d-fine

KI-basiertes Wissens-management in der Industrie

Effiziente Informationssuche mit Large Language Models

analytisch. technologisch. quantitativ.

Nov 2023

Inhalt

KI-basiertes Wissensmanagement in der Industrie, November 2023

01. Einleitung

In einer immer stärker digitalisierten Wirtschaft erzeugen und nutzen Unternehmen aller Branchen im Rahmen ihrer täglichen Arbeit eine ständig wachsende Menge an Daten, während die Mittel zur Erfassung, Speicherung und (potenziellen) Analyse dieser Daten immer mehr den Charakter einer austauschbaren „Commodity“ gewinnen. Es liegt auf der Hand, dass die Nutzung dieser Daten zur Gewinnung tieferer Einblicke und zur Herbeiführung fundierter Entscheidungen, ein Schlüsselfaktor für die Erlangung von Wettbewerbsvorteilen ist. Dies erfordert allerdings sowohl Disziplin bei der Verwaltung der Rohdaten als auch robuste Lösungen und Workflows, um daraus nützliches Wissen zu generieren. In diesem Whitepaper wird untersucht, wie Unternehmen moderne Large Language Models (LLMs) wie ChatGPT nutzen können, um wertvolle Erkenntnisse aus ihrem Corporate Body of Knowledge (CBK) zu gewinnen und mit ihrem häufig ungenutzten Know-How-Pool Ressourcenengpässe zu beheben und die Effizienz ihrer Abläufe zu steigern.

02. Wissensmanagement im Unternehmen – eine konstante Herausforderung

Der CBK eines Unternehmens variiert zwischen verschiedenen Industriezweigen, aber auch innerhalb einzelner Unternehmen erheblich hinsichtlich Inhalt, Form und zugeschriebener Wichtigkeit. Die technischen Methoden der Informationsverwaltung werden jedoch nur selten mit Blick auf die künftige Nutzbarkeit ausgewählt, was häufig zu einer Anhäufung von losen Dokumenten und unzusammenhängenden Datensilos führt, die nur selten aktiv genutzt werden.

„Stellen Sie sich vor, technische Reports, die dem Wartungsteam das Leben erleichtern, wären in allen Situationen zentral zugänglich - und das in fast allen Sprachen. So trägt die Wissensbasis des Unternehmens aktiv zur Wertschöpfung bei.“

– Frederick Blumenthal, Lead Expert Generative AI

Über „geistiges Eigentum“ im engeren Sinne hinaus, besteht der CBK u.A. aus:

- Anlagedaten (Handbücher, Wartungsdokumentation)
- Produkt- und Produktionsdaten (Rezepturen / Prozessrouten, Anlagenkonfiguration, Best Practices, Zeitpläne, Lastinformationen, Fristen...)
- Governance-Daten (Backoffice & IT-Prozessdokumentation, Sicherheits- & Compliance-Richtlinien, Trainingsunterlagen und Arbeitsanleitungen
- Lieferantendaten (Vertragsinformationen, Verfügbarkeit, Qualitätsprobleme)
- Kundendaten (Metadaten, Beschwerden, Feature-Anfragen)[1]

In Gesprächen mit Kunden aus Finanz- und Energiesektor, Gesundheitswesen, Produktion und Fertigung sowie anderen Bereichen haben wir eine Reihe von universellen Fragen identifiziert, die sich aus einer derart vielfältigen Wissensbasis ergeben:

1 Wir betrachten hier dediziert NICHT die Daten von Mitarbeitern oder persönliche Kundendaten. Für ein Infrastrukturkonzept, das speziell auf die Analyse hochsensibler Daten zugeschnitten ist, siehe z.B., https://www.eurodat.org/

- **Variabilität:** Know-how liegt verteilt auf verschiedene Systeme und in verschiedenen Formaten und Sprachen vor, und kann nicht ohne weiteres über alle Quellen hinweg kontextualisiert werden.
- **Qualität:** Veraltete Information wird nicht bereinigt oder aktualisiert – häufig deshalb, weil keine entsprechendene Prozesse oder Ressourcen existieren.
- **Verfügbarkeit:** Die Anfragenden wissen häufig nicht, dass relevante Informationen existieren – und wo sie sie finden können!
- **Verifizierbarkeit (in Bezug auf Datenqualität):** Der Grad, zu dem eine abgerufene Information verlässlich ist, variert.

03. LLMs in der Industrie

3.1 Ausgewählte Anwendungsbeispiele

ChatGPT und andere LLMs werden all diese Herausforderungen nicht auf magische Art und Weise lösen – insbesondere werden sie Probleme in der Datenqualität und nachlässige Datenmanagementprozesse nicht kompensieren. Wenn sie aber richtig aufgesetzt sind, vereinfachen sie einige wichtige Aufgaben:

- Konsolidierung und Kontextualisierung von Informationen aus verschiedenen Quellen, in verschiedenen Formaten und Sprachen
- Zentralisierung des Zugangs zu Daten ohne die Notwendigkeit, die präzise Quelle zu kennen
- Informationsabfrage in natürlicher Sprache mit substanziell reduzierter Zugangsbarriere, schnelleren Resultate und besserer Nutzerakzeptanz
- Substantiierung von Suchergebnissen durch Referenzierung von Originaldaten mit im Ergebnis gesteigertem Vertrauen in die KI

Aus unserer Projekterfahrung haben wir eine Reihe von Anwendungsfällen für LLMs abgeleitet, die wir in allen Sektoren, insbesondere aber für Industriekunden, für nützlich und realistisch umsetzbar halten. In Abbildung 1 ordnen wir diese Anwendungsfälle in einer (ordinalen) Matrix an, die unsere Einschätzung ihrer Durchführbarkeit und ihres Mehrwerts angibt, wobei beide Metriken von Fall zu Fall variieren können und insbesondere von der Verfügbarkeit von Daten abhängen[2].

3.2 Typische Herausforderungen

Weder sind LLMs ein Heilmittel für alle Datenmanagementthemen, noch sind sie ohne Aufwand und Kosten zu implementieren. Die zu überwindenen Hürden sind dabei technischer Natur oder beziehen sich auf die Verwaltung und die Akzeptanz des "digitalen Kollegen".

Technische Herausforderungen sind von vergleichbar Art wie bei anderen IT-Systemen:

- Um das Traing von LLMs und das anschließende Durchsuchen ihrer Inhalte zu ermöglichen, müssen Datenquellen durch geeignete Schnittstellen angebunden werden – und natürlich steigen gegenseitige Abhängigkeiten, Aufwand und Komplexität, sobald mehrere Quellen benötigt werden.

2 Generative Nutzungsszenarien wie Programmierhilfen oder technische Design-Assistenten weisen technische (und teils sicherheitsrelevante) Besonderheiten auf, auf die wir im Rahmen dieses Whitepapers nicht eingehen.

- Um das System auf dem neuesten Stand zu halten, ist ein ausgereifter ML-Ops-Prozess erforderlich, der den gesamten Lebenszyklus des Modells abdeckt, von Training und Test über die Bereitstellung und Wartung bis hin zur Nutzung und Stilllegung.
- Am wichtigsten – und oft auch am aufwendigsten ist die Sicherstellung der Datenqualität während Training und Nutzungsphase, damit das System dauerhaft nützlich bleibt.

Unter Governance-Gesichtspunkten sind drei Dinge von Bedeutung:

- eine klare Policy, welche Referenzarchitektur verwendet werden soll und wie ein LLM auf On-Prem- oder Cloud-basierten Unternehmens-IT-Strukturen gehostet werden soll
- eine sehr bewusst gefällte Entscheidung, welche Daten im Einklang mit den Datenschutzbestimmungen für das Training und die Suche verwendet werden sollen und welche NICHT
- ein Konzept für die Vergabe und Verwaltung von Berechtigungen zur Suche nach einer bestimmten Information

Folgende Maßnahmen gewährleisten die Nutzeraktzeptanz:

- Nutzertraining und klare Kommunikation hinsichtlich der Möglichkeiten und Grenzen der neuen Assistenzsysteme
- Richtlinien für deren Nutzung
- Möglichkeiten zur Verifizierung der vorgeschlagenen Antwort des smarten Suchassistenten

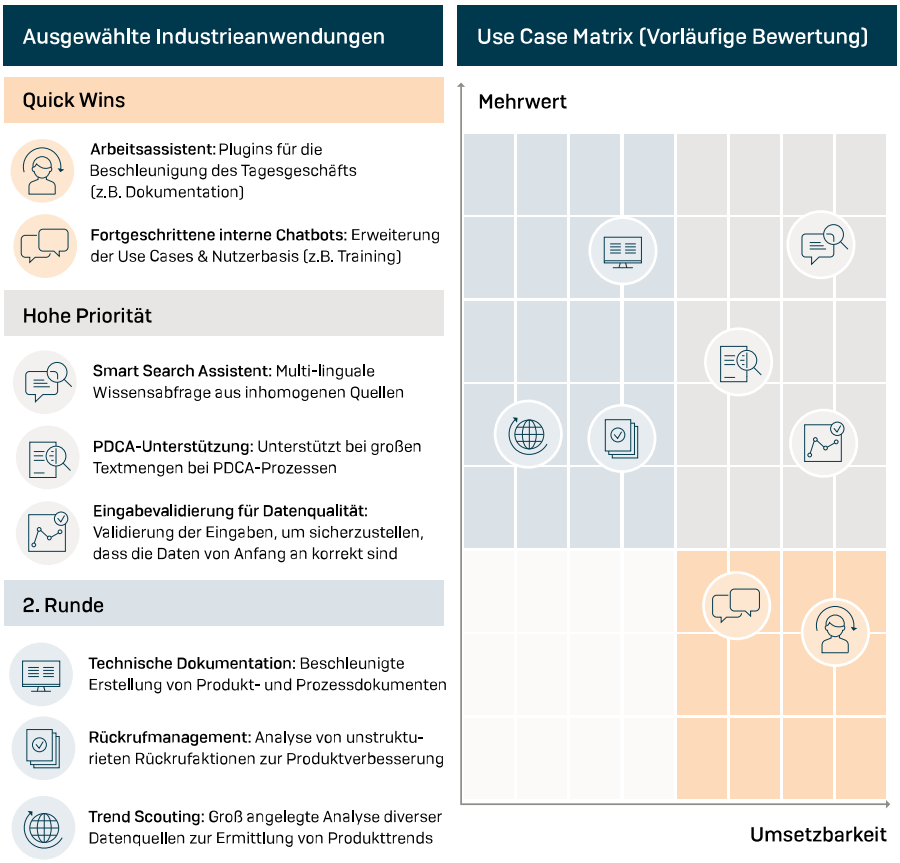

Abb. 1: Bewertungsmatrix für verschiedene LLM-Anwendungsfälle in der Industrie.

04.

Beispiel: Suche in technischer Literatur

Um für unsere Kunden eine Testumgebung bereitzustellen und ihnen eine praktische Nutzererfahrung zu ermöglichen, haben wir ein einfaches Demonstrationssystem eingerichtet, das in der Lage ist, Unternehmensinformationen in fast jeder gesprochenen Sprache zu suchen. Konkret beantwortet es Fragen zu einer Reihe von benutzerdefinierten Dokumenten, die zuvor in einem Microsoft Azure Search-Index gespeichert wurden. Das System erfordert kein Training oder Hosting eines eigenen LLM und basiert auf ChatGPT 4. Abbildung 2 zeigt ein Beispiel, bei dem wir das Open-Access-Buch "Elements of Robotics[3]" von Ben-Ari und Mondana, das unter einer Creative-Commons-Lizenz[4] verfügbar ist, nach einer Erklärung des Konzepts der Odometrie - einer gängigen Navigationsmethode für kostengünstige Roboter - befragt haben.

Trotz seines "Out-of-the-Box"-Charakters liefert das System mit geringfügigem Prompt-Engineering korrekte Antworten und in den meisten Fällen auch korrekte Verweise auf deren Herkunft in den gesuchten Dokumenten. Letzteres ist insofern wichtig, als kritische Nutzer die präsentierten Ergebnisse hinterfragen MÜSSEN und den Antworten der KI nicht wahllos vertrauen dürfen.

„LLM tragen dazu bei, den Wissensbestand eines Unternehmens zugänglich und praktisch nutzbar zu machen, z. B. durch die Erstellung von Anleitungen für Wartungsarbeiten auf dem Shop Floor.“

– Tassilo Christ, Partner d-fine Industrial Solutions

3 Ben-Ari,M.and Mondala, F.: Elements of Robotics, Springer Open, 2018; https://doi.org/10.1007/978-3-319-62533-1

4 http://creativecommons.org/licenses/by/4.0/

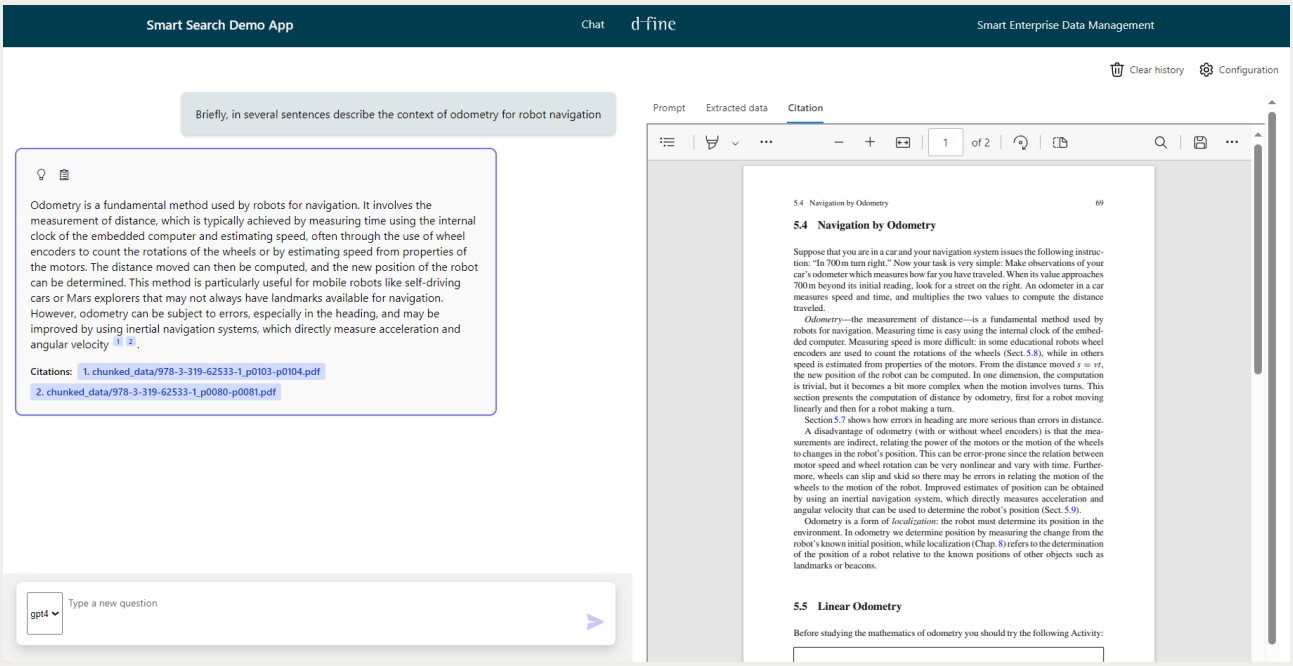

Abb. 2: Ergebnis der Suche in einem Lehrbuch nach Erläuterungen zu einem Navigationskonzept. Das System präsentiert eine Zusammenfassung mit Informationen aus mehreren Textabschnitten.

05.

Was ist Ihr Anwendungsfall – Wie können wir helfen?

KI-Lösungen von der Stange, die die “typischen” Anwendungsfälle abdecken und nur “ein wenig” Konfiguration benötigen, führen selten zu einem langfristigen Wettbewerbsvorteil. Unserer Erfahrung nach erfordert die Erzielung nachhaltiger Vorteile vielmehr einen maßgeschneiderten und fokussierten Ansatz, der tiefgreifendes Fachwissen auf der Kundenseite mit dem technologischen und methodischen Wissen eines Spezialistenteams kombiniert. Dies gilt umso mehr für KI-Themen, da die zugrundeliegende Technologie einen fundierten mathematischen Hintergrund erfordert, um einen messbaren Mehrwert zu liefern.

d-fine hat eine umfangreiche Erfolgsbilanz bei der gemeinsamen Entwicklung von maßgeschneiderten LLM-Lösungen mit Kunden aus dem Finanzwesen, der Industrie und anderen Branchen. Wir gehen weit über die häufig anzutreffenden "KI-PoCs" hinaus und unterstützen unsere Kunden dabei, Lösungen zur Einsatzsreife zu bringen. Abbildung 3 zeigt eine Auswahl erfolgreicher KI-Projekte, bei denen wir NLP-Systeme entwickelt haben, die auf die spezifischen Anforderungen unserer Kunden zugeschnitten sind.

Industrieunternehmen

Shop-Floor-Assistent: Smart Search Tool basierend auf MS Azure OpenAI-Services zur vereinfachten Informationssuche in technischen Dokumenten

Medizinische Forschung

Informations-Extraktion: Automatisierte, textbasierte Datensuche zur Gerätesicherheit zur Unterstützung von Expertenaufgaben bei einem Hersteller med. Geräte

Cloud Provider

Kunden-Chatbot: Daten-Pipelines zur Erweiterung von Chat GPT mit internen Dokumenten; Backend / API Entwicklung und Prompt Engineering

IT-Betrieb

Qualitäts-Screening von Dokumenten: Automatische Analyse und Überwachung von interner IT-Dokumentation hinsichtlich Vollständigkeit und Qualität

Öffentlicher Sektor

Wissenschaftliche Textanalyse: Forschungsprojekt zur Informationssuche bei der European Food Safety Authority; Integration der GPT-3-API and Prompt Engineering

Bankenregulierung

KI-basierte Suche: LLM-gestützte Suchplatt-form zur Informationsgewinnung aus einem großen Dokumentenkorpus aus dem Bereich Bankenregulierung

Abb. 3: Ausgewählte Referenzprojekte zu den Themen Natural Language Processing und Large Language Models in verschiedenen Industrien.

06.

Legen Sie los mit generativer KI

Um das Potenzial und die Grenzen generativer KI zu verstehen - und um Enttäuschungen zu vermeiden - müssen zunächst Anwendungsfälle für ihren Einsatz sorgfältig identifiziert und bewertet werden. Häufig beginnen unsere Projekte daher mit einer dedizierten Ideenfindungsphase für die Anwendungsszenarien oder mit einem Workshop, in den wir unsere breite, branchenübergreifende Erfahrung einbringen. Wir helfen unseren Kunden, Anwendungen für LLMs zu identifizieren und die technische Machbarkeit, die Kosten für die Implementierung und die potenziellen Auswirkungen zu bewerten. Um den Wert generativer KI zu verifizieren und eine steile Lernkurve zu ermöglichen, empfehlen wir, einen ausgewählten Pilotanwendungsfall mit hohem Mehrwert zu implementieren und die Lösung anschließend in einem agilen Vorgehensmodell auf andere Prozesse zu erweitern.

Unser Angebot: Wir sind sehr gespannt, welche Anwendungsfälle Sie in Ihrem Unternehmen finden! Dazu demonstrieren wir im Rahmen eines Workshops, welche Ergebniss unser Smart Search Tool-Stack mit Ihren Daten erzielt und diskutieren auf dieser Basis, welchen Mehrwert ChatGPT & Co. für Sie haben.

Autoren

Frederick Blumenthal
Manager und Expert NLP Solutions
d-fine GmbH, München
Frederick.Blumenthal@d-fine.com

Dr. Ulf Menzler
Senior Manager und Expert MLOps
d-fine GmbH, Düsseldorf
Ulf.Menzler@d-fine.com

Dr. Fedor Petrov
Manager und Expert Manufacturing Systems
d-fine GmbH, Frankfurt
Fedor.Petrov@d-fine.com

Dr. Tassilo Christ
Partner und Head of Industrial Solutions
d-fine GmbH, München
Tassilo.Christ@d-fine.com

d-fine

Berlin
d-fine GmbH
Kranzler Eck
Kurfürstendamm 21
10719 Berlin
Deutschland
berlin@d-fine.de

Düsseldorf
d-fine GmbH
Dreischeibenhaus 1
40211 Düsseldorf
Deutschland
duesseldorf@d-fine.de

Frankfurt
d-fine GmbH
An der Hauptwache 7
60313 Frankfurt
Deutschland
frankfurt@d-fine.de

Hamburg
d-fine GmbH
Am Sandtorpark 6
20457 Hamburg
Deutschland
hamburg@d-fine.de

London
d-fine Ltd
14 Aldermanbury Square
London, EC2V 7HR
United Kingdom
london@d-fine.co.uk

Mailand
d-fine s.r.l.
Via Giuseppe Mengoni 4
20121 Milano MI
Italien
milano@d-fine.com

München
d-fine GmbH
Bavariafilmplatz 8
82031 Grünwald
Deutschland
muenchen@d-fine.de

Stockholm
d-fine AB
Nybrogatan 17
114 39 Stockholm
Schweden
stockholm@d-fine.se

Utrecht
d-fine BV
Stadsplateau 7
3521 AZ Utrecht
Niederlande
utrecht@d-fine.nl

Wien
d-fine Austria GmbH
Seilerstätte 13
1010 Wien
Österreich
wien@d-fine.at

Zürich
d-fine AG
Brandschenkestrasse 150
8002 Zürich
Schweiz
zuerich@d-fine.ch