In diesem Fall wird der Anwender informiert und gebeten, seine Frage anders zu formulieren.



### OHNE SAUBERMACHEN GEHT ES NICHT

Das Query-Processing bildet das Tor zur Innenwelt des Companion for First. Was hier zur Weiterverarbeitung hineingelangt, muss frei von toxischen Rückständen sein. Böswillig eingebrachte Anfragen und Aufträge scheitern bereits an der strengen Toraufsicht. Speicherüberläufe durch überlange Inputs, unerlaubte Zeichencodes, manipulative SQL-Befehle, betrügerische Anweisungen an die KI u.v.m. werden aus dem Eingabestrom entfernt und dringen nicht bis zu den nachfolgenden Verarbeitungsstufen durch – damit das System jederzeit „sauber“ bleibt.

# Phase 2

## Werkzeug auswählen und Parameter ableiten

Ist die Begriffsaufklärung in der Phase 1 erfolgreich abgeschlossen, geht es in der nächsten Phase darum, das passende Werkzeug zu finden. Anschließend müssen die spezifischen Parameter für den Aufruf dieses Werkzeugs aus der Nutzeranfrage abgeleitet werden.

Für die konkrete Beantwortung von Nutzeranfragen kennt der Companion Hunderte von Werkzeugen (**Daten-Tools**), die Inhalte aus der First-Datenbank abrufen oder mit externen Datenlieferanten kommunizieren, jedes mit einer ganz spezifischen Aufgabe. Das LLM gibt die aufgelöste Nutzeranfrage an den Companion zurück und bittet diesen, über das Meta-Tool **search\_first\_tools()** passende Werkzeuge für diese Nutzeranfrage vorzuschlagen.

Die Auswahl erfolgt innerhalb der RAG-Pipeline auf Basis einer semantischen Suche über die aufgelöste Nutzerfrage. Die drei potenziell am besten geeigneten Werkzeuge werden mit ihren Namen und Beschreibungen an das LLM zurückgemeldet.

Dieses entscheidet sich im Vergleich zwischen der Nutzeranfrage und den Werkzeug-Beschreibungen für eines dieser Werkzeuge, wobei im Falle verbleibender Unsicherheiten immer das erstgenannte gewählt wird. Nun geht es für das LLM darum, die erforderlichen Parameter für den Aufruf des gewählten Werkzeugs in Erfahrung zu bringen. Dafür kann es sich eines weiteren Meta-Tools des Companions bedienen: **search\_fields()**

Mittels dieser Informationen bringt das LLM die geforderten Parameter mit den korrespondierenden Angaben aus der Nutzeranfrage zusammen. Teilweise müssen die Daten für diese Parameter selbst wieder mithilfe von **Daten-Tools** ermittelt werden (Stichwort: „alle Aktien“), wodurch sich ein **Tool-Graph** ergibt, der innerhalb der RAG-Pipeline aufgelöst und auf Korrektheit überprüft wird. Sobald alles passt, kann dieser Plan in der nächsten Stufe der RAG-Pipeline ausgeführt und so die gewünschte Antwort für den Nutzer gewonnen werden.

### DIE ROLLE DER DATEN- TOOLS



Die Daten-Tools bilden so etwas wie das Vokabular des Companion for First, jener Datenoperationen, die in seinem Wissensraum möglich sind und aus denen sich alle komplexen Abfragen zusammenfügen lassen. Es existieren mehrere hundert dieser Operationen, die grundsätzlich innerhalb des Companion ausgeführt werden, weil das LLM zu keinem Zeitpunkt an die First-Datenbank herangelassen wird.

Hier einige Beispiele:

#### **search\_assets()**

Suche nach Assets wie Anleihen, Wertpapiere etc.

#### **search\_positions()**

Suche nach Positionen in den Beständen.

#### **search\_results()**

Suche nach Ergebnissen wie Marktwerten zu einem gegebenen Datum und anderen Kennzahlen.

#### **search\_movements()**

Suche nach Bewegungen in den Beständen durch Käufe, Verkäufe und Abgänge.

# Ein konkretes Beispiel

## Schritt für Schritt



Montagvormittag: Sie möchten gut vorbereitet in das wöchentliche Teammeeting gehen und deshalb noch schnell einige relevante Kennzahlen abrufen. Im Companion for First geben Sie ein: „**Wie hoch ist der Marktwert unseres Aktienportfolios zum 31.9.2025?**“. Kaum haben Sie an Ihrem Kaffee genippt, erscheint auf dem Bildschirm bereits die Antwort: 178,4 Mio. EUR. Das sieht doch gut aus.

In den wenigen Momenten zwischen Eingabe und Ausgabe hat der Companion im Hintergrund eine ganze Phalanx von Operationen ausgeführt: Er hat Ihre Frage ausgewertet, die zur Beantwortung erforderlichen Daten identifiziert, einen Plan für den Abruf und die Weiterverarbeitung dieser Daten geschmiedet, den Plan auf Vollständigkeit und Plausibilität geprüft und schließlich in Form von Datenbankzugriffen und Ergebnisummutationen ausgeführt. Das Resultat sehen Sie vor sich. Doch wie muss man sich die einzelnen Phasen im Detail vorstellen?

### Schritt 1

#### Worum geht es?

Die semantische Suche und das externe Sprachmodell arbeiten beim Verstehen der Anfrage Hand in Hand. Sie erkennen, dass es in der Benutzeranfrage um folgende zentrale Elemente geht.

- **Aktienportfolio:** Gemeint sind offensichtlich alle Wertpapiere im Bestand vom Typ Aktien (interner Begriff: shares)
- **Marktwert:** Das System erkennt eine zugehörige Kennzahl mit dem Titel marketValueLC (für Market Value Local Currency). Darum soll es also gehen.
- **31.9.2025:** Dies ist ein Datumsfilter vom Typ dataDate. Auf dieses Datum muss sich die Abfrage also beziehen und auf keinen anderen Zeitpunkt.

### Schritt 2

#### Welche Daten werden benötigt?

Über die gelernten Datenbeziehungen erkennt das System, dass sich die benötigten Daten für „alle Aktien im Bestand“ aus drei Quellen speisen.

1. **Assets-Datenbank:** Welche Aktien hatten wir zum Stichtag im Bestand?
2. **Positionen-Datenbank:** Wie viele Stück jeder dieser Aktien lagen zum Stichtag in welchem Depot?
3. **Ergebnisse-Datenbank:** Was waren die Marktwerte dieser Aktien zum Stichtag?



### Schritt 3

## Wie sieht die optimale Reihenfolge aus?

Nun wird ein Plan gefasst, wie diese Daten abzurufen sind und vor allem, in welcher Reihenfolge, um tatsächlich alle Aktien zu berücksichtigen. Als Ausgangspunkt dafür dient wiederum das Graphenmodell.

1. **Finde alle Assets** mit dem Produkttyp „Aktien“
2. **Hole alle Positionen** zu diesen Assets
3. **Hole die Marktwerte** zum 31.09.2025 für diese Positionen
4. **Summiere die Ergebnisse** zum Endergebnis auf

### Schritt 4

## Passt der Plan?

Weil ein externes LLM in die Planerstellung involviert ist, macht es Sinn, den Plan vor der Ausführung unter verschiedenen Gesichtspunkten auf Vollständigkeit und Plausibilität zu prüfen. Sind alle benötigten Filter vorhanden? Sind alle Datenverknüpfungen korrekt? Fehlt kein Ausführungsschritt? Erst nach erfolgreicher Validierung geht der Plan in die finale Stufe. Das Query-Processing ist damit abgeschlossen.

### Schritt 5

## Planausführung und Ergebnisanzeige

Im Rahmen der sogenannten „Tool-Execution“ wird der ausformulierte Plan auf Basis der genannten Werkzeuge (**search\_assets()** etc., siehe oben) abgearbeitet. Die Ergebnisse werden dabei von Stufe zu Stufe weitergereicht und verknüpft. Datenbankabrufe münden in der Regel in korrespondierende SQL-Befehle für die First Datenbank, mathematische Operationen werden, wie im Plan definiert, ausgeführt. Beide Vorgänge, das Tool-Processing und die Ergebnisaufbereitung, werden wir in nachfolgenden Whitepapern noch genauer beleuchten.

Als Resultat der Benutzeranfrage steht schließlich ein Ergebnis, das vor der Ausgabe noch passend als Zahl, Tabelle oder Grafik aufbereitet wird. Die Verarbeitungskette im Companion for First kommt damit erst einmal zum Abschluss. Weitere Anfragen können aufgrund der asynchronen Verarbeitungsmöglichkeiten des Companion jedoch bereits in Arbeit sein. Sie müssen mit der Eingabe der nächsten Anfrage also nicht warten, bis eine vorangehende Abfrage abgeschlossen ist.



# Nächste Haltestelle: Tool Execution

Soweit zum Query-Processing im Companion for First. Wir bleiben der RAG-Linie noch für weitere Stationen treu und werden uns in der nächsten Folge mit der Tool-Execution beschäftigen. Dann wird es darum gehen, wie der Companion den erstellten Plan abarbeitet, um zu eindeutigen, belastbaren und reproduzierbaren Resultaten als Antwort auf die gestellten Fragen zu gelangen.

## Neugierig geworden?

Sie möchten mehr über First Cloud, den **Companion for First** und den Betrieb in der Fact Cloud erfahren? Dann schreiben Sie uns oder sprechen Sie direkt mit unserem Experten Aleksandar Ivezić. Er freut sich auf Ihre Anfrage:



**Aleksandar Ivezić**

[a.ivezic@fact.de](mailto:a.ivezic@fact.de)

+49 2131 777 238



### **Arbeitet der Companion for First vollständig KI-basiert?**

Nein! KI-Funktionen werden im Companion ausschließlich für die Sprachverarbeitung, die Erstellung von Ausführungsplänen und die visuelle Aufbereitung der Ergebnisse genutzt. Darum herum erfolgen viele Verarbeitungsschritte mit den Methoden der klassischen Informatik, weil diese häufig schneller, präziser und ressourcenschonender zum Ziel führen.

### **Muss ich beim Umgang mit dem Companion sehr auf meine Wortwahl achten?**

Nein, und das ist einer der maßgeblichen Vorteile von KI-Sprachmodellen. Außerdem beherrscht der Companion viele Synonyme und Fachbegriffe, mit denen er externe Sprachmodelle bei der Interpretation Ihrer Anfragen unterstützt. Es lohnt sich, ein wenig mit dem Companion herumzuexperimentieren und zu entdecken, wie man mit möglichst wenigen Worten zum gewünschten Ziel kommt.

### **Kann ich mit dem Companion ungewollt Daten löschen oder fremde Daten einsehen?**

Nein, das ist nicht möglich. Der Companion ruft derzeit nur Daten ab und berücksichtigt dabei die auf Nutzerebene definierten Zugriffsrechte.

### **Wie wird der Companion vor Betrug und Manipulation geschützt?**

Der Companion gibt Ihre Anfragen nicht ungefiltert an externe Sprachmodelle. Unerlaubte Zeichen, SQL- oder HTML-Befehle, manipulative Anweisungen an die KI und vieles mehr werden zuvor aussortiert.

### **Was geschieht, wenn der Companion mich nicht versteht?**

Dann wird er Sie freundlich auf die betreffende Stelle Ihrer Anfrage hinweisen. Gelegentlich wird er auch nachfragen und Ihnen Auswahlmöglichkeiten anbieten.

### **Kann ich mehrere Anfragen gleichzeitig an den Companion richten?**

Ja, das ist möglich, weil der Companion asynchron arbeitet und mehrere Anfragen in separaten Prozessen gleichzeitig ablaufen können.

### **Kennen externe Sprachmodelle die Struktur der First Datenbank?**

Nein, das genutzte LLM kennt ausschließlich die Namen der Werkzeuge, welche für die Datenabfrage innerhalb der RAG-Pipeline eingesetzt werden. Wie diese Werkzeuge die Daten konkret ermitteln und in welcher Verbindung die Daten zueinanderstehen, bleibt für diese externen Module transparent.

### **Fließen die vom Companion übermittelten Daten in das Training und damit in das Wissen der KI ein?**

Nein. Bei der bezahlten Nutzung eines externen LLM wie im Falle des Companion for First sichert der KI-Anbieter zu, diese Daten in keiner Weise zu speichern oder für das Training seines Modells zu verwenden. Sie werden sofort wieder gelöscht und damit vergessen.



#### Hauptsitz Neuss

Fact Informationssysteme & Consulting GmbH  
Hellersbergstraße 11 | 41460 Neuss | Tel.: +49 2131 777-0

#### Standort Frankfurt am Main

Fact Informationssysteme & Consulting GmbH  
Wilhelm-Leuschner-Straße 81 | 60329 Frankfurt am Main | Tel.: +49 69 8740313-121

E-Mail: [mail@fact.de](mailto:mail@fact.de)

Wir haben Ihnen dieses Whitepaper gerne zu unverbindlichen Informationszwecken überlassen. Bitte beachten Sie aber, dass die darin enthaltenen Informationen allgemeiner Natur sind und eine Beratung im konkreten Einzelfall nicht ersetzen können.

Die Fact Informationssysteme & Consulting GmbH hat diese Unterlage nach bestem Wissen erstellt und die Inhalte sorgfältig erarbeitet. Die Informationen werden ständig geprüft und aktualisiert. Gleichwohl können Fehler nicht ausgeschlossen werden. Bitte haben Sie deshalb Verständnis dafür, dass wir keine Garantie und/oder Haftung für die Aktualität, Richtigkeit und Vollständigkeit übernehmen. Infolgedessen haften wir nicht für direkte, indirekte, zufällige oder besondere Schäden, die Ihnen oder Dritten durch die Verwendung der Informationen dieser Unterlage entstehen.

Inhalt, Darstellung und Struktur dieser Unterlage sind urheberrechtlich geschützt und eine Nutzung, Verwendung, Reproduktion oder Weitergabe an Dritte – ganz oder teilweise – ist nur mit unserer ausdrücklichen vorherigen schriftlichen Zustimmung zulässig.

Öffentlich – Alle Rechte an den Inhalten dieses Whitepapers liegen bei der Fact Informationssysteme & Consulting GmbH.

Stand Dezember 2025



Folgen Sie uns  
auf LinkedIn



Alle im Whitepaper verwendeten Bilder wurden mit Unterstützung des KI-Tools Adobe Firefly generiert. Sie dienen ausschließlich der visuellen Veranschaulichung und haben keinen dokumentarischen Charakter.