



Integration von KI in der BI

Status Quo und 5 Thesen

Künstliche Intelligenz oder KI ist das auffälligste Trend-Thema, häufigste Buzzword und der größte Hype, den die Tech- und Business-Welt gegenwärtig kennt. Für die Allgemeinheit durch die Veröffentlichung von OpenAIs ChatGPT 3.5 im November 2022 schlagartig in den Fokus gerückt, gibt es schon seit einigen Jahren spannende KI-Entwicklungen in zahllosen Anwendungskontexten. Als Business Intelligence Consultants beschäftigt und begleitet uns dieses Technologie-Feld schon eine Weile. Es fällt durch zahllose Tool- und Feature-Releases nicht leicht, den Überblick zu behalten, wie KI-Technologien Business Intelligence und Data Analytics verändern. Daher versuchen wir in diesem Text einen Überblick über den Stand der Künstlichen Intelligenz in der Business Intelligence (BI) zu geben und formulieren fünf Thesen, wie sich die BI-Welt entwickeln könnte. Wir konzentrieren uns hierbei auf Tools, die sich klar der Business Intelligence zuordnen lassen und lassen andere KI-Anwendungen wie zum Beispiel ChatGPT oder Microsofts Image Creator weitgehend außer Acht.

Übergeordnete Zielsetzung und Wirkweisen von KI in BI-Tools



Übergeordnet konzentrieren sich Anbieter von BI-Software mit der Einbindung von KI in ihre Anwendungen bis dato auf drei Ziele:

1. **Steigerung der Qualität** von Prozessen und Analyseergebnissen bzw. Insights, um somit bessere, auf diesen Daten beruhende Entscheidungen zu ermöglichen.
2. **Steigerung der Effizienz und Geschwindigkeit**, mit der Insights generiert oder Daten bereitgestellt werden können. Damit geht auch eine optimierte Performance von Abläufen einher.
3. **Reduktion der Hürden, Insights durch BI-Tools zu generieren**, um einer möglichst breiten Menge von Anwenderinnen und Anwendern den Zugang zu datengetriebenen Entscheidungen zu ermöglichen.

Je nach Anwendungsgebiet und Produkt unterscheidet sich die Zielsetzung der KI-Einbindung natürlich stark. Zudem lässt sich beobachten, dass Anbieter von BI-Lösungen KI-Technologien auf drei verschiedene Arten und Weisen einsetzen oder einsetzen wollen, um diese Ziele zu erreichen (siehe auch Darstellung 1):

1. Zunächst wird KI-Technologie verwendet, um neue bzw. automatisierte Formen der **Datengenerierung bzw. -sammlung** zu ermöglichen (z.B. die Extraktion analytisch verwertbarer Daten aus Bild-, Text- oder Video-Dateien).
2. Weiterhin werden KI-Technologien eingesetzt, um die **Interaktion mit Daten** bzw. dem Tool, das die Daten verarbeitet, zu verändern bzw. zu verbessern. Dazu gehören insbesondere Natural Language Processing (NLP) Technologien, welche die Interaktion mit Daten vereinfachen sollen bzw. Abfragen in Form menschlicher Sprachmuster ermöglichen („Zeig mir alles...“ statt „Select * from ...“).
3. Zudem werden KI-Technologien verwendet, um **Hinweise** auf potenziell Relevantes (z.B. Ausreißer, Treiber von Trends, mögliche Fehler etc.) und **Empfehlungen** für das weitere Vorgehen zu machen (z.B. Vorschläge für Matching von Daten, zur Optimierung von ML-Modellen oder für die Bereinigung unvollständiger Informationen).



Künstliche Intelligenz in den Teilbereichen der Business Intelligence



Business Intelligence kann in fünf Teilbereiche untergliedert werden: Datensammlung, Datenintegration, Datenspeicherung, Datenaufbereitung sowie Datenanalyse und Informationsdarstellung. Im Folgenden werfen wir einen Blick in jeden dieser fünf Teilbereiche und geben jeweils einen Überblick über bestehende Anwendungen von KI darin.

Daten-
sammlung

Daten-
integration

Daten-
speicherung

Daten-
aufbereitung

Daten-
analyse und
Informations-
darstellung

Übergeordnet stellen wir fest, dass KI bis dato am erfolgreichsten an den „Außenpositionen“ der BI-Teilbereiche implementiert ist: Der Datensammlung bzw. -generierung und der Visualisierung & Analyse. Insbesondere im Bereich der Datengenerierung hat sich seit einigen Jahren ein erfolgreicher Markt von Anbietern entwickelt, die u.a. KI nutzen, um unterschiedlichste Daten in analysefähiger Form aus anderen Datenformaten zu gewinnen. Beispiele sind Bild-Analysen von Satellitenaufnahmen, Analysedaten von Videoaufzeichnungen von Sportereignissen oder Extraktion von Analyse-relevanten Daten aus Texten.

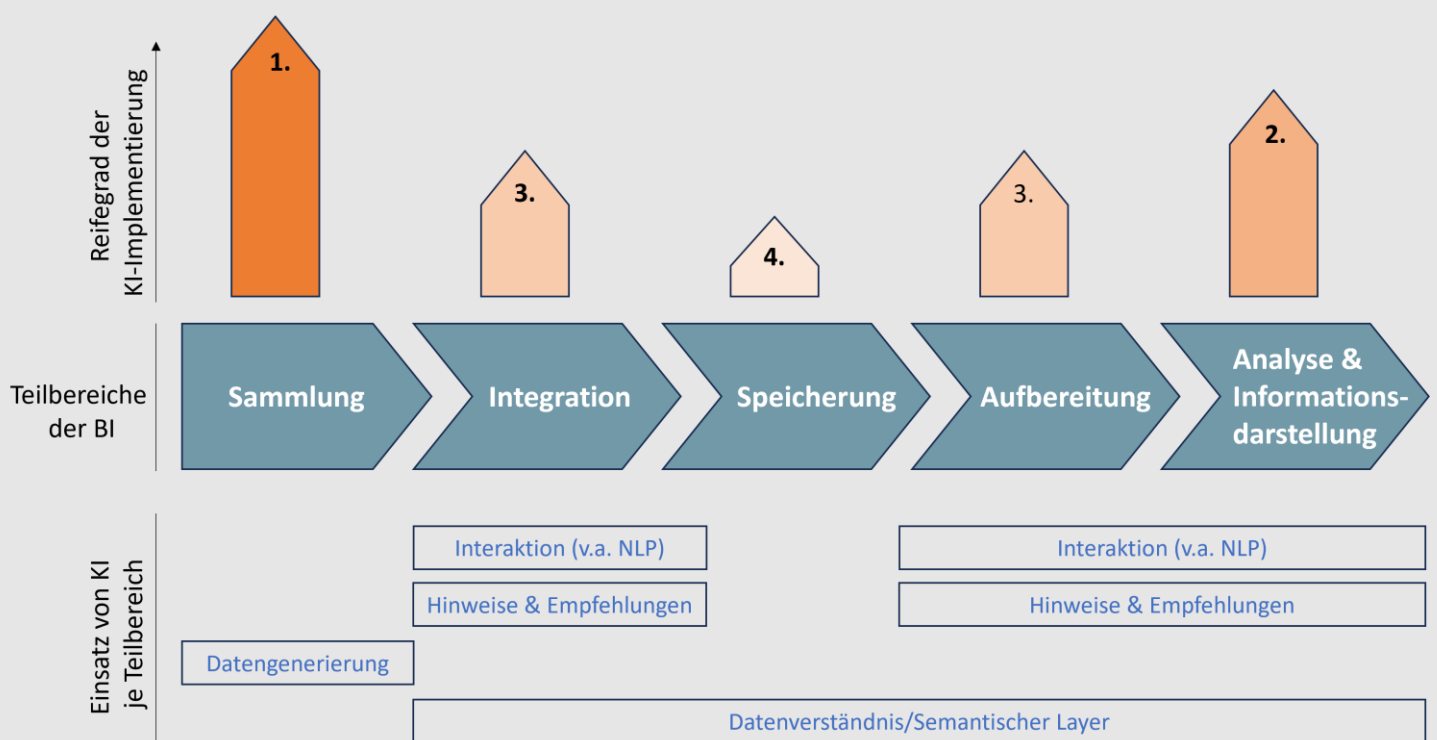
Gegenwärtig implementieren zudem viele Anbieter von BI-Applikationen KI in ihre Produkte zur Datenvisualisierung oder -analyse, insbesondere auf Basis von LLMs und NLP. Die meisten dieser Features fußen letztendlich darauf, die Interaktion mit Daten der menschlichen Sprache anzunähern und so auf Basis von Texteingaben nach menschlicher Logik Datenabfragen, -visualisierungen und -analysen aufzubauen.

Vergleichsweise gering ist die Ausprägung von KI in den Bereichen der Datenintegration, -speicherung und -aufbereitung. Auch hier bewegen sich einige Produkthanbieter in Richtung von LLM- und NLP-Implementation, tendenziell ist die Dynamik jedoch langsamer. Hintergrund ist vermutlich eine verhältnismäßig hohe Komplexität, die hinter den Datenverarbeitungsprozessen steckt und die schnelle Implementation von KI herausfordernd macht.

Künstliche Intelligenz in den Teilbereichen der Business Intelligence

In Darstellung 1 bilden wir den Rang der einzelnen Teilbereiche von BI anhand des Reifegrads der eingesetzten KI ab. Wie beschrieben wird in der Erschließung von Datenquellen seit Jahren erfolgreich KI eingesetzt. Im letzten Schritt der Datenanalyse & -präsentation wurden in kurzer Zeit viele Features auf Basis von Generativer KI veröffentlicht. Dazwischen befinden sich die meisten Features in einem sehr frühen Stadium und die Zahl der Use Cases ist begrenzt. Dabei bedeutet Rang 1 mitnichten eine Perfektion, hier ist KI lediglich am besten etabliert.

Die Einsätze von KI in den Teilbereichen wurden schon beschrieben. So sieht man beispielsweise, dass NLP oder Empfehlungssysteme an unterschiedlichen Stellen eingesetzt werden. Ein die Teilprozesse übergreifendes Thema ist das Metadatenmanagement, mit dem ein möglichst einheitliches Datenverständnis und ein semantischer Layer etabliert werden sollen, sobald die Daten in den Verarbeitungsprozess eintreten.



Darstellung 1: Reifegrad und Einsatzgebiete von KI in der BI

Künstliche Intelligenz in den Teilbereichen der Business Intelligence



1. Datensammlung

Datensammlung meint im BI-Kontext üblicherweise den Vorgang der Identifikation relevanter Daten und der Daten-Generierung, die in weiteren Prozessschritten verwendet werden sollen. Hierbei wird klassischerweise bzw. ohne Einsatz von KI zunächst bestimmt, aus welchen (meist internen) Systemen oder (meist externen) Quellen Daten verwendet werden sollen. Zu diesen wird dann über verschiedene Zugriffsmethoden ein Zugang gelegt und diese werden (zunächst unverändert bzw. als rohe Kopie) in einen Staging Layer eines Data Lakes / Data Warehouse o.ä. geschrieben.

Schon heute wurde je nach Kontext ein teilweise hohes Maß an manueller Arbeit insbesondere in der Datenerfassung durch KI-Anwendungen automatisiert. Beispiele sind etwa Bild-Analysen von Satellitenaufnahmen, Analysedaten von Videoaufzeichnungen von Sportereignissen oder Extraktion von Analyse-relevanten Daten aus Texten. Hierbei handelt es sich je nach Perspektive jedoch nicht um Business-Intelligence-Tools oder -Anbieter im herkömmlichen Sinn, sondern eher um Methoden oder Anwendungen zur Datengenerierung.

Andere Beispiele sind etwa [Infatica](#), die durch den Einsatz von KI bzw. NLP verändern, welche Daten durch Web Scraping generiert bzw. extrahiert werden können.

2. Datenintegration

Datenintegration beschreibt im BI-Kontext üblicherweise Vorgänge vor und in Data Warehouses oder Data Lakes, die Daten verschiedener Quellsysteme und Datentöpfe miteinander in Beziehung setzen und in eine vereinheitlichte Form bringen. Zielsetzung ist, darauf aufbauende Prozesse der Datenbereitstellung und -analyse durch die Vereinheitlichung und Qualitätssicherung zu vereinfachen. Je nach Verständnis des Begriffs kann Datenintegration lediglich den Vorgang des Datentransports von A nach B beschreiben oder sich auf die Herstellung von Beziehungen verschiedener Daten untereinander beziehen. Für die Betrachtung in diesem Text spielt diese Unterscheidung jedoch keine wesentliche Rolle.

Tools wie [Fivetran](#) nutzen KI-Technologie, um den Prozess der Anbindung weiterer Datenverbindungen zu ermöglichen. Hierbei wird jedoch keine KI-Technologie im Tool selbst eingesetzt, sondern KI-gestütztes Coden verwendet, um den Coding-Prozess selbst zu beschleunigen. Sobald die Daten einmal ins verarbeitende System gebracht sind, gibt es Anbieter, welche KI dazu verwenden, die weitere Verarbeitung zu optimieren und beschleunigen. Hierzu zählen Mechanismen zum automatisierten Erkennen, Definieren und Clustern von Daten (bspw. vorhanden bei [Alteryx](#), [RapidMiner](#), [Microsofts Azure ML](#) und [Googles Cloud AutoML](#)), aber auch Methoden zur Steigerung der Datenqualität wie automatisierte Erkennung von Anomalien und Pattern Recognition (bspw. [Dataiku](#) oder [Informaticas Data Quality platform](#)).

Zudem wird KI-Technologie eingesetzt, um das Mapping und Matching von Datenquellen zu automatisieren (z.B. in [Informaticas Intelligent Data Integration](#)). Auch die Erstellung von Datenmodellen (ERD) erfährt eine KI-Unterstützung durch den Einsatz von GenAI, was sich bei Softbuilders ERBuilder zeigt.

Künstliche Intelligenz in den Teilbereichen der Business Intelligence



3. Datenspeicherung

Datenspeicherung oder Datenhaltung bezeichnet im BI-Kontext üblicherweise Vorgänge und Infrastrukturen, die Daten speichern und für weitere Anwendungen bereithalten bzw. zugänglich machen, häufig in Form von Data Warehouses und Data Lakes. Im Unterschied zu allen weiteren hier besprochenen Teilbereichen der Business Intelligence, kommt Künstliche Intelligenz in der Kerntechnologie der Datenspeicherung bisher nicht vor. Daten werden, in welcher Form und Struktur auch immer, technisch gesehen weiterhin physisch oder virtuell gespeichert. Was sich jedoch stark ändert sind die Anforderungen an Skalierbarkeit und Performance durch KI-/ML-Workloads.



Dennoch spielt KI bei der Datenspeicherung eine Rolle: KI-Technologien werden eingesetzt, um das Metadaten-Management und die Metadaten-Generierung zu verbessern, also das Indizieren und Kategorisieren gespeicherter Daten auf Basis des Inhalts, der Struktur und Verwendung (bspw. [IBMs Watson Natural Language Understanding](#)). Ebenso wird KI verwendet, um das Durchsuchen großer Datenmengen durch den Einsatz menschlicher Sprache zu erleichtern (NLP), z.B. bei [ThoughtSpot](#). Zudem gibt es Ansätze, Datenintegration und Datenqualitätsprüfungen zu automatisieren (z.B. [Tredence](#)), durch KI-Technologien Query-Optimization für kosten- und performanceoptimierte Ressourcennutzung zu betreiben (z.B. [Airops](#)) und die Generierung von Schemata zu unterstützen (z.B. [TimeXTender](#)).

Des Weiteren finden sich Anwendungen, die das Ziel haben, die erfolgreiche Verwendung von Daten durch KI mittels Governance zu sichern, wie zum Beispiel [Collibras AI Governance](#) Feature.



Künstliche Intelligenz in den Teilbereichen der Business Intelligence



4. Datenaufbereitung

Datenaufbereitung beschreibt im BI-Kontext üblicherweise Vorgänge, die eine Teilmenge der zur Verfügung stehenden Daten für einen spezifischen Anwendungsfall aufbereiten. Wenn Daten beispielsweise zur Beantwortung einer analytischen Fragestellung verwendet werden sollen, wird häufig die relevante Teilmenge der Daten eines DWHs in eine möglichst unkompliziert nutzbare und möglichst performante Struktur gebracht, sowie durch eine Umbenennung der Felder für den letztendlichen Anwenderkreis leicht(er) verständlich gemacht. Je mehr Daten zur Verfügung stehen, desto wichtiger wird hierbei, den Überblick über die vorhandenen Daten und ihre Qualität zu wahren.

Eine Reihe von Tool-Anbietern verwenden KI-Technologien, um das Überblicken und Verstehen von Daten im Rahmen der Datenaufbereitung zu erleichtern. Beispielsweise nutzt Collibra im Rahmen der [Data Lineage Analysis](#) KI-Technologie, um die Bedeutung und den Inhalt von Feldern oder Tabellen zu ermitteln (gegenwärtig in der Beta-Phase).

Darüber hinaus wird KI-Technologie im Rahmen der Datenaufbereitung vor allem dazu eingesetzt, die eigentliche Aufbereitung zu erleichtern. Dazu gehören Features, die eine automatisierte und standardisierte Datenbereinigung anbieten (bspw. [Alteryx Designer Cloud](#)), die Vorschläge für Transformationen der Datenstruktur machen (z.B. Deduplikation/Data Matching, Empfehlungen durch Recommendation Engines für Analysen & Insight-Generierung) und welche durch NLP die Interaktion mit den Daten vereinfachen bzw. „menschlicher“ machen. Ein Beispiel hierfür ist etwa [Informaticas Data Quality Feature](#), mithilfe dessen übliche Schritte zur Sicherung der Datenqualität automatisch implementiert werden können, oder eine Funktionalität von [Snaplogic](#), um Datenverarbeitungsworkflows nach Befehlen in menschlicher Sprache zu erstellen. Viele Software-Lösungen aus dem BI-Kontext enthalten ähnliche, smarte Ansätze (etwa: [Qlik's Smart Data Preparation and Authoring Functions](#)), die jedoch bei näherer Betrachtung keine KI-Komponenten zu enthalten scheinen. Oft handelt es sich dabei um verhältnismäßig kleine Features einer Software, die das Arbeit mit den Daten durch praktische Hinweise erleichtern und beschleunigen sollen.

Künstliche Intelligenz in den Teilbereichen der Business Intelligence



5. Datenanalyse und Informationsdarstellung

Datenanalyse und Informationsdarstellung sind im BI-Kontext häufig der abschließende technologische Arbeitsschritt, bevor Anwenderinnen und Anwender die Ergebnisse des BI-Prozesses nutzen, um beispielsweise besser informierte, datengetriebene Entscheidungen zu fällen oder rechtlich erforderliche Berichte zu erstellen. Künstliche Intelligenz wird in verschiedenen Teilbereichen der Informationsdarstellung und -distribution eingesetzt, wobei unterschiedliche Tool-Anbieter verschiedene Ansätze und Schwerpunkte wählen.

KI-Technologie wird verwendet, um die **Interaktion** mit dem Tool bzw. mit den Daten einfacher, intuitiver und weniger technisch zu machen. Konkret bedeutet das, den Einsatz von NLP für die Erstellung von Analysen und die Formulierung von Fragen an Daten in menschlicher Sprache statt in Abfragesprachen. Beispiele sind etwa [Power BIs Q&A-Feature](#), [ThoughtSpots Feature SpotIQ](#), [Tableaus Feature Ask Data](#) (wird durch Tableau Pulse ersetzt) oder [Qliks Conversational Analytics](#), um Fragen an die Daten in menschlicher Sprache zu ermöglichen.

Des Weiteren werden KI-Technologien genutzt, um **Datenvisualisierungen und Handlungsempfehlungen** automatisiert zu erstellen oder die Erstellung durch menschliche Analysten zu unterstützen, in dem Vorschläge zur Darstellung oder Verbesserung gemacht werden und die **Exploration** von Daten und Zusammenhängen in den Daten durch Hinweise unterstützt wird. Beispiele sind das 2024 veröffentlichte Feature [Pulse von Tableau](#), um automatisiert Grafiken zu erstellen, die Trends im Zeitverlauf aufzeigen und Treiber hinter diesen Trends identifizieren können, [Power BIs Smart Data Discovery and Visualization Feature](#), mit dem ein Dashboard auf Basis einer Datenquelle automatisiert erstellt werden kann oder [Qliks Automated Insight Generation and Search-Based Discovery](#) und [Alteryx Auto Insights](#).

Außerdem werden KI-Technologien verwendet, um die Erstellung komplexerer statistischer Modelle zur **Vorhersage zukünftiger Entwicklungen** auch für Anwenderinnen und Anwender zu ermöglichen, die nicht über statistisches Fachwissen verfügen. Konkret werden z.B. Vorschläge gemacht, welche statistischen Modelle eine Vorhersage machen können und diese gegeneinander getestet, um Modelle mit dem besten Fit zu finden. Beispiele hierfür sind [Alteryx' AutoML Feature](#), welches User in die Lage versetzt, verschiedene (vom Tool vorgeschlagene) Machine Learning Modelle gegeneinander zu testen, ein bestimmtes Modell auszuwählen und dieses dann zu verfeinern. Ähnlich kann das [SAS Visual Data Mining and Machine Learning](#) eingesetzt werden, um ein am besten geeignetes ML-Modell zu bestimmen und im Nachgang zu optimieren.

5 Thesen, wie KI den Business-Intelligence-Kontext weiter verändern wird



1

Kurz- und mittelfristig bleibt KI ein (mächtiges) Helferlein an der Seite der User

Bis auf weiteres werden KI-gestützte Anwendungen im BI-Kontext im Wesentlichen weiter nach präziser Anweisung durch Anwenderinnen und Anwender Lösungen entwickeln bzw. aufbauen, die geprüft und implementiert oder freigegeben werden müssen. Dieser Trend lässt sich in allen oben beschriebenen Teilbereichen der BI erkennen und wird sich unserer Erwartung nach noch stärker etablieren. Eine echte Automatisierung von BI-Prozessen wird noch eine Weile auf sich warten lassen.

2

Langfristig wird die KI Menschen im BI-Kontext Entscheidungen abnehmen

Ob, wie und wann KI-gestützte Systeme in der Lage sein werden, nach Definition eines Bedarfs eigenständig eine Lösung zu entwickeln, die inhaltlich und logisch korrekt ist und diesen Bedarf zufriedenstellen deckt, ist gegenwärtig noch nicht zu erkennen. Zudem deuten der kürzlich veröffentlichte [AI Act der EU Kommission](#) und eine [Entscheidung des Gerichtshofs der Europäischen Union über automatisiertes Credit Scoring der Schufa](#) an, dass wir erst am Beginn einer gesellschaftlichen Auseinandersetzung darüber stehen, in welchem Maße KI selbstständig Entscheidungen treffen darf. Gleichzeitig wäre es verwunderlich, wenn nicht ein Teil der zahllosen, täglich zu fallenden Entscheidungen innerhalb der BI-Sphäre vollständig automatisiert würde.

3

Den größten Impact wird die KI im DWH entfalten

Das Speichern, Integrieren und Aufbereiten von Daten sind im BI-Kontext nicht nur die zentralsten und komplexesten Schritte, sondern auch die aufwändigsten und damit teuersten Vorgänge. Es wird mutmaßlich noch ein bisschen dauern, bis KI in diesen Bereichen deutlich Fuß fasst, da der Kontext komplex ist und die Auswirkungen von Fehlern weitreichend wären. Sobald KI jedoch von den Herstellern in diese Bereiche integriert werden wird, werden Organisationen massiv profitieren, weil sie schneller und qualitativ besser agieren können.

5 Thesen, wie KI den Business-Intelligence-Kontext weiter verändern wird



4

Was bis heute in den Analyse- und Visualisierungstools angekündigt oder implementiert wurde, verändert die Branche nicht wesentlich

ML-Modelle mit ein paar Klicks erstellen, Visualisierungen oder ganze Dashboards auf Zuruf aufbauen und Handlungsempfehlungen von der KI per Push aufs Handy bekommen. Was bis dato versprochen und teilweise schon in BI-Tools angeboten wird, senkt die Handlungsschwelle hier und da, revolutioniert die BI-Welt im Großen und Ganzen jedoch wenig. Weiterhin werden Data Engineers, Database Admins, BI Developer und Data Analysts den Großteil der Arbeit im BI-Kontext tun und sind darin aufgrund ihrer Erfahrung und des Business-Verständnisses besser und schneller als smarte KIs.

5

Governance als Schlüssel zur erfolgreichen Nutzung von KI in der BI

Was sich jedoch verändern müssen wird, ist die Governance im BI-Kontext. Auch kleinere KI-gestützte Anwendungen, die BI in einer Organisation nicht grundlegend verändern, können die Hürde zur Arbeit mit Daten erheblich senken. Ähnlich wie es Self-Service BI-Tools vor einigen Jahren getan haben, werden KI-gestützte Tools BI noch weiter in die Organisationen hineinragen und fachfremde Anwenderinnen und Anwender befähigen, noch mehr mit Daten zu arbeiten. Dafür wird eine gute, sich dynamisch entwickelnde Governance benötigt werden, die einerseits Freiräume lässt und andererseits kontrolliert, dass keine Fehler passieren.



DIE AUTOREN



BENJAMIN SCHALK

Principal Consultant Business Intelligence
benjamin.schalk@theinformationlab.de



MORITZ SCHIRMBÖCK

Principal Consultant Business Intelligence
moritz.schirmboeck@theinformationlab.de

Disclaimer: Für die Recherchen, die diesem Text zugrunde liegen, wurde unter anderem ChatGPT verwendet.
Für die Erstellung der Bilder wurde DALL·E verwendet.

**Sie möchten gern mehr zu unserer
Beratung und unseren Services im Bereich
Data Analytics & Business Intelligence erfahren?**

Dann kontaktieren Sie uns gern unter info@theinformationlab.de

The Information Lab Deutschland
Kaiser-Wilhelm-Straße 93 – 20355 Hamburg

theinformationlab.de
+49 40 2109 18180
info@theinformationlab.de