

Generative Künstliche Intelligenz in ERP-Systemen Aktueller Stand und Ausblick

Stefan Wittenberg, Birte Malzahn, Christina Kratsch,
Rene Marienburg

Die Fähigkeit generativer Künstlicher Intelligenz (Gen-KI), über die Nutzung leistungsfähiger Large-Language-Modelle (LLMs) menschliche Sprache zu verstehen, komplexe Fragen zu beantworten, Programmcode zu schreiben oder Wissen zusammenzufassen, bietet neue Möglichkeiten für die Automatisierung von Geschäftsprozessen [1]. Die zunehmende Multimodalität der LLMs im Generieren und Interpretieren von Audio-, Bild- und Videodaten eröffnet dabei auch vielversprechende Anwendungsgebiete jenseits von reinen textbasierten Aufgaben [2].

Studien sagen durch den Einsatz von Gen-KI deutliche Produktivitätssteigerungen insbesondere in Tätigkeiten von hochqualifizierten „Wissensarbeitern“ im „White-Collar-Bereich“ voraus, die bisher aufgrund eines geringeren Standardisierungsgrades der Aufgaben im Vergleich zu „Blue-Collar-Tätigkeiten“ typischerweise nicht Gegenstand von Automatisierungsbemühungen waren [3, 4, 5].

Gen-KI wird aktuell im Unternehmenskontext vorwiegend in „Office-nahen“ Bereichen eingesetzt. Beispiele sind die Erstellung von E-Mails, Textdokumenten, Präsentationen, die Zusammenfassung von Meetings oder die Generierung von Marketinginhalten [6]. Durch die Integration von Gen-KI in die 365-Office-Umgebung von Microsoft über den „MS 365 Copilot“ und MS Teams können bereits zahlreiche Unternehmen die Funktionen generativer KI standardisiert „out-of-the-box“ nutzen [7].

Retrieval-Augmented Generation (RAG)

Aus Sicht von Unternehmen stellen Halluzinationen – das „Erfinden“ oder Verfälschen von Inhalten durch Gen-KI – ein großes Problem dar. Halluzinationen treten auf, wenn das LLM aufgrund falscher oder nicht vorhandener Trainingsdaten Inhalte „erfindet“, die nicht der Realität entsprechen. Da ein LLM typischerweise die internen Daten eines Unternehmens nicht „kennt“, weil diese nicht Inhalt der im Internet verfügbaren Trainingsdaten waren, steigt die Gefahr der Halluzinationen bei Anfragen, die spezifische Abläufe im Unternehmen betreffen. Um Halluzinationen zu vermeiden, gewinnt Retrieval-Augmented Generation (RAG) aktuell an Bedeutung. Im Rahmen von RAG werden in einem Retrieval-Prozess die Eingabe-Prompts für das LLM mit vertrauenswürdigen unternehmensinternen und externen Daten angereichert [8].

Ablauf von RAG [8]:

1. Der Nutzende gibt einen Prompt im Eingabefenster des Gen-KI-Systems ein.
2. Die Inhalte des Prompts werden im Rahmen einer semantischen Suche mit Inhalten aus internen und externen Datenbanken, Dokumenten oder Applikationen abgeglichen. Die Inhalte mit der größten semantischen Ähnlichkeit werden abgerufen und stehen nun als „Kontext“ bereit.

Lesen Sie

- welche Möglichkeiten Gen-KI für die Automatisierung von Geschäftsprozessen bietet
- wie Gen-KI nach dem Vorgehensmodell der HTW Berlin systematisch eingesetzt werden kann

3. Die Nutzeranfrage wird um den bereitgestellten Kontext erweitert und an das LLM gesendet.
4. Das LLM nutzt den Kontext und kann so eine spezifischere Antwort bereitstellen, die über die reinen Trainingsdaten des LLMs hinausgeht und zudem tagesaktuelle Informationen berücksichtigen kann. Ein Verweis auf die genutzten Quellen ermöglicht den Nutzenden eine Validierung der Ergebnisse.

Bild 1 stellt den Ablauf grafisch dar. Als Speichermedium für unstrukturierte Daten eines Unternehmens (z. B. Word, E-Mails, PDFs) für RAG kommen vorwiegend Vektordatenbanken zum Einsatz, in denen einzelne Abschnitte der Informationen („Chunks“) in mehrdimensionalen Vektoren mit semantischen Bedeutungen abgespeichert werden. Diese Form der Abspeicherung – das „Embedding“ – ermöglicht auf Basis des Eingabeprompts die Suche nach semantisch ähnlichen Inhalten und unterscheidet sich daher deutlich von einer reinen Schlagwortsuche [9].

RAG in ERP-Systemen

Enterprise-Resource-Planning (ERP)-Systeme orchestrieren seit Jahrzehnten abteilungsübergreifende End-to-End-Prozesse und integrieren Daten in einem zentralen System. Typische Anwendungsfälle sind etwa die Abbildung der Lead-to-Order-, Order-to-Cash-, Purchase-to-Pay- oder Issue-to-Solution-Prozesse, die jeweils mehrere Abteilungen und Module der ERP-Systeme durchlaufen. Leistungsfähige ERP-Systeme wie SAP S/4 bieten schon seit einigen Jahren KI-Funktionen wie Vorhersagemodelle im Bereich der Sales- oder Produktionsplanung oder etwa zum Matching von Wareneingangspositionen zu Bestell- und Rechnungspositionen [10]. Diese breite Datenbasis bietet LLMs grundsätzlich eine reichhaltige Ressource für RAG. Vor diesem Hintergrund stellt sich die Frage, wie Daten aus ERP-

Systemen im Rahmen von RAG für Gen-KI nutzbar gemacht werden können.

ERP-Systeme verwalten in Form relationaler Tabellen strukturierte Informationen von Objekten wie Materialstämme, Kunden- und Lieferantenstämme und Bewegungsdaten (z. B. Kundenaufträge, Bestellungen, Wareneingänge) [11]. Die hersteller-spezifischen, teilweise „kryptischen“ Bezeichnungen von Tabellen und Feldern in ERP-Systemen sind für LLMs nicht intuitiv verständlich, auch wenn grundsätzliche Strukturen bekannter ERP-Anbieter wie SAP in den Trainingsdaten führender LLMs wie GPT (Open AI), Llama (Meta) oder Claude (Anthropic) enthalten sind. Spätestens jedoch dann, wenn durch unternehmensspezifische Anpassungen der ERP-Systeme unternehmensspezifische Tabellen oder Feldbezeichnungen angepasst oder ergänzt werden, hat ein LLM Schwierigkeiten mit dem Verständnis der Inhalte. LLMs sind zudem typischerweise auf die Verarbeitung von natürlicher Sprache trainiert und daher nur begrenzt in der Lage, mit Informationen in Tabellenform umzugehen [12].

Die strukturierten Daten in ERP-Systemen werden im Arbeitsalltag häufig mit unstrukturierten Daten aus Dokumentenmanagementsystemen oder Office-Dokumenten verknüpft. Beispiele sind etwa Materialspezifikationen und Lieferantenangebote zu einem Bestellvorgang. Häufig sind diese ergänzenden unstrukturierten Daten notwendig, um einen Geschäftsvorfall inhaltlich zu „verstehen“.

Ein aktuell vielfach diskutierter Ansatz, um strukturierte und unstrukturierte Daten einem LLM als Kontext für RAG aufzubereiten, sind Knowledge-Graphen. Knowledge-Graphen ermöglichen es in diesem als „Graph RAG“ bezeichneten Verfahren, die oben genannten strukturierten und unstrukturierten Objekte als Entitäten mit Attributen sowie deren Beziehungen untereinander als Kanten explizit zu modellieren. Die Extraktion der Knowledge-Graphen in Textform kann dann als Kontext im RAG-Ablauf bereitgestellt werden [13, 14].

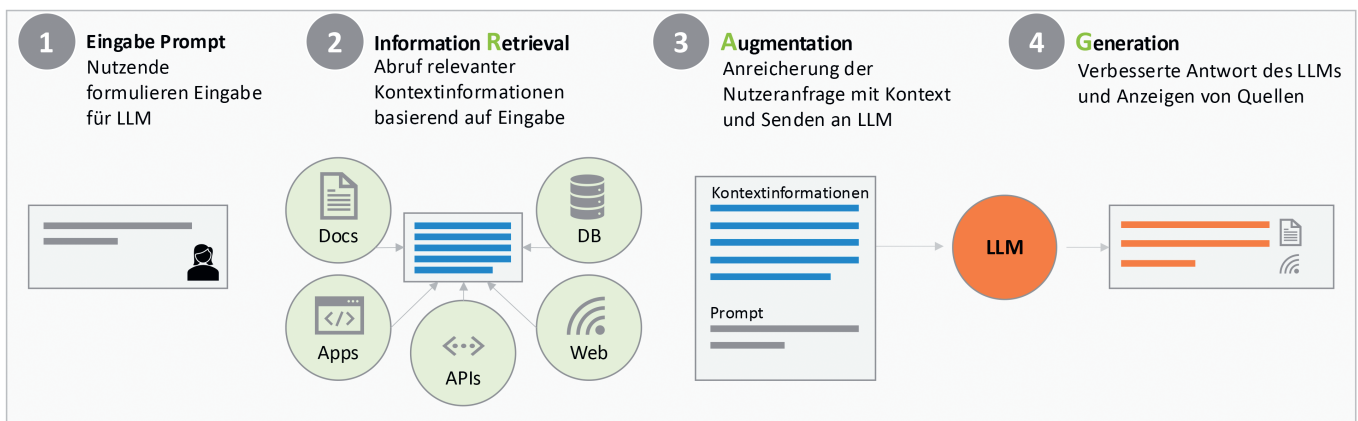


Bild 1: Ablauf Retrieval-Augmented Generation (RAG).

Gen-KI-gestützte Wartung in ERP-Systemen: NBB Netzgesellschaft Berlin-Brandenburg (NBB)

Die NBB als Tochtergesellschaft des GASAG-Konzerns verantwortet den Betrieb von Gas- und Wasserstoffnetzen in Berlin und Brandenburg. Mit Unterstützung der KI-Werkstatt der Hochschule für Technik und Wirtschaft Berlin und in Kooperation mit der SAP SE wurde ein gemeinsames Pilotprojekt zur Nutzung generativer KI im S/4-HANA-System der NBB gestartet. Als Pilotbereich wurde der Bereich der Instandhaltung ausgewählt.

Für Zwecke der Wartung und Reparatur sind im SAP-Modul Plant Maintenance (PM) der NBB strukturierte Daten als Stammdaten hinterlegt (z. B. technische Plätze, Equipments, Ersatzteile). Ergänzend zu den Stammdaten in SAP existieren auch unstrukturierte Stammdaten (z. B. CAD-Zeichnungen, Fotos der Anlagen, Spezifikationen von Ersatzteilen, Wartungshinweise), die u. a. im SAP-Dokumentenmanagementsystem (DMS) mit den entsprechenden Objekten verknüpft sind.

Ist eine Wartung oder Reparatur durchzuführen, wird eine Instandhaltungsmeldung und anschließend ein Instandhaltungsauftrag angelegt, auf den eine Rückmeldung durchgeführt wird. Beide Objekte werden als strukturierte Bewegungsdaten in Tabellenform abgespeichert. Zusätzliche Checklisten, Protokolle oder Nachweise der Durchführung von Wartungsarbeiten durch Dienstleister werden wiederum als zugehörige unstrukturierte Daten im SAP-DMS abgespeichert. Bild 2 stellt die verschiedenen Datentypen exemplarisch dar.

In einem Ideation-Workshop wurden mit Mitarbeitenden der NBB folgende Ideen für Gen-KI-Use-Cases entwickelt:

Use-Case-Beispiel 1: Wartungsvorbereitung

Für die sorgfältige Vorbereitung einer Wartung einer technischen Anlage (in SAP: technische Plätze und Equipments) gestaltet sich die Sichtung aller relevanten Daten in SAP sowie begleitender Dokumente sehr zeitaufwendig. Um etwa zu ermitteln, ob es im Rahmen der letzten Wartung zu Auffälligkeiten oder Fehlern gekommen ist, müssen u. a. Instandhaltungsaufträge, Instandhaltungsmeldungen und -nachweise von Wartungsdienstleistern manuell kontrolliert werden. Hieraus entstand der

Use-Case-Ansatz, von einem LLM einen komprimierten Bericht zur Wartungshistorie erzeugen zu lassen, der sowohl strukturierte als auch unstrukturierte Informationen einbezieht.

Use-Case-Beispiel 2: Überprüfung von Stammdaten

Die SAP-Stammdaten enthalten Angaben über Ort, Länge und Spezifikationen der technischen Plätze und Equipments. So haben Gasrohre etwa spezifizierte Längen und Durchmesser. Angaben auf CAD-Plänen oder in Spezifikationen können durch Übertragungsfehler oder Änderungen in Konstruktionen, die im SAP-System möglicherweise nicht angepasst wurden, von den SAP-Stammdaten abweichen. Im Sinne der Sicherung der Stammdatenqualität ist vorgesehen, über ein LLM die unstrukturierten Dokumente mit den SAP-Stammdaten abzugleichen.

Use-Case-Beispiel 3: Auswertung von Wartungsprotokollen

Bei der Auswertung von Wartungsprotokollen sollen über das LLM Empfehlungen oder sogar automatische Aktionen ausgelöst werden – z. B. die Eröffnung einer neuen Instandhaltungsmeldung bei Feststellen von Mängeln im Wartungsprozess oder Schließen eines Instandhaltungsauftrages, wenn keine Auffälligkeiten sichtbar sind. Weiterhin können Wartungsprotokolle sowie Nachweise und/oder Rechnungen von Wartungsdienstleistern auf Plausibilität geprüft werden, z. B. auf Abweichungen zwischen abgerechneter Leistung und Angaben in den Stammdaten. Aktuell befindet sich das Pilotprojekt in der Detailplanung zur technischen Umsetzung. Es ist ein Einsatz der Business Technology Plattform (BTP) von SAP als Connector zwischen SAP S/4 und einem LLM in einem europäischen Datacenter vorgesehen, um Datenschutz und -sicherheit zu gewährleisten. Neben dem Einsatz einer Vektordatenbank soll auch der Mehrwert einer Knowledge-Graphen-Datenbank geprüft werden.

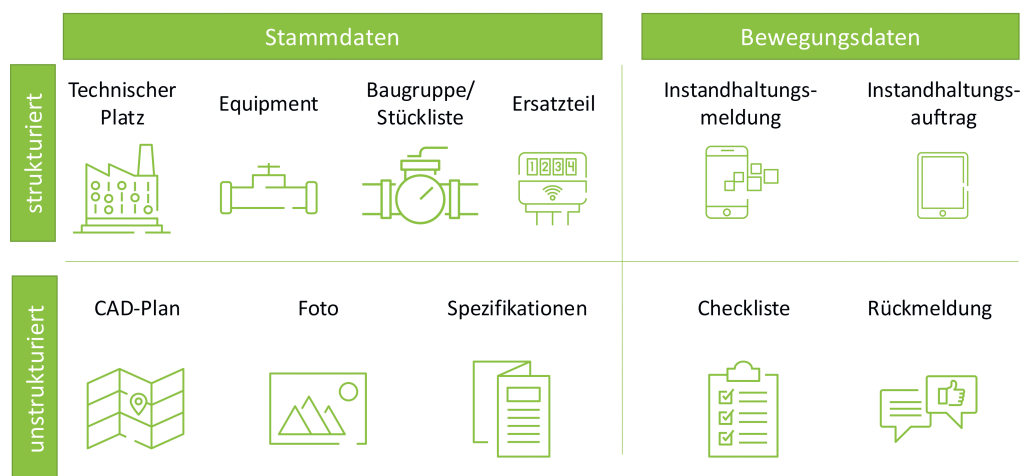


Bild 2: Strukturierte und unstrukturierte Stamm- und Bewegungsdaten.

Vorgehensmodell zur Einführung generativer KI

Das Beispiel der NBB verdeutlicht, dass Gen-KI nicht nur einfache, generische Office-Anwendungen, sondern auch anspruchsvolle, spezifische Geschäftsprozesse unterstützen kann. Die Integration generativer KI in Geschäftsprozesse und ERP-Systeme ist jedoch eine komplexe Aufgabe, die weit über technologische Aspekte hinausgeht. So sind etwa die Auswahl und Priorisierung von Use Cases auf Basis unternehmerischer Ziele, die Schulung von Mitarbeitenden und die gezielte Auseinandersetzung mit Sicherheit- und Governance-Fragestellungen zentrale Bausteine für den Erfolg von Gen-KI-Projekten. Hier empfehlen die Autoren ein systematisches Vorgehen zum Einsatz generativer KI, wie es etwa im Vorgehensmodell zur Einführung generativer KI der HTW Berlin beschrieben ist. In diesem Modell werden, ausgehend von einer Top-Down-Aufnahme strategischer Ziele für Gen-KI aus Managementsicht, über die Ausarbeitung von Use-Case-Ideen und Soll-Prozessen in Bottom-Up-Workshops aus Mitarbeitendensicht, die LLM-Modellauswahl, die Pilotierung bis hin zum späteren Betrieb wichtige Fragestellungen erörtert. Im Schritt vier des Vorgehensmodells (Technologieauswahl) stellt sich im Kontext der ERP-Systeme beispielsweise die Frage, ob die integrierten Gen-KI-Funktionen von ERP-Herstellern „out-of-the-box“ genutzt werden sollten oder lediglich die Daten aus dem ERP „abgezogen“ und über Schnittstellen mit LLM-Systemen externer Anbieter bzw. On-Premises-betriebener Open-Source-LLMs verknüpft werden sollten. Die integrierte Lösung würde dem Anspruch der ERP-Systeme zur Bereitstellung unter-

nehmensweiter Lösungen gerecht und ex-ante den Aufbau neuer, zusätzlicher Gen-KI-Systeme vermeiden, die wiederum zu neuen Schnittstellen und Doppelbetrieb von Systemen führen würden. Dagegen könnten jedoch das begrenzte Angebot der durch ERP-Hersteller zur Verfügung gestellten LLMs und die erhöhten Kosten der Nutzung sprechen, da die Hersteller von ERP-Systemen häufig auch die LLM-Modelle und Rechenzentrumsleistungen von US-amerikanischen Herstellern „einkaufen“ und diese dann mit möglicherweise hohen Aufschlägen „weiterverkaufen“.

Diskussion und Ausblick

Der Markt der bereits integriert verfügbaren LLM-Lösungen von ERP-Herstellern befindet sich derzeit noch am Anfang der Entwicklung. SAP als Vorreiter stellt hier bereits in der S/4 Public Cloud den produktübergreifenden Gen-KI-Assistenten „Joule“ bereit, verfügt über den SAP-Knowledge-Graphen und arbeitet an der Bereitstellung von LLM-Modellen, die auch im Umgang mit strukturierten Daten qualitativ hochwertige Ergebnisse liefern. Während SAP bereits zahlreiche Gen-KI-Funktionen produktiv anbietet, zeigte eine erste Analyse der Webseiten von 168 in Deutschland verfügbaren mittelständischen ERP-Lösungen ein ernüchterndes Bild. Dort wurden zum Stand 01/2025 nur bei 12 ERP-Lösungen konkrete Gen-KI-Funktionen identifiziert. Diese Analyse ist nur ein Schlaglicht, da sie rein auf der Analyse der Webseiten der Hersteller und keiner Befragung der Anbieter bzgl. interner Gen-KI-Roadmaps beruhte.

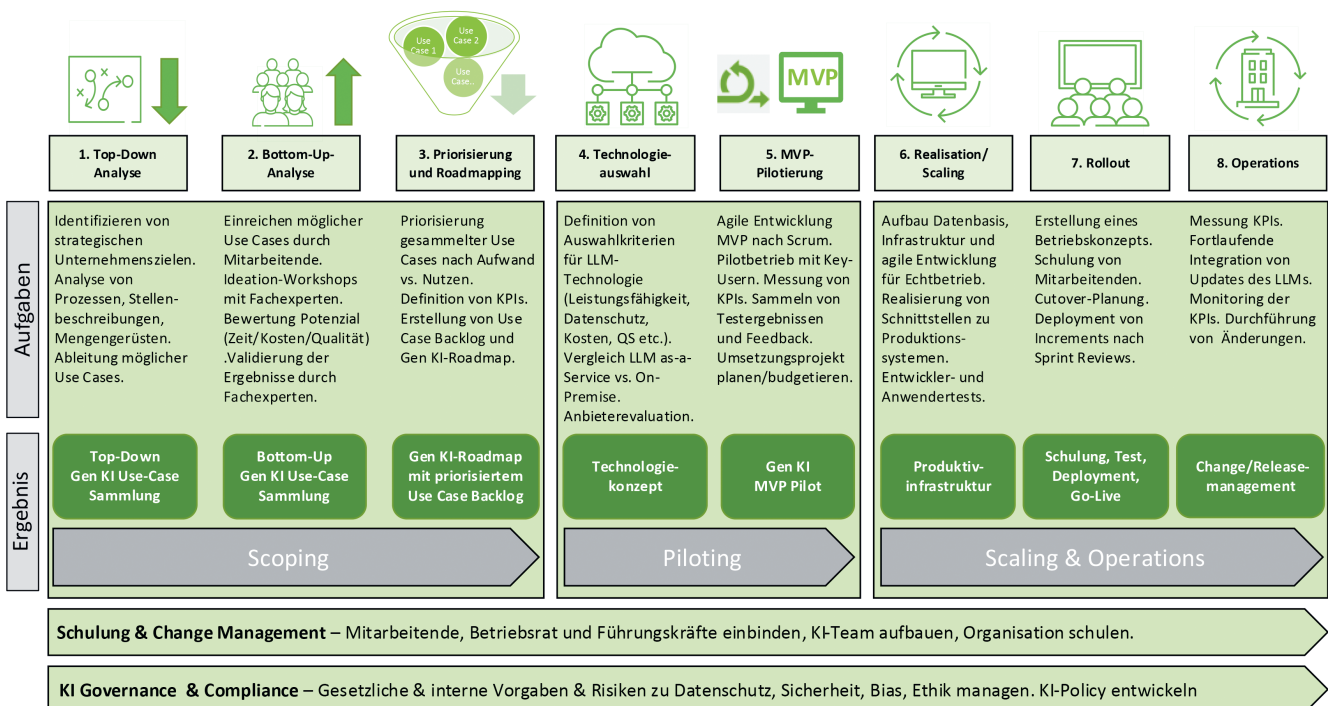


Bild 3: Vorgehensmodell zur Einführung generativer KI [15].

Allerdings stellt sich die Frage, ob mittelständische ERP-Anbieter, die oft seit Jahrzehnten die Produkte inkrementell weiterentwickelt haben, in kurzer Zeit aufholen und auch mit hoher Geschwindigkeit neuartige Gen-KI-Innovationen in die Produkte integrieren können. Aktuell zeigt sich etwa in Gen-KI ein Trend zur „Agenten- und Reasoningfähigkeit“ [16] – also einerseits die Fähigkeit, auf eine konkrete Rolle oder Persona zu fokussieren und diese autonom auszuführen, und andererseits eigenständig komplexe Lösungsstrategien für neue Problemstellungen zu finden. Die Geschwindigkeit des Wandels „mitzugehen“, wird eine Herausforderung für mittelständische ERP-Hersteller mit sehr begrenzten Kapazitäten an KI-Experten darstellen. Die Fähigkeit, neueste Gen-KI-Innovationen als ERP-Hersteller mitanzubieten, könnte jedoch zukünftig ein wichtiges Entscheidungskriterium bei der Auswahl eines ERP-Anbieters werden.

Literatur

[1] Feuerriegel, Stefan/Jochen Hartmann/Christian Janiesch/Patrick Zschech: Generative AI, in: Business & Information Systems Engineering, Bd. 66, Nr. 1, 12.09.2023, [online] doi: 10.1007/s12599-023-00834-7, S. 111–126.

Prof. Dr. Stefan Wittenberg ist seit 2016 Professor für ERP-Systeme und Geschäftsprozessmanagement an der Hochschule für Technik und Wirtschaft Berlin. Zuvor war er als Leiter des Auftragszentrums und als SAP-Projektleiter in der Bundesdruckerei GmbH tätig. Seine fachlichen Schwerpunkte liegen u. a. in der Integration generativer KI in betriebliche Geschäftsprozesse und ERP-Systeme.

Prof. Dr. Birte Malzahn ist seit 2012 Professorin für Wirtschaftsinformatik an der HTW Berlin. Zuvor war sie mehrere Jahre als SAP-Beraterin tätig. Ihre fachlichen Schwerpunkte liegen u. a. in den Bereichen Geschäftsprozessmanagement und betriebliche Anwendungssysteme.

Prof. Dr. Christina Kratsch ist seit 2022 Professorin für Künstliche Intelligenz und Software-Engineering an der HTW Berlin. Zuvor war sie als Consultant und Projektleiterin in den Bereichen Data Science, KI und Strategieentwicklung tätig. Ihre fachlichen Schwerpunkte liegen u. a. in Anwendungsfällen und Operationalisierung von generativer KI.

Rene Marienburg leitet in der NBB Netzgesellschaft Berlin-Brandenburg mbH & Co. KG die Abteilung „Netzsupport“. In dieser Funktion ist er vollumfänglich für die Betreuung der spezifischen IT-Systeme eines Netzbetreibers verantwortlich. Dazu zählen unter anderen SAP-PM/SAP IS-U, mobiles Workforce-Management, Geografische Informationssysteme und Elektronische Dokumentenmanagementsysteme. Er hat langjährige Erfahrungen als SAP-Consultant und ist vor mehr als 10 Jahren zur NBB gewechselt.

Stefan Wittenberg
Hochschule für Technik und Wirtschaft Berlin
Treskowallee 8, 10318 Berlin
E-Mail: stefan.wittenberg@htw-berlin.de
<https://kiwerkstatt.f2.htw-berlin.de/projekte/gen-ai>

- [2] Akhtar, Zarif Bin: Unveiling the Evolution of Generative AI (GAI): A Comprehensive and Investigative Analysis toward LLM models (2021–2024) and Beyond, in: Journal of Electrical Systems and Information Technology, Bd. 11, 2024, [online] doi:10.1186/s43067-024-00145-1.
- [3] Gmyrek, Pawel/Janine Berg/David Bescond: Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality, International Labour Organization (Hrsg.), Genf, Schweiz, 2023.
- [4] Eloundou, Tyna/Sam Manning/Pamela Mishkin/Daniel Rock: GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, 2023, [online] <https://arxiv.org/abs/2303.10130>.
- [5] Chui, Michael/Eric Hazan/Roger Roberts/Alex Singla/Kate Smaje/Alex Sukharevsky/Lareina Yee/Rodney Zempel: The Economic Potential of Generative AI: The next productivity frontier, McKinsey & Company (Hrsg.), 2023.
- [6] Gozalo-Brizuela, Roberto/Eduardo C. Garrido-Merchán: A Survey of Generative AI applications, 2023, [online] <https://arxiv.org/abs/2306.02781>.
- [7] Vasilescu, Cristina/Militaru Gheorghe: Improving the Performance of Corporate Employees through the Use of Artificial Intelligence: The Case of Copilot Application, in: Proceedings of the International Conference on Business Excellence, Bd. 18, 2024, [online] doi:10.2478/pic-be-2024-0153, S. 1819–1830.
- [8] Gao, Yunfan/Yun Xiong/Xinyu Gao/Kangxiang Jia/Jinliu Pan/Yuxi Bi/Yi Dai/Jiawei Sun/Meng Wang/Haofen Wang: Retrieval-Augmented Generation for Large Language Models: A Survey, 2024, [online] <http://arxiv.org/abs/2312.10997>.
- [9] Lewis, Patrick S. H./Ethan Perez/Aleksandra Piktus/Fabio Petroni/Vladimir Karpukhin/Naman Goyal/Heinrich Küttler/Mike Lewis/Wen-Tau Yih/Tim Rocktäschel/Sebastian Riedel/Douwe Kiela: Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020, S. 9459–9474.
- [10] Sarferaz, Siar: Embedding Artificial Intelligence into ERP Software: A Conceptual View on Business AI with Examples from SAP S/4HANA, Springer Cham, 2024.
- [11] Drumm, Christian/Bernd Scheuermann/Stefan Weidner: Einstieg in SAP S/4HANA: Geschäftsprozesse, Anwendungen, Zusammenhänge – Erklärt am Beispielunternehmen Global Bike, Bonn: Rheinwerk Verlag GmbH, 2023.
- [12] Fang, Xi/Weijie Xu/Fiona Anting Tan/Jiani Zhang/Ziqing Hu/Yanjun Qi/Scott Nickleach/Diego Socolinsky/Srinivasan Sengamedu/Christos Faloutsos: Large Language Models (LLMs) on Tabular Data: Prediction, Generation, and Understanding – A Survey, 2024, [online] <https://arxiv.org/abs/2402.17944>.
- [13] Edge, Darren/Ha Trinh/Newman Cheng/Joshua Bradley/Alex Chao/Apurva Mody/Steven Truitt/Jonathan Larson: From Local to Global: A Graph RAG Approach to Query-Focused Summarization, 2024, [online] <https://arxiv.org/abs/2404.16130>.
- [14] Pan, Shirui/Linhao Luo/Yufei Wang/Chen Chen/Jiapu Wang/Xindong Wu: Unifying Large Language Models and Knowledge Graphs: A Roadmap, in: IEEE Transactions On Knowledge And Data Engineering, Bd. 36, 2024, [online] <https://doi.org/10.1109/tkde.2024.3352100>, S. 3580–3599.
- [15] KI Werkstatt der HTW Berlin: o. D., [online] <https://kiwerkstatt.f2.htw-berlin.de/>
- [16] OpenAI: Introducing Deep Research: o. D., [online] <https://openai.com/index/introducing-deep-research/> (abgerufen am 02.02.2025).

Stichwörter:

Generative KI, Retrieval-Augmented Generation, Embedding, Knowledge Graph, KI-Agent