Nr. 3 | August 2025 | ISSN 2366-7915 | www.doag.org

Red Stack Magazin

DOAG SOUG swiss oracle user group AOUG AUSTRIAN ORACLE USER GROUP

Aus der Praxis

Wie Large Language Models Arbeitsprozesse transformieren

Im Interview

Stefan Latuski, CIO, Bundesagentur für Arbeit; CEO, IT-Systemhaus der BA

KI

Datenanalyse mit ChatGPT im Praxistest

e3mag.com

DEUTSCH

Information und Bildungsarbeit von und für die SAP®-Community

ENGLISH

The global independent platform for the SAP® community

ESPAÑOL

La plataforma global e independiente para la comunidad SAP®

FRANÇAIS

La plateforme indépendante mondiale de la communauté SAP®

Oliver Szymanski
DOAG Vorstand KI, iJUG, JUG Nürnberg

Liebe Mitglieder, liebe Leserinnen und Leser,

Künstliche Intelligenz ist gekommen, um zu bleiben. Was lange Vision war, ist heute Bestandteil unserer Werkzeuge, Arbeitsprozesse – und Diskussionen auf nahezu jeder IT-Konferenz. Wer die Entwicklungen der letzten Monate beobachtet hat, weiß: Es geht nicht mehr um „ob", sondern um „wie" und „wofür".

Mit dieser Spezialausgabe widmen wir uns dem „Red Stack" der KI. Und das ist wörtlich zu nehmen: Vom Einsatz großer Sprachmodelle bei der SQL-Generierung (Patrik Graf & Stefan Winkler), über die Wege zur passenden KI-Strategie (Thomas Keßler) bis hin zu der Frage, warum trotz aller Euphorie so viele Projekte scheitern (Alexander Hendorf) – diese Ausgabe liefert Denkanstöße, praktische Erfahrungen und offene Worte. Auch ein spannendes Interview mit Stefan Latuski zu KI in der Bundesagentur für Arbeit erwartet uns.

Auch die regulatorische Seite fehlt nicht: Der AI Act der EU wird ein Gamechanger – nicht nur juristisch, sondern auch technisch. Wie sich Governance, Compliance und Entwicklung verändern müssen, beleuchtet Benedikt Backhaus. Dazu gibt es spannende Einblicke in Datenanalyse und Security – schließlich bleibt der Datenstapel das Rückgrat jeder KI-Anwendung.

Und ja, auch diesmal kann ich es mir nicht verkneifen, meinen Husky-Vergleich zu bringen: KI ist wie ein Husky – beeindruckend, lernfähig, voller Energie. Aber wehe, man lässt ihn unbeaufsichtigt laufen. Dann zerlegt er einem in fünf Minuten den Garten – oder schreibt mit einem LLM fehlerhafte SQL-Statements direkt in die Produktionsdatenbank. Der Punkt ist: Wer KI nutzt, braucht klare Leitplanken sowie Training – und jemanden, der regelmäßig nach dem Rechten sieht.

Dieses Heft ist eine Einladung, mit Neugier und kritischem Blick in die Zukunft zu schauen. Und vielleicht erkennen wir dabei auch: Die Reise mit der KI ist nicht nur eine technische, sondern auch eine kulturelle.

Viel Freude beim Lesen!

Oliver Szymanski

Ausgabe Nr. 3/2025
auf Abruf!

Sprich mit den Daten: Datenanalyse mit ChatGPT im Praxistest

„KI als neuer Mitarbeiter" – Wie Large Language Models Arbeitsprozesse transformieren

Einleitung

KI

Security

Development

Datenbank

News

Intern

TIMELINE

13. bis 15. Mai 2025

Die APEX connect 2025 – die Fachkonferenz rund um die Low-Code-Plattform Oracle APEX – findet erstmals im Europa-Park in Rust statt. 352 Teilnehmende (darunter 12 Studierende) komen im Confertainment Center zusammen, um sich über die neuesten Entwicklungen in den Bereichen APEX, SQL & PL/SQL sowie Solutions auszutauschen. Mit 64 Sessions, darunter drei ganz unterschiedlichen, aber allesamt äußerst inspirierenden und sympathischen Keynotes von Martin Bach, Samuel Nitsche und Simon Hunt, vier DevTalks und zahlreichen 1:1-Sessions, ist für jeden Low-Code-Enthusiasten etwas dabei. Die Konferenz bietet Tech-Talks, Networking, eine Ausstellung, Party und leckeres Essen – alles bei bestem Frühlingswetter im Ambiente eines Freizeitparks.

14. bis 15. Mai 2025

Parallel zur APEX connect 2025 findet die DOAG 2025 Datenbank mit Cloud Infrastructure an gleicher Stelle statt. Die rund 150 Besucherinnen und Besucher können aus rund 50 Vorträgen zu ihren Lieblingsthemen Datenbank, Engineered Systems und Cloud Infrastruktur ihre Favoriten wählen. Die Teilnehmenden erfreuen sich am persönlichen Wiedersehen und der Möglichkeit, sich auszutauschen und spannende Sessions von internationalen Speakern live auf der Bühne erleben zu können. Fachlichen Austausch gibt es auch bei den 13 Ausstellern und Sponsoren, die ihre Lösungen im Ausstellungsort Dome präsentieren. Am Abend kommen nach dem gemeinsamen Abendessen mit den Teilnehmenden der APEX connect Achterbahn-Freunde auf ihre Kosten und können Park-Attraktionen wie den Eurosat – CanCan Coaster in der silbernen Kuppel und den spektakulären Adrenalin-Booster, die Achterbahn Silver Star, ausprobieren.

19. Mai 2025

Die Roadshow Highway 2 CloudLand des CloudLand Festivals startet in Coburg und ist zu Gast bei der HUK Coburg. Organisiert von der deutschsprachigen Cloud Native Community gibt es im Tourbus eine Mischung aus Party und Bier, Pizza und Networking, zwei Impulsvorträge und Diskussionen – eine unschlagbare Kombination aus Unterhaltung und Wissensaustausch.

20. Mai 2025

Der rote Bus macht auf der Highway to CloudLand einen Stopp in Nürnberg. Nach den Vorträgen sind die Teilnehmenden zu Erfahrungsaustausch und Networking ins DOAG-Bus-Bordbistro zu frisch gebackener Pizza, kühlem Bier und alkoholfreien Getränken eingeladen.

20. Mai 2025

Das Regionaltreffen München//Südbayern findet statt. Es gibt zwei Vorträge zu den Themen „Low Code Apps mit JSON-Daten? JSON Sources in APEX 24W.2!“ und „Document Generator – Die Alternative zu AOP?“.

22. Mai 2025

Beim Regionaltreffen Hamburg stehen gleich vier Vorträge auf der Agenda. Zum Thema Development spricht Jan Gorkow über „PL/SQL – Laufzeitoptimierung durch Parallelisierung“ und Carolin Krützmann über „Datenbankentwicklung mit SQL Developer for VSCode“. Des Weiteren referieren zum Thema Infrastruktur Benjamin Kurschies über „Database patching mit dem autoupgrade Tool“ sowie Dr. Benjamin Linnik über „Wie man Anomalien erkennt und ihre Ursachen mit KI in Echtzeit analysiert“.

26. Mai 2025

Die WebSession „DOAG & Low-Code Association – Teil 1 von 4“ findet statt. Zwei Organisationen, ein Ziel: Gemeinsam für die Zukunft der Softwareentwicklung. Die DOAG und die Low-Code Association laden ihre Mitglieder herzlich zu einem gemeinsamen Webinar ein. Im Mittelpunkt steht die neue Kooperation – und die Frage, wie wir gemeinsam die Zukunft der Softwareentwicklung gestalten können. Die Teilnehmenden erfahren mehr über die Ziele dieser Zusammenarbeit und erhalten exklusive Einblicke in geplante Aktivitäten – darunter der German Low-Code Day 2025 sowie der Low-Code Creator. Die WebSession bietet außerdem die Gelegenheit zum offenen Austausch über technologische Entwicklungen, strategische Schwerpunkte und mögliche Synergien – speziell zugeschnitten auf unsere DOAG-Mitglieder.

2. und 3. Juni 2025

Die KI Navigator lädt in die Hauptstadt ins Mercure Hotel MOA Berlin ein. Die Inhalte des kleinen, aber feinen Programms überzeugen die rund 200 Teilnehmerinnen und Teilnehmer. Ohne technisches Vorwissen finden Fach- und Führungskräfte hier Informationen mit Mehrwert und Orientierung in der komplexen Welt der KI-Anwendungen.

3. bis 5. Juni 2025

„APEX - 3 Tage kompakt“ heißt das Event der AOUG in Wien. Das dreitägige Seminar richtet sich an Personen mit keinen oder geringen APEX-Vorkenntnissen. Im Kurs lernen die Besucherinnen und Besucher die Konzepte und Techniken der Low-Code- Anwendungsentwicklung mit APEX kennen. Neben einer ausführlichen Einführung in die zu Grunde liegenden Modelle anhand von praktischen Beispielen bekommen Sie einen Überblick über das weitreichende Einsatzspektrum von Oracle APEX.

4. Juni 2025

In einer zweiten WebSession gibt es Einblicke in die Kooperation von DOAG und der Low-Code Association. Es werden aktuelle Veranstaltungen vorgestellt und gezeigt, welche Schwerpunkte beide Organisationen setzen. Mit dabei sind Mitgliedsunternehmen der Low-Code Association, die ihre Plattformen präsentieren und einen Einblick geben, wie sie und ihre Kunden Low-Code einsetzen.

5. Juni 2025

Im DOAG DevTalk mit Matthias Schulz und Christian Schwitalla dreht sich alles um das Thema „Modernes SQL: Views in der DB-Modellierung“.

11. Juni 2025

Die WebSession „DOAG & Low-Code Association – Teil 3 von 4“ steht an. Erneut werden aktuelle Veranstaltungen vorgestellt und gezeigt, welche Schwerpunkte beide Organisationen setzen sowie von Mitgliedsunternehmen Plattformen präsentiert. Mit dabei sind „Low Code Turbo durch nahtlose ALM-Integration“ von Eckhard Herdt, „Structr – KI nach dem Hype“ mit Michael Volpert und Axel Morgner sowie „Erstellung eines komplexen Fachverfahrens (u.a. am Beispiel Valikom) ohne vorherige Feinspezifikation mit der Allisa LCP“ von Christof Langer.

11. Juni 2025

Das Regionaltreffen Berlin Brandenburg bietet drei Vorträge. „Nutzung einer Nicht-Oracle-Datenbank als Persistenz-Schicht in APEX am Beispiel von PostgreSql“ und „Konzepte zur Historisierung von Datenänderungen“ von Jan Gorkow sowie „Barrierefreiheit in APEX“ von Janek Schweda.

12. Juni 2025

Der Tourbus der Highway to CloudLand hält in Chemnitz und ist zu Gast bei der Firma pentacor. Nach den Vorträgen gibt es wieder Erfahrungsaustausch und Networking im Bordbistro des DOAG-Busses.

13. Juni 2025

In einer DB WebSession präsentiert Andrew Lacy „Oracle On-Premises – eine Roadmap!?“

16. und 17. Juni 2025

Auf der AOUG Anwenderkonferenz 2025 – AI starts with Data der Austrian Oracle User Group wird den Besucherinnen und Besuchern in Wien präsentiert, wie Oracle mit seinen leistungsstarken Lösungen in den Bereichen Datenmanagement und Künstliche Intelligenz neue Maßstäbe setzt. Geboten werden spannende Anwenderberichte, exklusive Einblicke in aktuelle Trends und praktische Lösungsansätze, die zeigen, wie Oracle-Technologien Ihre Daten in intelligente Entscheidungen verwandeln können.

18. Juni 2025

Die WebSession „DOAG & Low-Code Association – Teil 4 von 4“ findet statt. Darin stellen sich erneut Mitgliedsunternehmen der Low-Code Association vor. Was verbindet die DOAG mit der Low-Code Association? In diesem gemeinsamen Webinar wird gezeigt, welche Ziele unsere Zusammenarbeit verfolgt, welche Themen uns verbinden – und welche Veranstaltungen aktuell anstehen. Die Plattformen Scopeland, formcycle und Oracle APEX werden präsentiert.

24. und 25. Juni 2025

Das Expertenseminar „Pluggable Database Einführung/Migration/Administration/Best Practices“ mit Marco Patzwahl findet in Berlin statt.

25. Juni 2025

Beim Regionaltreffen München/Südbayern gibt es den Vortrag „AI Vector Search in der Oracle Datenbank 23ai“ mit Referent Markus Kißling.

1. bis 4. Juli 2025

Die grenzenlosen Möglichkeiten der Cloud und viele spannende Innovationen können die Teilnehmerinnen und Teilnehmer der CloudLand erleben. Das Cloud Native Festival öffnet zum ersten Mal im Heide Park Soltau die Türen zu einer digitalen Zukunft, in der die Cloud das Herzstück bildet. Vernetzung mit Experten und das Entdecken der neuesten Trends der IT-Welt stehen im Mittelpunkt dieses Events. Mehr als 200 Sessions, 8 interaktive Beiträge und 27 Workshops und ein sommerliches Abendevent bieten an vier Tagen ein abwechslungsreiches Programm.

2. Juli 2025

Das Regionaltreffen Osnabrück/Bielefeld/Münster findet in Paderborn statt. Auf der Agenda stehen zwei Vorträge über

„Cloud Control 24ai“ von Ralf Durben und „MySQL Hochverfügbarkeit und Disaster Recovery“ von Matthias Jung.

11. Juli 2025

Die DB WebSession „23ai Features für Entwickler“ mit Marco Pachaly-Mischke findet als Wiederholung statt.

22. Juli 2025

Das Regionaltreffen München/Südbayern findet in München statt. Auf dem Programm stehen Sessions zu den Themen „Was gibt es Neues bei MySQL 8.4 und MySQL 9? “ von Mario Beck und „MySQL HeatWave – MySQL in der Cloud mit integrierter Analyse Engine, und noch einiges mehr“ von Vincent Trutschler.

23. Juli 2025

Der DOAG DBTalk zum Thema „Datenbank-Migration“ mit Bruno Cirone und Matthias Jung findet statt. Behandelt werden die Überlegungen von der initialen Planung und Risikobewertung bis hin zu Ausführungsstrategien und der Validierung nach der Migration.

CLOUD NATIVE FESTIVAL

im Heide Park Soltau

CloudLand
www.cloudland.org

19. – 22. MAI 2026

#CLOUDLAND2026

„Im Zuge der Automatisierung und Digitalisierung hat sich die BA ein Wertefundament gegeben: die BA-Datenethik"

Martin Meyer, Redaktionsleiter des Red Stack Magazin, sprach mit Stefan Latuski, CIO der BA sowie CEO des IT-Systemhauses der BA, über das Thema Künstliche Intelligenz und Digitalisierung, deren verantwortungsvollen Einsatz bei der Bundesagentur für Arbeit und über Datenschutz.

Bitte stellen Sie sich unseren Lesern vor. Wer sind Sie und was ist Ihre Aufgabe bei der Bundesagentur für Arbeit?

Ich bin 41 Jahre alt, habe Wirtschaftswissenschaften an der Bergischen Universität Wuppertal studiert und bin gebürtiger Gelsenkirchener. Als CIO treibe ich die Digitale Transformation der BA voran – in Doppelfunktion: ich bin zugleich CEO des IT-Systemhauses, wo meine BA-Reise im Juli 2021 startete. Vor meinem Wechsel in den öffentlichen Sektor war ich über 15 Jahre in verschiedenen IT-Bereichen für den Siemens-Konzern tätig, zuletzt als CIO von Siemens Mobility.

Was war ihr persönlicher Wow-Effekt bezüglich KI und warum?

Mein persönlicher Wow-Effekt im Umgang mit KI war im Rahmen meines berufsbegleitenden EMBA-Studiums. Ein Professor ermutigte uns, GPTs aktiv zu nutzen – nicht als Abkürzung, sondern als intelligente Ergänzung. Das ist gerade im akademischen Kontext ein neues Verständnis von Lernen, Denken und Arbeiten mit KI – zumindest im Vergleich zu meinem ersten Studium vor vielen Jahren. Besonders spannend fand ich auch, ein eigenes Promptverzeichnis anzulegen. Das hat mir noch einmal verdeutlicht, wie sehr der Output von der Qualität der Interaktion abhängt – und wie wichtig es ist, den Umgang mit KI aktiv zu gestalten.

Wie sieht Ihre Vision für den Einsatz von KI in der Bundesagentur für Arbeit aus und wie trägt sie zur Erfüllung der Mission der Agentur bei?

Unsere Vision ist es, als „digitale Behörde" zur modernsten öffentlichen Dienstleisterin in Europa zu werden. Technologien wie Künstliche Intelligenz und Maschinelles Lernen ermöglichen es uns, dabei neue Wege zu gehen: der Einsatz von KI in Assistenzsystemen, die Nutzung moderner Automatisierungstechnologien wie zum Beispiel der Einsatz von Low-Code/No-Code oder Robotic Process Automation (RPA) und der Ausbau interner Automatisierungsprozesse schaffen neue Services für unsere Kundinnen und Kunden und entlasten unsere Beschäftigten von zeitintensiven Routineaufgaben. Auf den sinnvollen Einsatz von Automatisierung kann die BA auch deshalb nicht verzichten, weil wir vor großen demografischen Herausforderungen stehen: Bis 2032 werden rund 35 Prozent unserer Beschäftigten in den Ruhestand gehen. Diese Lücke wird sich kaum durch Nachrekrutierungen schließen lassen. Automatisierung und insbesondere der Einsatz von KI sind Teil der Lösung und helfen uns dabei, die personellen Abgänge zu kompensieren.

Allerdings können wir die Automatisierungspotenziale wegen bestehender Gesetze und Regularien noch nicht voll ausschöpfen, da rechtliche Rahmenbedingungen die durchgängige Automatisierung von Prozessen erschweren (zum Beispiel Vier-Augen-Prinzip bei Auszahlungsanordnungen).

Welche konkreten Anwendungsfälle sehen Sie für KI in Bezug auf die Arbeitsvermittlung und -verwaltung?

Ein Beispiel ist der BERUFEBOT. Der Chatbot ist eine generative KI zur dialogbasierten Wissenserschließung zum Thema Berufe. Der BERUFEBOT umfasst Berufsinformationen zu über 3.500 Berufen und wird natürliche Sprache verarbeiten, also so, wie die Menschen wirklich sprechen. Er gibt dann qualitätsgesicherte Antworten sowie gezielte Quellenverweise. 2025 ist der Rollout an Mitarbeitende geplant und perspektivisch soll der BERUFEBOT auch für unsere Kundinnen und Kunden zur Verfügung stehen.

Im Arbeitgeberservice wurde eine automatische Verarbeitung beziehungsweise Übernahme von durch Unternehmen übermit-

telten Stellenangeboten in das Fachverfahren der BA eingeführt. Weitere KI-Einsatzmöglichkeiten sind Voice-Bots und Speech-to-text-Möglichkeiten. Im Herbst letzten Jahres haben wir mit Aleph-Alpha einen Rahmenvertrag abgeschlossen. Zusammen mit dem Heidelberger Start-up wird die BA KI als Schlüsseltechnologie auf ihrer Automatisierungsreise in den kommenden Jahren implementieren.

Wie stellen Sie sicher, dass der Einsatz von KI in der Bundesagentur für Arbeit verantwortungsbewusst, ethisch und transparent ist, insbesondere im Hinblick auf Datenschutz und Diskriminierung?

Im Zuge der Automatisierung und Digitalisierung hat sich die BA ein Wertefundament gegeben: die BA-Datenethik. Wir haben eine Leitlinie entwickelt und richten unser Handeln an sieben Datenethik-Prinzipien aus. So stellen wir sicher, dass die BA algorithmische Entscheidungssysteme entsprechend unserer gesellschaftlichen Werte einsetzt – und zum Beispiel keine unbeabsichtigten Vorurteile oder Verzerrungen implementiert werden. Über die Einhaltung der datenthetischen Maßnahmen wachen ein Datenethik-Expertenteam und ein Datenethik-Gremium.

Welche Prioritäten hat die BA-IT im Kontext Digitalisierung und KI?

Bürgerinnen und Bürger erwarten zunehmend digitalen Zugang zu staatlichen Leistungen, das heißt schnellere, effizientere und passgenauere Services – jederzeit und überall.

Wir haben eine Digitalisierungsagenda aufgesetzt, die wir Schritt für Schritt umsetzen. Neben steigenden Erwartungen treiben uns dabei weitere Faktoren. Allen voran der bereits erwähnte demografische Wandel. Deshalb entwickeln wir Lösungen, die sicherstellen, dass wir unsere Aufgaben auch weiterhin erfüllen können. Chancen liegen hier ganz klar in der Digitalisierung und Automatisierung. Beides betreiben wir „human friendly“, also mit und im Sinne der Menschen. Zudem gewinnt die Resilienz der Geschäftsprozesse an Bedeutung, um auch in unerwarteten Notfällen wie etwa Cyberangriffen oder Naturkatastrophen handlungsfähig zu bleiben. Wir müssen aber auch den gesetzlichen Anforderungen gerecht werden. Die Nutzung entsprechender Cloud-Dienste wird unverzichtbar, denn ein digitales Frontend reicht nicht aus – wir müssen auch die Prozesse im Hintergrund digitalisieren. Damit ist die Cloud-Transformation, die eng mit der Nutzung von KI-Lösungen zur Effizienzsteigerung und Entlastung unserer Mitarbeitenden verzahnt ist, ein Fokusthema in den nächsten Jahren. Wir setzen bei der Automatisierung auf „Human Friendly Automation“, das heißt, wir automatisieren dort, wo es sinnvoll und möglich ist und schaffen so für unsere Mitarbeiterinnen und Mitarbeiter durch Automatisierung repetitiver Tätigkeiten mehr Freiräume für die persönliche Beratung und Betreuung.

STEFAN LATUSKI

Stefan Latuski ist seit August 2023 CIO der Bundesagentur für Arbeit (BA) und bereits seit 2021 Vorsitzender der Geschäftsführung (CEO) des IT-Systemhauses der Bundesagentur für Arbeit. In seiner Rolle treibt er die digitale Transformation der größten deutschen Behörde voran. Vor seinem Wechsel in den Public Sector war der Wirtschaftswissenschaftler 15 Jahre in verschiedenen IT-Bereichen für den Siemens-Konzern tätig, zuletzt als CIO von Siemens Mobility.

DOAG 2026 Datenbank

mit Cloud Infrastructure

+ APEX connect

DOAG

Heide Park Soltau

SUPER-SAVER

Bis 31.10.2025

18. – 19. Mai 2026

datenbank.doag.org

Sprich mit den Daten: Datenanalyse mit ChatGPT im Praxistest

Fabian Heidenstecker, Opitz Consulting

Die Analyse strukturierter Daten gehört heute zum Alltag vieler IT-Abteilungen. Ob Kundenverhalten, Vertriebserfolge oder Betriebskennzahlen – überall entstehen Daten, die es zu verstehen und zu nutzen gilt. Doch klassische Werkzeuge wie SQL, Excel oder Business-Intelligence-Tools setzen meist technisches Know-how voraus. Generative KI, etwa in Form von ChatGPT, verspricht eine neue Form der Interaktion: Natürlichsprachliche Abfragen, automatische Visualisierungen, sogar erste Machine-Learning-Modelle – direkt im Chatfenster. Doch wie gut funktioniert das in der Praxis? In diesem Erfahrungsbericht zeige ich Schritt für Schritt, wie ich mit ChatGPT ein vollständiges Data-Analytics-Projekt durchgeführt habe – ohne eigene Entwicklungsumgebung, allein mit Prompts und einem realitätsnahen Unternehmensdatensatz.

Wie leistungsfähig ist die GenAI-basierte Datenanalyse?

Traditionell war tiefgreifende Datenanalyse Expertinnen und Experten vorbehalten – mit spezialisierten Tools, komplexer Syntax und oft steiler Lernkurve. Doch mit der Verfügbarkeit leistungsfähiger Sprachmodelle wie ChatGPT verändert sich dieser Paradigmenwechsel: Strukturierte Daten können heute dialogbasiert analysiert, visualisiert und interpretiert werden – ohne tiefgreifende Programmierkenntnisse oder dedizierte Software. Aber wie gut und verlässlich sind diese Modelle?

In diesem Projekt demonstriere ich anhand eines praxisnahen Szenarios, wie sich eine Datenanalyse durch GenAI anfühlt und ob die Ergebnisse wirklich halten, was sie versprechen. Die Basis bildet ein synthetischer Unternehmensdatensatz mit Informationen zu Kunden, Produktkategorien und Kaufverhalten – ein typisches Setup, wie es in vielen Unternehmen existiert.

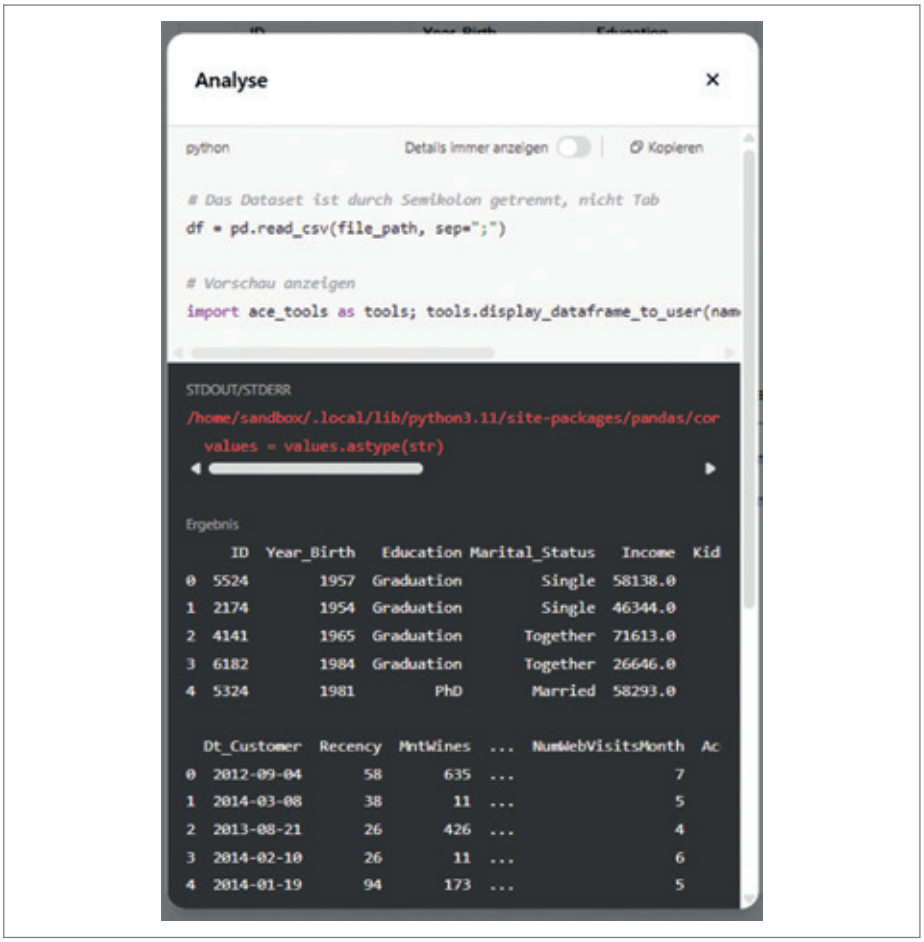


Abbildung 1: Dateianalyse mit generierten Kommentaren (Quelle: Fabian Heidenstecker)

Meine Zielsetzung

Ziel meiner Analyse ist es, zwei zentrale Fragen zu beantworten:

1. Wie leistungsfähig sind GenAI-Tools wie ChatGPT bei typischen Aufgaben der Datenanalyse? Dazu gehören das Laden, Transformieren, Visualisieren und Interpretieren strukturierter Daten.
2. Wie gut lassen sich datenbasierte Erkenntnisse in kontextualisierte Aussagen überführen – etwa in Form von Data-Driven Personas oder einfachen Vorhersagemodellen?

Die Analyse erfolgte vollständig im Dialog mit ChatGPT – ohne IDE, rein über Prompts im Chatfenster.

Hinweis: Die im Artikel abgebildeten Prompts wurden teilweise zur besseren Lesbarkeit gekürzt. Die vollständigen Versionen sind über die Quellangaben verlinkt.

Datengrundlage: Marketingkampagnen und Kundendaten

Der verwendete Datensatz enthält Informationen zu über 2.000 Kundinnen und Kunden – von soziodemografischen Angaben wie Alter, Bildungsgrad oder Einkommen bis hin zu Kaufverhalten über verschiedene Vertriebskanäle. Enthalten sind unter anderem Ausgaben für Warengruppen wie Wein, Fleisch oder Gold, Reaktionen auf frühere Kampagnen und genutzte Vertriebskanäle wie Online, Filiale oder Katalog. Die Struktur ist typisch für CRM-Systeme oder Data Warehouses im Retail-Umfeld – damit ein ideales Testfeld für KI-gestützte Datenanalyse.

Die Daten spiegeln typische Strukturen realer Vertriebs- und Marketingdaten wider – ideal, um das Potenzial generativer KI zu demonstrieren.

Daten einlesen mit ChatGPT

Über die Dateiupload-Funktion in ChatGPT (GPT-4o mit Advanced Data Analysis)

Abbildung 2: Boxplots der Ausreißer (Quelle: Fabian Heidenstecker)

Ausgaben Und Kundenanzahl Je Kohorte

	Kohorte	Anzahl Kunden	Ø Gesamtausgabe (Mnt*)	Gesamtausgabe (Mnt*)
1	1940	107	933.6448598130842	99900
2	1950	460	673.6195652173913	309865
3	1960	506	623.4110671936759	315446
4	1970	740	523.222972972973	387185
5	1980	363	532.2506887052342	193207
6	1990	61	810.5737704918033	49445

Abbildung 3: Erste Kennzahlen nach Kohorte (Quelle: Fabian Heidenstecker)

ließ sich die CSV-Datei direkt einlesen – inklusive automatischer Erkennung des Semikolon-Trennzeichens *(siehe Abbildung 1)*.

Schon nach dem Laden machte ChatGPT Vorschläge für mögliche nächste Schritte. Ich entschied mich zunächst für eine Prüfung der Datenqualität. Hier überzeugte mich vor allem die automatische tabellarische Darstellung mit Hinweisen auf fehlende Werte und potenzielle Ausreißer – ganz ohne manuelle Rückfragen.

Die Ausreißer der Geburtsjahre ließ ich mir als Boxplot darstellen *(siehe Abbildung 2)*.

Erste Kennzahlen und Gruppenanalysen

Mit einfachen Prompts wie
„Wie viele Kunden enthält der Datensatz und wie hoch ist das durchschnittliche Einkommen?“ und
„Wie viel gibt ein alleinstehender Kunde durchschnittlich für Wein aus?“
lieferte ChatGPT schnell Ergebnisse. Auch gruppierte Analysen waren möglich, etwa:

„Bitte gruppiere die Weinausgaben nach Alterskohorten (10-Jahres-Schritte nach Geburtsjahr).“

Die Resultate wirkten plausibel – und hielten auch dem Excel-Gegencheck stand (dort habe ich die Werte nachgerechnet, was sich für den überschaubaren Datensatz angeboten hat).

Ermutigt von den Ergebnissen, stellte ich fortgeschrittenere Fragen, zum Beispiel: „Wie hoch sind die Ausgaben für Wein, gruppiert nach dem Geburtsjahr?“ Außerdem ließ ich mir die Geburtsjahre in 10-Jahres-Schritten gruppieren, also in Alterskohorten *(siehe Abbildung 3)*. Auch hier war ich mit dem Ergebnis zufrieden.

Wichtig zu verstehen

ChatGPT liefert auf Rechenfragen häufig korrekte Antworten, was den Eindruck erweckt, es könne rechnen. Tatsächlich basiert die Ausgabe jedoch auf Wahrscheinlichkeiten, die aus Sprachmustern im Training abgeleitet wurden. Als Sprachmodell führt es keine echten Berechnungen durch, sondern erzeugt bei Bedarf Code, der von einer extra Rechenumgebung ausgeführt werden kann. Die Fähigkeit zur scheinbaren Berechnung ist somit eine Folge der Mustererkennung, nicht mathematischer Kompetenz.

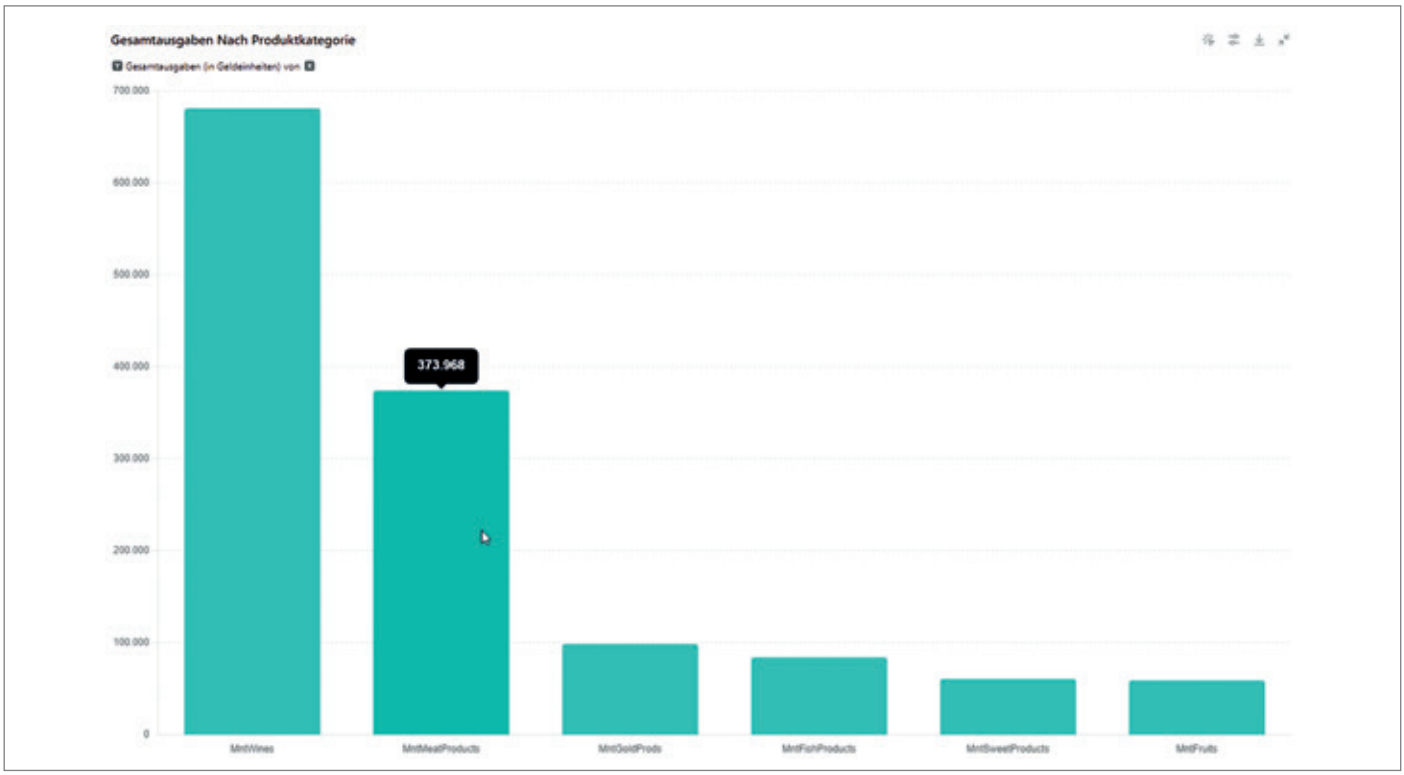


Abbildung 4 Interaktives Säulendiagramm (Quelle: Fabian Heidenstecker)

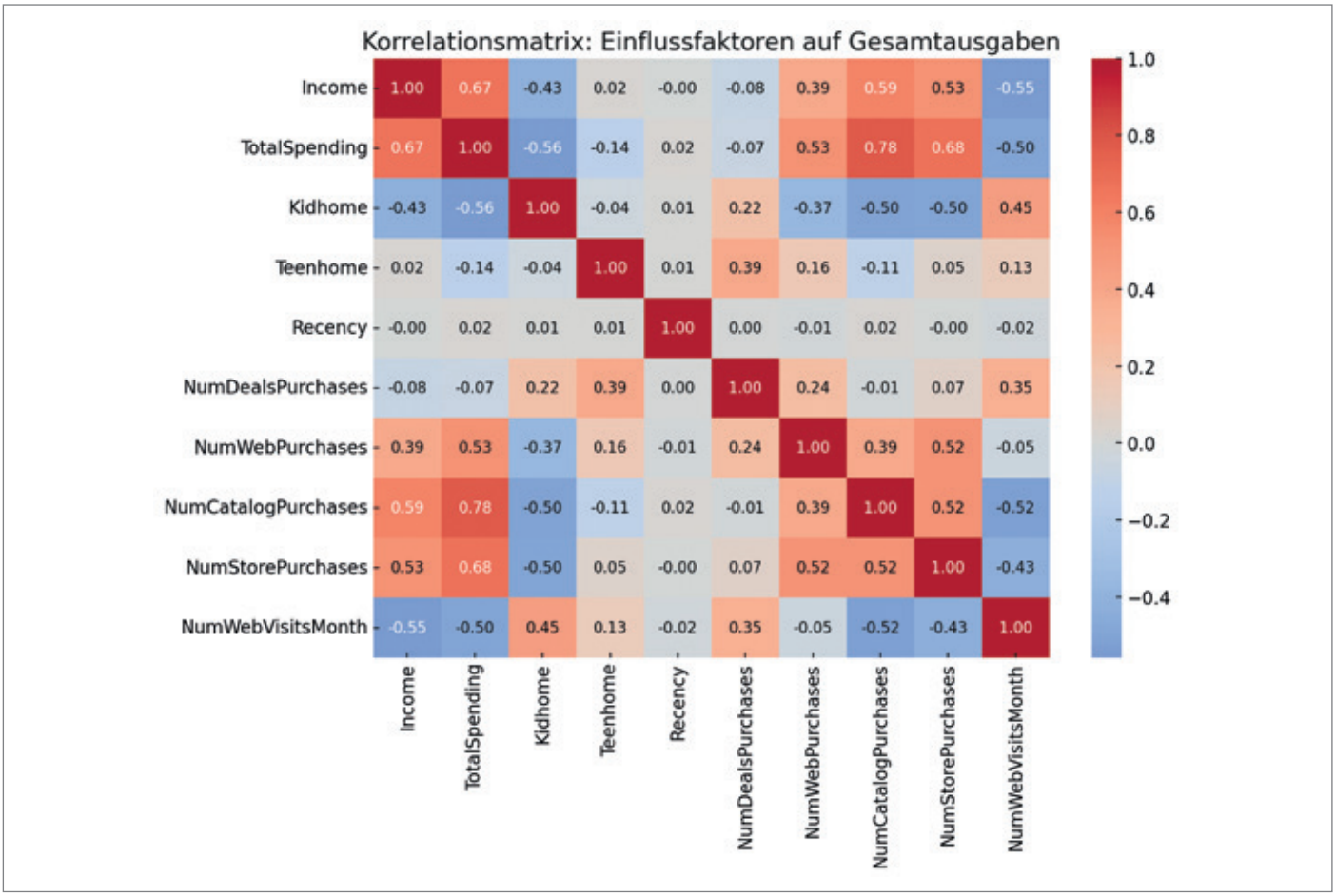


Abbildung 5: Korrelationsmatrix (Quelle: Fabian Heidenstecker)

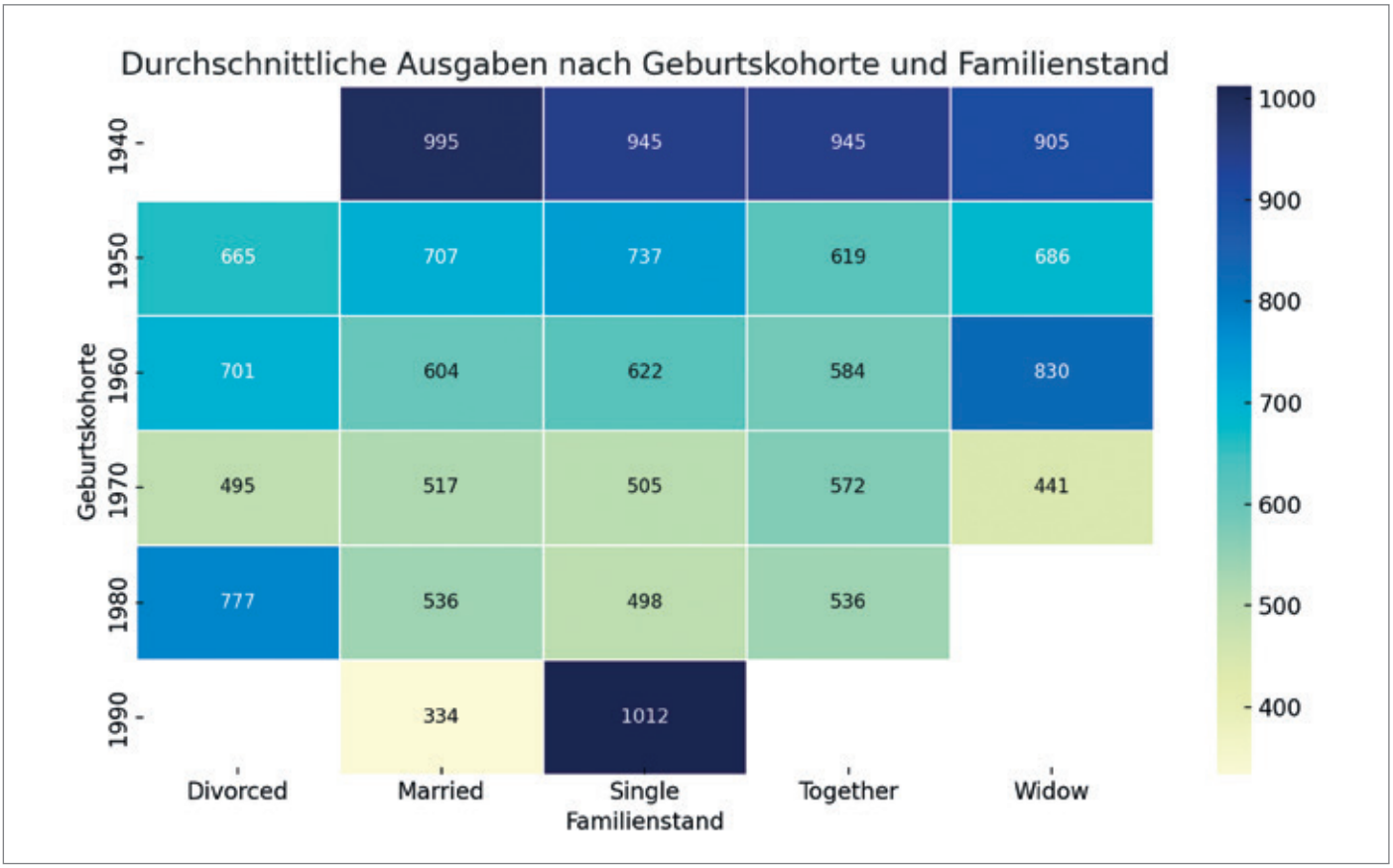


Abbildung 6: Heatmap (Quelle: Fabian Heidenstecker)

Bildung	Familienstand	Einkommen	Kinder (klein/Teen)	Geburtsjahr	Beschwert	Prognose Ausgabe
Graduation	Married	50.000 €	1 / 1	1980	Nein	251,70
PhD	Together	75.000 €	0 / 0	1970	Nein	1.507,24
Master	Single	30.000 €	2 / 0	1990	Ja	108,75

Tabelle kopieren

Abbildung 7: Prognose-Tabelle (Quelle: Fabian Heidenstecker)

modellvergleich.py > ...

```python
# ----------------------------
# 4. Evaluation pro Modell
# ----------------------------
results = []

for name, model in models.items():
    print(f"Trainiere Modell: {name}")

    pipeline = Pipeline(steps=[
        ("preprocessor", ColumnTransformer(
            transformers=[("cat", OneHotEncoder(drop="first"), cat_features)],
            remainder="passthrough"
        )),
        ("regressor", model)
    ])

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)

    pipeline.fit(X_train, y_train)
    y_pred = pipeline.predict(X_test)

    mae = mean_absolute_error(y_test, y_pred)
    rmse = np.sqrt(mean_squared_error(y_test, y_pred))

    results.append({
        "Modell": name,
        "MAE": round(mae, 2),
        "RMSE": round(rmse, 2)
    })

# ----------------------------
# 5. Ergebnisse anzeigen
# ----------------------------
results_df = pd.DataFrame(results).sort_values("MAE")
print("\nModellvergleich (nach MAE sortiert):\n")
print(results_df.to_string(index=False))
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Trainiere Modell: Linear Regression
Trainiere Modell: Random Forest
Trainiere Modell: Random Forest
Trainiere Modell: Gradient Boosting
Trainiere Modell: XGBoost

Modellvergleich (nach MAE sortiert):

          Modell    MAE   RMSE
         XGBoost 201.14 307.44
   Random Forest 201.32 300.09
Gradient Boosting 212.61 309.82
Linear Regression 293.33 391.15
```

Abbildung 8: Python-Code für lokale Ausführung (Quelle: Fabian Heidenstecker)

Visualisierungen direkt im Chat

Auch die Visualisierung funktionierte erstaunlich gut. Ob Säulendiagramm, Heatmap oder Boxplot – die Diagramme wurden automatisch generiert und waren direkt im Chatfenster anklickbar und exportierbar.

Beispiel „Säulendiagramm":

Prompt: „Erzeuge ein Säulendiagramm, das die Verkaufsvolumina der Warengruppen zeigt."

Ergebnis: Die Visualisierung war ansprechend, aber optisch musste ich da noch mal dran. Was mir leicht gemacht wurde, denn direkt auf der Oberfläche konnte ich Details nachjustieren – etwa Farbgebung oder Skalierung *(siehe Abbildung 4)*. Besonders hilfreich: Der interaktive Modus mit Tooltips, die präzise Werte beim Mouse-over anzeigen.

Als nächstes wollte ich herausfinden, welches grundlegende Verständnis ChatGPT von der Visualisierung hat und bat um Vorschläge zur Darstellung des Merkmals „Familienstand". ChatGPT lieferte zwei Optionen – Balkendiagramm für absolute, Kreisdiagramm für relative Häufigkeiten – und verwies explizit auf Vor- und Nachteile. Fachlich fundiert, stilistisch sauber.

Beispiel „Boxplot"

Prompt: Um eine fortgeschrittene Visualisierungsmethode zu testen, bat ich ChatGPT, einen Boxplot zu erzeugen. Boxplots sind in der Statistik beliebt, um die Verteilung von Werten zu zeigen. Dieser Diagrammtyp ist nicht so gängig und auch nicht in allen Visualisierungstools (wie zum Beispiel Power BI) als Standard verfügbar.

Ergebnis: Auch hier lieferte das Modell ein solides Ergebnis, inklusive Interpretation: Höherer Bildungsgrad korreliert mit höherem Einkommen. Ausreißer werden auf Wunsch entfernt.

Korrelationen und Zusammenhänge

Ein Highlight war für mich die automatische Erstellung einer Korrelationsmatrix – inklusive Interpretation. So wurde beispielsweise der erwartete Zusammenhang zwischen Einkommen und Ausgaben direkt erkannt, erläutert und grafisch aufbereitet. Hier zwei Beispiele:

- **Korrelationsmatrix erstellen**

  Prompt: „Erstelle eine Korrelationsmatrix aller numerischen Variablen und interpretiere die stärksten Zusammenhänge."

  Ergebnis: ChatGPT lieferte nicht nur die Matrix selbst, sondern auch gleich eine textuelle Auswertung – etwa den erwartbaren Zusammenhang zwischen Einkommen und Ausgaben oder zwischen Bildung und Kaufverhalten *(siehe Abbildung 5)*.

- **Zusammenhang zwischen kategorischen Merkmalen**

  In der weiter oben beschriebenen Heatmap hatte ChatGPT visualisiert, wie Alter (Geburtsjahr) und Familienstand der Personen aus meinem Datenset zusammenhängen – mit Farbcodierung nach durchschnittlichen Ausgaben. Besonders hilfreich: Die Gruppengrößen ließen sich direkt als Zusatzinformation in Klammern darstellen.

  Prompt: „Zeige den Zusammenhang zwischen den Geburtsjahren und dem Familienstand, bezogen auf die Gesamt-Ausgaben. Ich hätte gerne eine Heatmap."

  Ergebnis: Auch hier erzeugte ChatGPT die passende Visualisierung samt Kontextualisierung in natürlicher Sprache *(siehe Abbildung 6)*.

Vorhersagemodell mit Machine Learning

Eine Frage, über die aktuell noch viel spekuliert wird: Wie gut beherrscht ChatGPT Machine-Learning-Algorithmen, beziehungsweise ist es in der Lage, ein Vorhersagemodell zu erstellen?

Prompt: Erstelle ein Vorhersagemodell. Ich möchte verschiedene demographische Attribute und das Einkommen eingeben und eine Vorhersage der Gesamtausgaben für alle Warengruppen erhalten (siehe Abbildung 7).

Ergebnis: Heraus kam ein vollständiges Modell inklusive Berechnung des Mean Absolute Error (~201 €) – durchaus solide für einen ersten Wurf. Es wurde ein Random-Forrest-Algorithmus gewählt. Die Prognose erfolgte direkt im Chat – ohne externe IDE. Auch das Ausführen des Modells mit zufälligen Testdaten war direkt im Chat machbar und das Ergebnis wurde als Tabelle dargestellt, die sich per Knopfdruck auch noch exportieren ließ.

An dieser Stelle wich ich von meiner ursprünglichen Zielstellung ab. Ich wollte nämlich herausfinden, ob der Code für die Vorhersage auch lokal ausgeführt werden kann. Denn für einen realen Anwendungsfall würde man die Werte beispielsweise in einem Nachtlauf berechnen lassen, um sie dann zum Beispiel in das DWH zu integrieren und gegebenfalls in einem Report anzeigen zu lassen. Den Code übernahm ich per Copy & Paste in eine Python-Umgebung. Als kleinen Service generierte ChatGPT den passenden Befehl, um die notwendigen Python-Bibliotheken zu installieren *(siehe Abbildung 8)*.

Der circa 100 Zeilen lange Code ließ sich problemlos ausführen, es gab keine Fehlermeldungen oder Warnungen. ChatGPT lieferte einen gut strukturierten Code, auch Kommentare waren enthalten. Der Code war syntaktisch und strukturell sauber – ideal für produktionsnahe Tests.

Um hier noch mehr Eindrücke zu gewinnen und zu prüfen, ob die Vorhersagequalität noch besser werden kann, habe ich den Code um eine Schleife erweitern lassen. Diese trainierte nacheinander verschiedene Modelle auf Basis bekannter Algorithmen wie etwa XGBoost, und gab die entsprechenden Qualitätskennzahlen aus.

Ergebnis: Insgesamt war ich aber mehr als beeindruckt. Mit nur ein paar Prompts erstellte man Vorhersagemodelle in Python. Natürlich habe ich ein Grundverständnis von den Algorithmen, weiß was Python-Bibliotheken sind, und kann mir lokal eine virtuelle Umgebung einrichten. Der springende Punkt ist, dass ich mich nicht sehr gut mit der Syntax auskenne und hier am meisten Zeit spare.

Klar wurde aber auch: Ohne Grundwissen zu ML und Modellinterpretation stößt man beim Prompting schnell an Grenzen.

Data-Driven Personas und Data Storytelling

Im Marketing sind Personas ein etabliertes Mittel, um Zielgruppen greifbar zu machen und zum Beispiel die Kommunikationsmaßnahmen besser auf ihre Bedürfnisse zuzuschneiden. Ich ließ mir zunächst generische Personas basierend auf dem Datensatz erstellen – später auch spezifische Varianten pro Warengruppe, sogenannte Data-Driven Perso-

Abbildung 9: Infografik einer Persona (Quelle: Fabian Heidenstecker)

nas. Statt auf subjektiven Einschätzungen basieren diese auf realen Daten.

Nachdem ich grundsätzliche Informationen zum Aufbau einer guten Data Story erhalten hatte, experimentierte ich mit einigen Prompts zur Erzeugung der Personas. Dies waren für meinen Praxistest die umfangreichsten Prompts, daher dient der folgende Prompt als verkürztes Beispiel.

Prompt: „Erzeuge 5 Data-Driven Personas basierend auf dem Kaufverhalten von Kund:innen."

Ergebnis: tabellarische Personas mit Attributen wie Alter, Einkommen, bevorzugtem Kanal und Ausgabenverhalten. Ergänzend generierte ChatGPT auch passende Handlungsempfehlungen für das Marketing.

Da ChatGPT auch Bilder erstellen kann, ging ich noch den letzten Schritt und erstellte eine Infografik *(siehe Abbildung 9)*.

Zwischenfazit: Die Ergebnisse sind nützlich – wenn auch mit Einschränkungen:

- Mehrere Infografiken gleichzeitig lassen sich nicht generieren.
- Stil und Inhalt der generierten Bilder sind inkonsistent.
- Textliche Beschreibung und Attributwerte passen nicht immer zusammen.

Der Einsatz von ChatGPT bietet in diesem Fall einen guten Startpunkt für die datengetriebene Segmentierung.

Grenzen und Verantwortung

ChatGPT verändert die Art, wie wir mit Daten interagieren. Statt sich durch komplexe BI-Tools zu klicken oder SQL-Abfragen zu schreiben, reicht oft eine einfache Frage – mit teils überraschend guten Ergebnissen. Für technisch versierte Anwender bedeutet das vor allem: Geschwindigkeit, Flexibilität und neue Spielräume für explorative Analysen.

Doch die Risiken sind nicht zu unterschätzen: KI erfindet mitunter Informationen, rechnet nicht wirklich und ist abhängig von klarer Kommunikation. Gerade in sicherheitskritischen oder regulierten Umgebungen ist ein sorgfältiger Review unerlässlich. Auch der Datenschutz muss mitgedacht werden – besonders bei sensiblen Informationen.

Sensible Daten schützen

Auch wenn in meinem Beispieldatensatz keine Klarnamen oder direkte Identifikatoren enthalten sind, handelt es sich bei den Daten um sensible Informationen – schließlich geben Merkmale wie Einkommen, Alter oder Kaufverhalten Rückschlüsse auf reale Personen. Der verantwortungsvolle Umgang mit solchen Daten ist daher unerlässlich.

Bei der Nutzung generativer KI zur Analyse sollte daher immer sichergestellt sein, dass keine personenbezogenen Daten ungeschützt verarbeitet oder weitergegeben werden. Der Einsatz generativer KI – insbesondere cloudbasierter Lösungen – erfordert eine präzise Betrachtung hinsichtlich DSGVO, Speicherort, Zugriffskontrolle und Anonymisierung.

Tipp: Frühzeitig Datenschutzmaßnahmen wie Pseudonymisierung oder On-Premise-Verarbeitung einplanen.

Fehlerquellen im Blick behalten

Die GenAI operiert mit Wahrscheinlichkeiten. Von Mathematik besitzt sie nur ein begrenztes Verständnis. Insbesondere bei statistischen Analysen kommt sie daher an ihre Grenzen, was sie so direkt nicht unbedingt zugibt.

Wir dürfen nicht vergessen, dass die Sprachmodelle in erster Linie dazu da sind, flüssige und gut klingende Sprache zu generieren. Ob die Inhalte korrekt sind, ist für sie dabei zweitrangig, teilweise werden diese auch einfach frei erfunden. Das kennen wir dann unter den Begriff „Halluzinationen".

Was bedeutet das zum Beispiel für unseren Python-Code?

Ein klassisches Beispiel ist die Berechnung von Durchschnittswerten. Die sind zuweilen gar nicht so intuitiv berechenbar, da sich der Durchschnitt immer auf etwas bezieht: Durchschnitt pro

Zeiteinheit, Durchschnitt über alle Kunden, Durchschnitt aller Zahlen und vieles mehr.

Drückt man sich nicht klar genug aus, dann kann hier eine falsche Annahme erfolgen, und ein syntaktisch richtiger Python-Code kann trotzdem ein ungenaues oder sogar falsches Ergebnis liefern.

Tipp: Nicht blind auf die KI vertrauen, sondern das eigene Zahlenverständnis und die eigene Einschätzung mit einsetzen! Im Zweifel würde ich dazu raten, den Code extern auszuführen und zum Beispiel Zwischenergebnisse der Berechnungen ausgeben zu lassen, um diese nachzuprüfen.

Fazit: KI kann viel – aber nicht alles

Generative KI ermöglicht heute bereits leistungsstarke, intuitive Datenanalysen – mit hoher Zugänglichkeit und erstaunlicher Effizienz. Doch: Sprachmodelle operieren mit Wahrscheinlichkeiten, nicht mit „Verständnis". Fehlinterpretationen und Halluzinationen sind keine Seltenheit.

Gerade bei numerischen Werten ist Präzision gefragt. Unklare Prompts können zu korrektem, aber semantisch falschem Code führen. Deshalb gilt: Ergebnisse validieren, Code nachvollziehen und bei Bedarf lokal prüfen.

Trotz aller Faszination bleibt also ein realistischer Blick wichtig:

- „Rechnen" heißt nicht „verstehen" – Durchschnittswerte können falsch interpretiert werden, wenn der Prompt unklar ist.
- Code ist nicht immer korrekt – auch syntaktisch einwandfreier Code kann logisch falsch sein.
- Datenschutz ist Pflicht – gerade bei cloudbasierten KIs darf nie mit personenbezogenen Daten gearbeitet werden

Wo stehen wir aktuell?

Für technisch versierte Anwenderinnen und Anwender bietet ChatGPT ein mächtiges Werkzeug zur explorativen Analyse und Visualisierung von Daten – gerade in frühen Projektphasen. Die Fähigkeit, in natürlicher Sprache komplexe Auswertungen durchzuführen, spart enorm Zeit und senkt die Einstiegshürden.

Ausblick: Integration in Enterprise-Umgebungen

Die natürlichsprachliche Interaktion mit Daten wird zukünftig eine wichtige Rolle spielen – auch in Oracle-zentrierten Architekturen. Denkbare Einsatzszenarien:

- Generative KI als Add-on für Oracle Analytics Cloud
- Prompt-basierte Analyse in Kombination mit Autonomous Data Warehouse
- Ergänzung zu klassischen Tools wie APEX

Entscheidend ist: No Data – No AI. Daten müssen gut aufbereitet und verfügbar sein, um auch in einer explorativen Datenanalyse nutzbar zu sein.

Wichtig ist dabei immer: Die KI ersetzt keine Fachlichkeit – aber sie reduziert Hürden und beschleunigt Prozesse.

Mit zunehmender Reife der Werkzeuge wird die Grenze zwischen Analyse und Kommunikation weiter verschwimmen – und genau hier liegt das strategische Potenzial von GenAI.

Quellen

Prompts, Dataset und komplette Analyse:
https://github.com/Fabster79/redstack_eda

Über den Autor

Fabian Heidenstecker ist seit circa 20 Jahren in der IT unterwegs. Die letzten 15 Jahre in unterschiedlichsten Positionen im Consulting. Ursprünglich startete er im CRM- Umfeld und kam dort erstmals mit dem Thema BI und Analytics in Berührung. So wurden Daten seine Leidenschaft, welcher er in zahlreichen Projekten nachgehen konnte. In den letzten Jahren kamen noch die Themen Machine Learning und künstliche Intelligenz hinzu.

Heute ist er Senior Manager für Solutions bei Opitz Consulting und begleitet mit seinem Team Kunden auf dem Weg zur Data-Driven Company.

Fabian Heidenstecker
fabian.heidenstecker@
opitz-consulting.com

Wege zur passenden KI-Strategie: KI systematisch und gewinnbringend im Unternehmen verankern

Thomas Keßler, Genie Enterprise Deutschland & bitExpert

Die Implementierung von Künstlicher Intelligenz (KI) ist für Unternehmen aller Branchen längst keine ferne Zukunftsmusik mehr, sondern eine aktuelle Notwendigkeit, um wettbewerbsfähig zu bleiben und neue Potenziale zu erschließen. Doch der Weg zu einer erfolgreichen KI-Nutzung ist oft mit Herausforderungen gepflastert. Dieser Artikel beleuchtet, wie Unternehmen KI systematisch und gewinnbringend einsetzen können. Er analysiert typische Hürden, stellt ein praxiserprobtes Reifegradmodell vor, definiert notwendige Fähigkeiten und skizziert konkrete Vorgehensweisen.

Die strategische Bedeutung von KI

Unternehmensstrategien legen die langfristige Ausrichtung fest und ermöglichen die Erreichung definierter Ziele. Eine spezifische KI-Strategie ist dabei unerlässlich, um die Potenziale dieser transformativen Technologie gezielt zu nutzen und nicht im Aktionismus zu verharren. Sie schafft einen Rahmen für Investitionen, Kompetenzaufbau und die Integration von KI in Geschäftsprozesse.

Typische Herausforderungen auf dem Weg zur KI-Nutzung

Unternehmen sehen sich bei der Einführung von KI oft mit ähnlichen Problemen konfrontiert:

- **Datenlandschaft als Nadelöhr:** Daten sind häufig in Silos verteilt und schwer zugänglich. Die Qualität und das Management von Daten gelten als zentrale Hürden, denn KI-Modelle sind nur so gut wie ihre Datengrundlage. Ein optimiertes Informationsmanagement und eine sorgfältige Datenbereinigung sind entscheidend, um diese Herausforderungen zu bewältigen.
- **Organisatorische Trägheit und Kultur:** Traditionelle, risikoscheue Organisationsstrukturen und etablierte Prozesse können KI-Initiativen ausbremsen. Organisationaler Widerstand, oft aus Sorge vor Arbeitsplatzverlust, ist ein signifikanter Faktor. Die Entwicklung einer agilen, fehlerfreundlichen und innovationsfördernden Kultur ist daher essenziell.
- **Der „Skill Gap“:** Fehlendes KI-Know-how: Ein Mangel an Mitarbeitenden mit fundierten KI- und Data-Science-Kenntnissen ist eine weit verbreitete Herausforderung. Die Nachfrage nach KI-Skills steigt rasant und viele Unternehmen haben Defizite bei der internen Weiterbildung.
- **Fehlende Vision und ungesteuerter Aktionismus:** Ohne eine klare Vision und Strategie für den KI-Einsatz kommt es oft zu isolierten Einzelprojekten oder einer ungesteuerten Nutzung von KI-Tools durch einzelne Mitarbeitende, ohne dass ein echter Unternehmenswert entsteht. Mangelnde Ausrichtung von Talenten, Daten und Technologien auf priorisierte Anwendungsfälle behindern den Erfolg.

Das KI-Reifegradmodell: Den eigenen Standpunkt bestimmen

Um den Weg zur erfolgreichen KI-Nutzung strukturiert anzugehen, hat sich die Orientierung an einem Reifegradmodell bewährt *(siehe Abbildung 1)*. Dieses Modell hilft Organisationen, ihren aktuellen Stand der KI-Implementierung und -Nutzung zu bewerten und die notwendigen Schritte zur Weiterentwicklung zu definieren.

1. **KI-Bewusstsein:** Individuelle, oft unkoordinierte Nutzung von KI-Tools. Das Unternehmen steht am Anfang seiner KI-Reise.
2. **Experimentell:** Erste Pilotprojekte werden gestartet, oft in einzelnen Teams, und KI-Prototyping findet statt.
3. **Funktionale Integration:** KI-Lösungen sind in einzelnen Geschäftsbereichen oder spezifischen Verfahren erfolgreich implementiert und etabliert.
4. **Erweiterte Kompetenzen:** KI wird systematisch eingesetzt, Prozesse sind standardisiert und werden kontinuierlich optimiert. Das Unternehmen verfügt über breite KI-Kompetenzen.
5. **Transformative Nutzung:** KI bildet die Grundlage für neue, datengetriebene Geschäftsmodelle und verändert das Unternehmen fundamental.

Es ist zu erwarten, dass eine starke Korrelation zwischen dem KI-Reifegrad eines Unternehmens und dessen Geschäftserfolg zu beobachten sein wird. Interessanterweise weisen aktuelle Zahlen auf einen leichten Rückgang des durchschnittlichen Reifegrads hin, was die Komplexität der nachhaltigen KI-Integration unterstreicht, und die Notwendigkeit strukturierter Ansätze hervorhebt.

Fundament für den Erfolg: Notwendige Fähigkeiten und Abhängigkeiten

Die erfolgreiche Etablierung von KI im Unternehmen erfordert den Aufbau und die Orchestrierung von Fähigkeiten in vier zentralen Bereichen *(siehe Abbildung 2)*:

- **Struktur (Organisatorische Fähigkeiten):**
  Ein klar definiertes Management-Konzept für KI ist entscheidend, um Prozesse transparent und effizient zu gestalten. Dazu gehört auch die sorgfältige Überwachung von Berechtigungen und Zugängen, um Daten und Modelle sicher zu schützen. Ebenso wichtig ist die strukturierte Organisation von KI-bezogenen Zuständigkeiten und Rollen, um Verantwortungsbereiche klar abzugrenzen. Ein weiterer Schwerpunkt liegt im systematischen Umgang mit dem KI-Anforderungs- und Technologieportfolio, um langfristige Strategien zu ermöglichen.
- **Technik (Technologiebezogene Fähigkeiten):**
  Die erfolgreiche Umsetzung von KI-Projekten erfordert zunächst die Be-

Abbildung 1: Das KI-Reifegradmodell (Quelle: Thomas Keßler)

reitstellung eines passenden Software-Stacks, der die notwendigen Tools und Plattformen für Modellentwicklung und -ausführung umfasst. Dazu gehört auch die Sicherstellung der zugrunde liegenden Hardware und Infrastruktur, wobei sowohl On-Premise-Lösungen als auch Cloud-Angebote flexibel kombiniert werden müssen. Ebenso entscheidend ist die Gewährleistung von Datenverfügbarkeit und -qualität, die durch robuste Data-Governance-Maßnahmen und effektive Datenintegration sichergestellt werden muss, um die Grundvoraussetzungen für leistungsstarke KI-Systeme zu schaffen.

- **Menschen (Soziokulturelle Fähigkeiten):**
  Die Etablierung eines unterstützenden Wertesystems und einer KI-freundlichen Kultur bildet die Grundlage für nachhaltige Innovation. Dazu gehört auch das aktive Einbinden von Stakeholdergruppen, um Interessen und Anforderungen transparent und kooperativ abzustimmen. Ebenso zentral ist die Entwicklung von Personen durch gezielte Aus- und Weiterbildungsmaßnahmen, die den notwendigen Kompetenzzuwachs sichern. Ein weiterer Schwerpunkt liegt im Empowerment der Mitarbeitenden, die durch gezielte Unterstützung und Ressourcen in die Lage versetzt werden, KI aktiv zu nutzen und eigenständig Kompetenzen aufzubauen.
- **Anwendung (Anwendungsbezogene Fähigkeiten):**
  Die erfolgreiche Integration von KI beginnt mit der Identifikation und einem klaren Verständnis relevanter Use-Cases, um Anwendungspotenziale präzise zu erfassen. Darauf aufbauend ist eine professionelle Implementierung von KI-Lösungen erforderlich, die technische und fachliche Anforderungen gleichermaßen berücksichtigt. Ein weiterer zentraler Aspekt ist das Management des gesamten KI-Lebenszyklus, um die nachhaltige Entwicklung, Bereitstellung und Überwachung von Modellen sicherzustellen.

Konkrete Vorgehensweisen zur Entwicklung der KI-Reife und -Strategie

Abhängig vom individuellen Reifegrad und den Unternehmenszielen können verschiedene Ansätze zur Entwicklung der KI-Strategie und zur Umsetzung von Projekten gewählt werden:

- **KI-Lotse:** Leichtgewichtige Workshops zur Sensibilisierung, Identifikation erster Einsatzgebiete sowie Chancen und Risiken. Vermittlung von Grundlagen des Reifegradmodells und etablierter KI-Tools
- **Konzeptentwicklung und Pilotierung:** Systematische Konzepterstellung, beginnend mit einem Assessment des aktuellen Reifegrads, Identifikation und Durchführung von Pilot-Use-Cases bis hin zur Begleitung von Umsetzungsprojekten und der Ausbildung interner Multiplikatoren.
- **Entwicklung KI-Strategie:** Vertiefende Workshops je Reifegrad-Dimension zur Erstellung eines detaillierten Reifegrad-Reports und zur Ableitung konkreter Entwicklungsschritte.
- **Parallelisierung von Strategie und Praxis:** Es ist nicht immer notwendig, zuerst eine vollumfängliche Strategie zu entwickeln, bevor erste Projekte starten. Oft ist eine parallele Vorgehensweise sinnvoll: Während Pilotprojekte erste Erfahrungen und Erfolge liefern, wird die übergeordnete KI-Strategie iterativ entwickelt und verfeinert. Dieser Ansatz wird auch von aktuellen Forschungserkenntnissen gestützt, die dazu raten, klein anzufangen und dann zu skalieren.

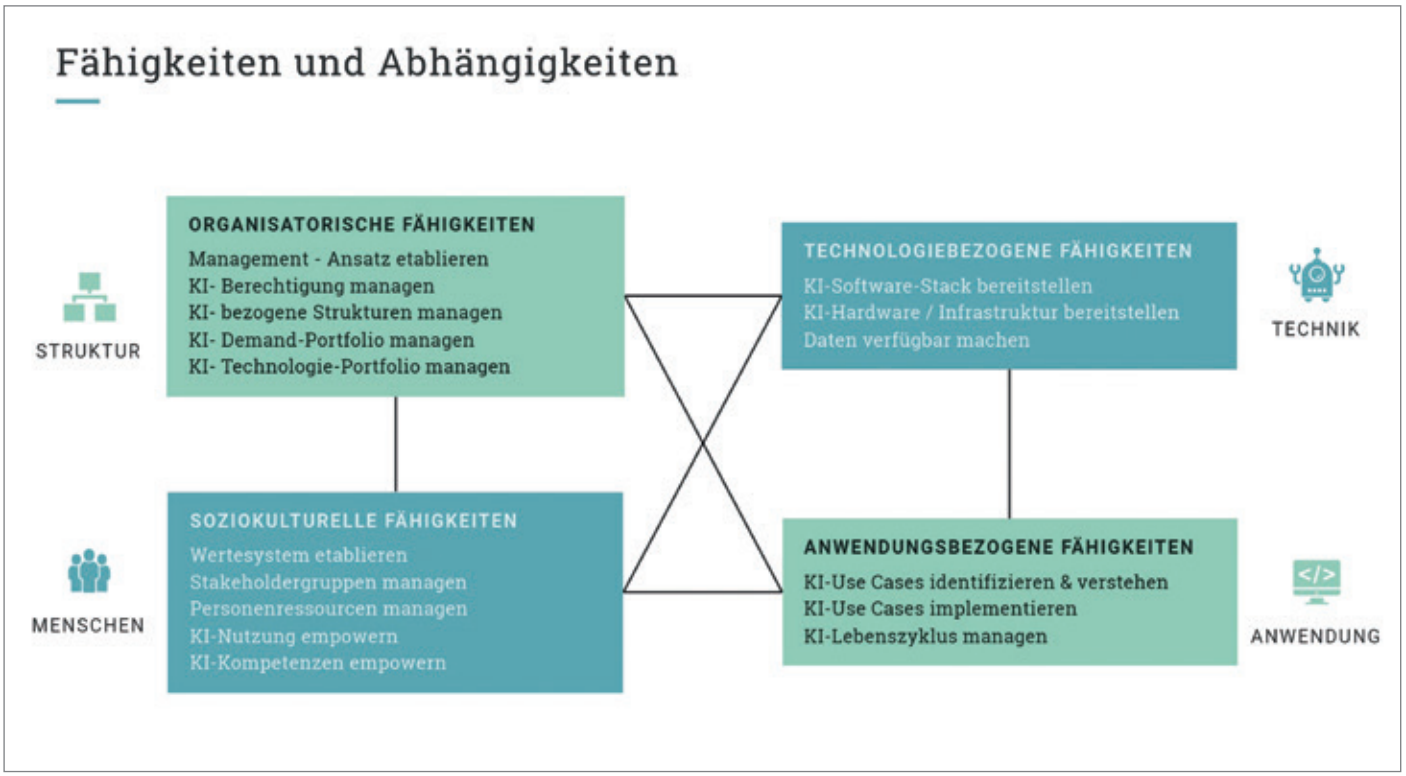


Abbildung 2: Der Aufbau und die Orchestrierung von Fähigkeiten in vier zentralen Bereichen (Quelle: Thomas Keßler)

Die Rolle von Ethik und Compliance nicht unterschätzen

Ein oft vernachlässigter, aber für den langfristigen Erfolg und die Akzeptanz von KI entscheidender Aspekt ist die Berücksichtigung von Ethik und Compliance. AI Governance Frameworks, Risikobewertungen und interne Ethikrichtlinien gewinnen an Bedeutung, um Vertrauen zu schaffen und regulatorische Anforderungen (z. B. EU AI Act) zu erfüllen.

Fazit: Jetzt die Weichen für eine KI-gestützte Zukunft stellen

Die Einführung und Skalierung von Künstlicher Intelligenz ist eine komplexe, aber lohnende Aufgabe:

a) **Jetzt loslegen:** Ob zuerst die Strategie verfeinert wird, erste Pilotprojekte gestartet werden oder beides parallel erfolgt – entscheidend ist, den ersten Schritt zu tun und Momentum aufzubauen.
b) **Alle Aspekte berücksichtigen:** Eine erfolgreiche KI-Strategie ist mehrdimensional und umfasst technologische, organisatorische, kulturelle und ethische Aspekte – nicht nur den reinen Fokus auf Technik oder ROI.
c) **Unterstützung nutzen:** Der Weg muss nicht allein beschritten werden. Externe Expertise durch KI-Lotsen, Berater bei der Umsetzung von Piloten oder bei der Erarbeitung der Gesamtstrategie kann wertvolle Impulse liefern und helfen, Fallstricke zu vermeiden.

Unternehmen, die diese Aspekte berücksichtigen und einen strukturierten, an ihrem individuellen Reifegrad orientierten Weg einschlagen, können die Potenziale der Künstlichen Intelligenz systematisch erschließen und gewinnbringend für ihre Zukunft nutzen.

Über den Autor

Thomas Keßler ist als CTO bei Genie Enterprise tätig, einem KI-Thinktank, der sich auf die Forschung an KI-Verfahren und die Entwicklung von KI-Systemen für Unternehmen spezialisiert hat. Darüber hinaus verantwortet er den KI-Bereich der bitExpert AG, spezialisiert auf die Entwicklung von individuellen Webapplikationen und Geschäftsanwendungen, einschließlich Lösungen in den Bereichen ERP, CRM, KI und BI, unter Einsatz von Low-Code und Pro-Code-Technologien. Mit seiner Expertise in verschiedenen Branchen wie Finanzen, LegalTech und Medizin treibt er die systematische und gewinnbringende Implementierung von KI-Lösungen voran.

Thomas Keßler
tkessler@genie-enterprise.com
t.kessler@bitexpert.de

Oracle Datenbanken Monthly News

Auf dem deutschsprachigen Oracle-Blog ist die Juli-Ausgabe der News-Serie erschienen.

DOAG Online

Es ist wieder so weit: die neue Ausgabe ist online! Das sechsköpfige Redaktionsteam von Oracle Deutschland hat wieder Neuigkeiten rund um die Oracle-Datenbank für On-Premises und Cloud-Installation zusammengestellt.

Alles wird wieder in einem Video präsentiert.

In der aktuellen Ausgabe wird wieder ein zusätzliches Quick Link Posting (in Englisch) zur Verfügung gestellt, um einen schnellen Zugriff auf die zugehörigen Beiträge zu gewährleisten.

https://www.doag.org/de/home/news/oracle-datenbanken-monthly-news-47/

„KI als neuer Mitarbeiter" – Wie Large Language Models Arbeitsprozesse transformieren

Christian Harms, merlin.zwo digital concept

Large Language Models (LLMs) wie GPT-4, LLaMA, Claude oder Mistral haben sich innerhalb kurzer Zeit zu leistungsfähigen Werkzeugen entwickelt, um textbasierte Prozesse im Unternehmenskontext neu zu denken. Ihr Mehrwert zeigt sich besonders dort, wo aus Freitext strukturierte Informationen generiert werden sollen – sei es in der Kundenkommunikation, der Verwaltung oder im medizinischen Bereich.

Was LLMs dabei von klassischen regelbasierten Ansätzen unterscheidet, ist nicht die Tatsache, dass sie Kategorien zuweisen oder Entscheidungen vorbereiten – sondern wie sie das tun: Sie arbeiten auf Basis eines tieferen Sprach- und Kontextverständnisses, nicht auf Grundlage starrer, manuell codierter Regeln. Die resultierende Klassifizierung wirkt dadurch nicht „intelligent" – sie ist es, weil sie auf sprachlich-logischen Zusammenhängen und semantischer Interpretation beruht.

In unseren Projekten ging es explizit darum, strukturierte Ausgaben zu erzeugen: E-Mails sollten nach Thema, Dringlichkeit und Stimmungslage eingeordnet werden. Medizinische Texte sollten konkrete Entitäten wie Diagnose, Maßnahme oder Fachbereich liefern. Das Ziel war also eine systematische, wiederholbare Klassifikation – jedoch auf einem Niveau, das klassische Automatisierung nicht leisten kann.

Die zentrale Steuerungskomponente dabei ist das *Prompt-Engineering.* Statt Logik in Code zu gießen, wird sie sprachlich modelliert. Anforderungen, Einschränkungen, Formate und Beispiele werden in einem Prompt formuliert, der das Modell präzise instruiert – und je nach Domäne iterativ geschärft werden muss.

Dieser Artikel stellt zwei praxiserprobte Szenarien vor:

Zum einen die automatisierte Vorverarbeitung von Support-Mails mit Sentiment-Analyse und zum anderen die Analyse medizinischer Akten mit einem lokal betriebenen LLM – inklusive datenschutzkonformer Architektur. Beide Anwendungsfälle zeigen, wie sich LLMs gezielt in produktive Workflows integrieren lassen, wenn man ihre Stärken richtig adressiert: Kontextverstehen, semantische Einordnung und flexible, aber kontrollierbare Ausgabeformate.

Praxisbeispiel 1: Kundenstimmung erkennen – Sentiment-Analyse im Supportpostfach

Ein eindrucksvolles Beispiel für den erfolgreichen Einsatz von KI in realen Geschäftsprozessen ist ein Projekt, das wir gemeinsam mit einem Kunden aus dem IT-Dienstleistungssektor als Proof-of-Concept (PoC) umgesetzt haben. Ziel war es, die Bearbeitung eingehender E-Mails im zentralen Supportpostfach durch den gezielten Einsatz von LLMs zu beschleunigen und qualitativ zu verbessern.

In klassischen Supportstrukturen ist das Postfach oft ein Nadelöhr: Jeden Tag treffen dort zahlreiche Anfragen ein, deren Relevanz und Dringlichkeit zunächst manuell eingeschätzt werden müssen. Die Folge: hohe Reaktionszeiten, unklare Prioritäten und mitunter übersehene kritische Eskalationen. Hier setzt unsere Lösung an – mit einer intelligenten Textanalyse, die automatisiert die inhaltlichen Eckpunkte einer E-Mail erkennt und die Stimmung des Absenders interpretiert.

Mithilfe eines sorgfältig entwickelten Prompt-Engineerings analysiert ein LLM jede eingehende Nachricht, beziehungsweise gegebenenfalls die gesamte vorhergegangene Korrespondenz und extrahiert dabei relevante Informationen wie:

- **Kategorie des Anliegens** (zum Beispiel technisches Problem, Anfrage, Beschwerde),
- **Priorität** (niedrig, mittel, hoch),

- **Status** (offen, gelöst, Rückfrage erforderlich),
- **Tonalität und Stimmung** (neutral, unzufrieden, verärgert).

Besonders spannend ist die automatische Einschätzung der Kundenzufriedenheit anhand sprachlicher Nuancen. So erkennt das Modell beispielsweise verschärfte Formulierungen, Wiederholungen von Beschwerden oder einen fordernden Ton – und kann darauf basierend Hinweise auf potenzielle Eskalationen geben. Diese Informationen werden in einem strukturierten Format an nachgelagerte Prozesse übergeben und helfen dem Support-Team dabei, gezielt und priorisiert zu reagieren.

Die Vorteile dieser Lösung sind bereits im PoC deutlich geworden:

- Reaktionszeiten konnten verkürzt werden, weil dringende Fälle automatisiert nach oben priorisiert wurden.
- Supportmitarbeiter erhielten eine fundierte Vorbewertung, ohne selbst lange Texte lesen zu müssen.
- Die Stimmungserkennung ermöglichte ein frühzeitiges Eingreifen bei drohender Unzufriedenheit.

Aktuell wird das System produktiv eingeführt und in bestehende Workflows integriert – mit dem Ziel, langfristig nicht nur die Effizienz zu steigern, sondern auch die Servicequalität messbar zu verbessern. Der Erfolg des Projekts zeigt: LLMs sind in der Lage, Sprache nicht nur zu verarbeiten, sondern wirklich zu verstehen – und das macht sie zu einem mächtigen Werkzeug im Kundenkontakt.

Praxisbeispiel 2: Strukturierte Informationen aus komplexen Dokumenten – KI in der medizinischen Aktenanalyse

Ein besonders spannender und anspruchsvoller Anwendungsfall für den Einsatz von Large Language Models ist die Analyse medizinischer Patientenakten. Gemeinsam mit einem Kunden aus dem Gesundheitswesen haben wir im Rahmen eines Proof-of-Concepts eine Lösung entwickelt, um medizinische Dokumente automatisiert zu analysieren und Informationen strukturiert zu extrahieren, sowie dem zuständigen Fachbereich zur Verfügung zu stellen.

Medizinische Texte wie OP-Berichte, Entlassungsschreiben oder klinische Verlaufsdokumentationen umfassen häufig Dutzende bis Hunderte von Seiten. Sie sind unstrukturiert, verwenden hochspezialisierte Fachsprache und enthalten kritische Informationen, die für nachgelagerte Prozesse – zum Beispiel Abrechnung, Qualitätssicherung oder medizinische Auswertung – zentral sind. Die manuelle Aufbereitung dieser Dokumente durch menschliche Mitarbeiter ist ressourcenintensiv und fehleranfällig.

Ziel unseres Ansatzes war es, mithilfe eines LLM aus solchen Dokumenten automatisch die wichtigsten Kopfdaten zu extrahieren – etwa die Fallnummer, Patientendaten, Diagnosen, durchgeführte Maßnahmen, Behandlungen und – insbesondere – daraus den zuständigen Fachbereich zu ermitteln. Die große Herausforderung bestand nicht nur im Erkennen dieser Inhalte, sondern in ihrer präzisen, strukturierten und validierbaren Aufbereitung, sodass sie anschließend in bestehende Prozesse und Systeme integriert werden können.

Datenschutz als zentrales Entscheidungskriterium

Von Beginn an war klar: In einem medizinischen Kontext steht der Schutz personenbezogener und besonders sensibler Gesundheitsdaten an oberster Stelle. Die Nutzung eines öffentlichen Online-Modells – etwa über eine Cloud-basierte API – war daher ausgeschlossen. Die Anforderung lautete: Die Daten dürfen das Unternehmen zu keinem Zeitpunkt verlassen.

Die Lösung: Der Einsatz eines lokal betriebenen LLMs, das innerhalb der eigenen IT-Infrastruktur betrieben wird. Dadurch bleiben alle Daten vollständig unter der Kontrolle des Unternehmens. Gleichzeitig ermöglicht dieses Setup maximale Flexibilität bei der Anpassung des Modells und bei der Integration in bestehende Sicherheitssysteme und Prozesse.

Natürlich bringt ein lokales Modell zusätzliche Herausforderungen mit sich – insbesondere hinsichtlich Ressourcenbedarf und technischer Komplexität. Doch der Datenschutz, sowie der Zugewinn an Datensouveränität und Compliance war für unseren Kunden ausschlaggebend. Zudem zeigt die Entwicklung im Bereich Open-Source-LLMs, dass leistungsfähige Modelle mittlerweile auch lokal einsetzbar sind – nahezu ohne Abstriche bei der Qualität.

Der Weg zum funktionierenden Prompt

Technisch gesehen waren Modelle wie LLaMA und Qwen out of-the-box in der Lage, medizinische Inhalte zu analysieren. Entgegen der Erwartung ein ressourcenintensives Finetuning durchführen zu müssen, konnten wir also direkt mit den Modellen arbeiten. Doch erst durch ein systematisch entwickeltes Prompt-Engineering wurden die Ergebnisse praxistauglich. Die Lernkurve dabei war steil. Es wurde schnell deutlich, dass ein erfolgreicher Einsatz nicht nur technisches Know-how, sondern auch fachliches Verständnis und viel Iteration erfordert.

Folglich flossen erhebliche Anstrengungen in die Entwicklung eines präzisen, robusten Prompts. Dabei waren fünf Faktoren entscheidend:

- natürlichsprachliche, klar formulierte Anforderungen,
- eine hohe Spezifität der Vorgaben,
- explizite Ausschlüsse möglicher Fehlerquellen,
- die Arbeit mit konkreten Beispieltexten und
- genaue Formatvorgaben für die Ausgabe.

Nach mehreren Optimierungszyklen konnte das Modell zuverlässig strukturierte Daten liefern – zum Beispiel zur Klassifizierung von Diagnosen, Zuordnung von Fachbereichen oder Erkennung von operativen Maßnahmen. Besonders hilfreich war die Möglichkeit, Negativbeispiele zu definieren und das Modell gezielt auf Ausnahmefälle vorzubereiten.

Die Sache mit der Kontextgröße

Die Verarbeitungskapazität von Large Language Models (LLMs) ist durch eine begrenzte Kontextlänge eingeschränkt. Es liegt auf der Hand, dass umfangreiche Dokumente, etwa hunderte Seiten lange PDFs, nicht vollständig und in einem Schritt an ein LLM übermittelt werden können. Konventionelle Ansätze wie Retrieval-Augmented-Generation (RAG),

die darauf abzielen, relevante Informationen aus großen Dokumenten mithilfe von LLMs zu extrahieren, stoßen hierbei schnell an ihre Grenzen. Dies zeigt sich insbesondere bei medizinischen Unterlagen, wie Patientenakten, in denen sich auf nahezu jeder Seite zahlreiche personenbezogene Angaben wie Namen und Adressen finden. Doch welche dieser Informationen gehören tatsächlich zum Patienten?

Zur Lösung dieses Problems wurde ein zweistufiges Vorgehen entwickelt: Zunächst wird das vollständige PDF seitenweise durch ein LLM analysiert, um eine strukturierte Übersicht der enthaltenen Informationen zu generieren. Auf Basis dieser Vorverarbeitung können anschließend gezielt die relevanten Bereiche identifiziert und dem LLM zur Extraktion spezifischer Inhalte – etwa des Patientennamens oder der Diagnosen – übergeben werden.

Das Modell der Wahl

Die Auswahl eines geeigneten Large Language Models geht weit über einen reinen Leistungsvergleich hinaus. Entscheidend ist ein fundiertes Verständnis der jeweiligen Stärken und Grenzen – etwa in Bezug auf Sprachverständnis, Genauigkeit und Kontextverarbeitung. Eine zentrale Rolle spielt zudem der Datenschutz: In sensiblen Anwendungsbereichen wie der Medizin oder dem Rechtswesen ist der Einsatz eines lokalen Modells oft unumgänglich, um regulatorische Anforderungen wie die DSGVO zuverlässig zu erfüllen.

Cloud-basierte LLMs – beispielsweise von OpenAI – bieten hohe Genauigkeit, kontinuierliche Weiterentwicklung und eine einfache Integration über standardisierte APIs. Gleichzeitig erfordern sie jedoch eine sorgfältige Prüfung hinsichtlich der Datensicherheit und -übermittlung, insbesondere bei personenbezogenen Informationen. Lokale Modelle wie LLaMA oder Qwen ermöglichen den Betrieb in vollständig kontrollierten Umgebungen und bieten maximale Datenhoheit. Sie setzen allerdings eine leistungsfähige Infrastruktur, technisches Know-how sowie einen erhöhten betrieblichen Aufwand voraus.

Grundsätzlich gilt: Je sensibler die Daten und je strenger die Compliance-Vorgaben, desto stärker spricht die Entscheidung für ein lokal betriebenes LLM. Bei allgemeinen Anwendungsfällen ohne besondere Datenschutzanforderungen überwiegen hingegen oft die Vorteile cloudbasierter Lösungen.

Der Schlüssel zur Qualität: Der richtige Prompt

Der Einsatz eines Large Language Models steht und fällt mit der Qualität der Anweisungen, die es erhält. Anders als bei klassischer Programmierung definiert man bei LLMs nicht einen festen Ablauf, sondern beschreibt möglichst präzise, was das Modell tun soll – und wie das Ergebnis aussehen soll. Dieses sogenannte *Prompt-Engineering* ist der zentrale Stellhebel für zuverlässige, konsistente und praxisnahe Resultate.

Im Rahmen unseres medizinischen Anwendungsfalls hat sich gezeigt, dass ein leistungsfähiger Prompt das Ergebnis systematischer Arbeit ist. Fünf Aspekte haben sich dabei als besonders erfolgskritisch erwiesen:

1. Natürlichsprachliche und strukturierte Anforderungen

Der Prompt muss in klarer, verständlicher Sprache formuliert sein – so, wie man auch einem Menschen eine Aufgabe erklären würde. Dabei sollte der Fokus auf Verständlichkeit und Eindeutigkeit liegen. Unklare Formulierungen oder mehrdeutige Begriffe führen schnell zu ungenauen Ergebnissen. Es hat sich als hilfreich erwiesen, die Anforderungen als Aufzählung zu formulieren. Am Beispiel der Patientenakte:

„Ermittle aus dem folgenden Text diese Informationen:

1. *„NAME_KLINIK“: Name der Klinik, in der der Patient behandelt worden ist.*
2. *„NAME_PATIENT“: Name des behandelten Patienten.*
3. *„ADRESSE_PATIENT“: Adresse des behandelten Patienten.*
4. *„GEBURTSDATUM_PATIENT“: Geburtsdatum des behandelten Patienten.*

2. Spezifität der Vorgaben

Je genauer der Prompt beschreibt, welche Informationen gesucht sind, desto besser sind die Resultate. Allgemeine Fragen wie „Was steht im Text?“ führen zu unbrauchbaren Ausgaben. Stattdessen sollte genau festgelegt werden, welche Inhalte extrahiert werden sollen – beispielsweise: *„Nenne die durchgeführte Maßnahme im medizinischen Eingriff und gib zusätzlich die zugehörige Diagnose an.“*

Wichtig sind zudem Vorgaben und Regeln, wo die Informationen zu finden sind. So befinden sich in einem Arztbrief Namen und Adressangaben im Briefkopf und in der Fußzeile, hierbei handelt es sich aber nicht um Informationen zum Patienten.

3. Explizites Ausschließen von Fehlern

Ein oft unterschätzter Faktor: Es reicht nicht, nur zu sagen, was das Modell tun soll, man muss auch klar benennen, was es nicht tun darf. Beispielsweise: „Gib keine Vermutungen ab“, „Vermeide Interpretationen bei unklaren Formulierungen“, oder „Wiederhole keine Inhalte, die bereits genannt wurden.“ Diese Negativbedingungen erhöhen die Konsistenz der Ausgaben erheblich.

4. Arbeiten mit Beispielen

LLMs arbeiten zwar auch allein mit expliziten Anweisungen sehr gut, jedoch profitiert die Qualität der Ergebnisse auch erheblich von Beispielen. Daher ist es hilfreich, dem Prompt konkrete Beispieltexte inklusive der gewünschten Ausgabeformate beizufügen. Auf diese Weise kann das Modell besser abstrahieren, was gemeint ist – und das Vorgehen auf andere Texte übertragen.

5. Formatvorgaben und Struktur

Ein weiterer kritischer Punkt ist das gewünschte Ausgabeformat. LLMs sind flexibel – manchmal zu flexibel. Wird nicht klar definiert, in welcher Struktur die Antwort zurückgegeben werden soll (zum Beispiel als JSON, Tabelle, Listenformat oder Klartext mit festen Überschriften), entstehen uneinheitliche Ausgaben, die nur schwer automatisiert weiterverarbeitet werden können. Beispiel:

„Erstelle ein JSON-Objekt mit folgenden Attributen: x, y, z, … Du darfst keine Anführungszeichen als Präfix oder Suffix verwenden und auch sonst keine einleitenden Worte in deiner Antwort ausgeben.“

Diese fünf Prinzipien waren das Ergebnis intensiver Arbeit mit dem Modell und

zahlreicher Testläufe. Mit dieser Vorgehensweise konnten wir verlässliche und reproduzierbare Ergebnisse erzielen – sowohl im medizinischen als auch im E-Mail-Szenario.

Fazit und Ausblick: KI als Mitgestalter der digitalen Arbeitswelt

Die beiden Praxisbeispiele zeigen eindrucksvoll, wie LLMs nicht nur theoretisch faszinieren, sondern ganz konkret Mehrwert im Arbeitsalltag schaffen. Ob im Supportbereich oder im Gesundheitswesen: Überall dort, wo Sprache verarbeitet, Informationen extrahiert und Entscheidungen vorbereitet werden müssen, kann KI zum produktiven Mitgestalter werden.

Dabei kommt es weniger auf die reine Rechenleistung an, sondern vielmehr auf die richtige Herangehensweise: Ein gut durchdachter Prompt ist oft wertvoller als ein noch leistungsfähigeres Modell. Die Erfahrung aus unseren Projekten zeigt, dass ein enger Austausch mit den Fachabteilungen, iterative Tests und das kontinuierliche Nachschärfen der Anforderungen essenziell sind, um von der beeindruckenden Leistungsfähigkeit der LLMs zu profitieren.

Der Blick in die Zukunft ist vielversprechend: Mit der wachsenden Verfügbarkeit lokaler LLM-Modelle, verbesserter Datenschutzmechanismen und leistungsstarker APIs rücken noch mehr Anwendungsfelder in greifbare Nähe. Gleichzeitig entstehen neue Herausforderungen – etwa bei der Validierung von Ergebnissen, der Integration in bestehende Systeme oder der Akzeptanz durch Anwender.

Doch eines ist bereits heute klar: KI ist kein Ersatz für menschliche Expertise – sie ist ein mächtiges Werkzeug, das Menschen entlastet, unterstützt und ihnen hilft, sich auf das Wesentliche zu konzentrieren. Und genau deshalb lohnt es sich, Künstliche Intelligenz als das zu begreifen, was sie zunehmend wird: ein neuer, wertvoller Mitarbeiter im digitalen Team.

Über den Autor

Christian Harms ist seit über 20 Jahren als Berater für Oracle-Datenbanken, Data-Warehouse-Systeme und Business Intelligence tätig. Sein aktueller Schwerpunkt liegt auf dem Einsatz Künstlicher Intelligenz und Large Language Models in Verbindung mit komplexen Datenbanksystemen, um bestehende Kundensysteme gezielt zu erweitern und manuelle Prozesse zu automatisieren.

Seine Erkenntnisse teilt er regelmäßig als Sprecher auf Fachkonferenzen und als Autor von Fachbeiträgen.

Christian Harms
christian.harms@merlin-zwo.de

SUPER-SAVER

Bis 31.10.2025

Heide Park Soltau

APEX connect

18. - 20. Mai 2026

+

DOAG 2026 Datenbank

mit Cloud Infrastructure

18. - 19. Mai 2026

It's the Data, Stupid! – Wie wir den Erfolg von KI-Initiativen greifbar machen können

Arne Wellnitz

In deutschen Unternehmen sind KI-Initiativen oft von Misserfolgen gezeichnet, die sich auf eine Vielzahl von Hürden zurückzuführen lassen. Dieser Artikel betrachtet häufig anzutreffende Archetypen, die uns auf dem teils langen Weg zum KI-Erfolg begegnen und zeigt, wie wir mit ihnen umgehen können, um ein Gelingen unserer KI-Ansätze dennoch in greifbare Nähe zu rücken.

Was ist eigentlich unsere Motivation für den Einsatz von KI in Unternehmen? Natürlich mag sich diese im Detail unterscheiden, aber es gibt bestimmt einen abstrakten gemeinsamen Nenner. Alle Unternehmen, getrieben von Wettbewerb und Innovationsdruck, suchen kontinuierlich nach Möglichkeiten, ihre komplexen Geschäftsprozesse qualitativ und quantitativ effizienter zu machen. Zudem haben Unternehmen zahlreiche, zum Teil schwerwiegende Entscheidungen zu treffen, welche – insbesondere in einer zunehmend digitalen Welt – durch große Datenmengen unterstützt werden können. Des Weiteren möchten wir die menschlichen Arbeitskräfte auch gerne von langweiligen, sich wiederholenden Routineaufgaben entbinden, die immer wieder dem gleichen Schema beim Auswerten dieser Daten folgen. Das Potenzial KI-basierter Automatisierung erschöpft sich dabei längst nicht bei einer linearen Verbesserung einzelner Arbeitsschritte, sondern kann Organisationen die notwendige Flexibilität und Skalierbarkeit geben, die sogar in kurzer Zeit Produktivitätssteigerungen mehrerer Magnituden ermöglicht. Ob eine Firma beispielsweise im nächsten Quartal drei weitere Kundenaufträge annehmen kann, ob sie auch nach der Implementierung drei weiterer Features in ihrer Software noch eine Wartbarkeit gewährleisten kann, oder ob sie dabei an die Grenzen ihrer „Bio"-Arbeitskräfte stößt, hängt von eben diesen Faktoren ab. Wir möchten also eine datengetriebene Entscheidungshilfe, Produktivitätssteigerung und Kostensenkung, die Entbindung von repetitiven Routineaufgaben sowie Skalierbarkeit von Geschäftsprozessen.

Wir wissen relativ genau, was wir von der KI erwarten, aber warum ist die Umsetzung häufig so schwierig? Im Folgenden betrachten wir sechs Archetypen, die uns den Weg zum KI-Erfolg erschweren und vielleicht dem ein oder anderen schon begegnet sind.

„Archetypen" nach Carl Gustav Jung sind universelle, wiederkehrende Urbilder oder Symbole, die tief im kollektiven Unbewussten der Menschheit verankert sind und sich in Mythen, Geschichten, Träumen und Kulturen weltweit zeigen. Sie repräsentieren grundlegende menschliche Erfahrungen und Rollen wie den Helden, die Mutter oder den Weisen.

Die Kims : Jedes Team ein eigenes Königreich

Als Erstes haben wir die Kims. Benannt nach der nordkoreanischen Herrscherfamilie, führen die Kims ein unbeschwertes Leben in ihrem hermetisch abgeriegelten Königreich *(siehe Abbildung 1)*.

Abbildung 1: Kim Jong-un (Quelle: Grok-Generation)

Es sind Teams, die schon seit Jahrzehnten die Verantwortung für strategisch wichtige Unternehmensprozesse haben, welche sich jedoch im Laufe der Zeit kaum verändert haben. Entsprechend arbeiten sie schon seit Ewigkeiten mit den gleichen Datenschemata. Über die Jahre sind in diese Datenschemata mehr Ausnahmen als Regeln hineingewachsen. Letztlich ist das aber nur schwer nachzuvollziehen, denn der letzte Stand der Datendokumentation befindet sich noch irgendwo auf einer Diskette, ganz tief im Schrank neben dem ausgemusterten Faxgerät – oder „Ersatz-Faxgerät", wie man in einem deutschen Amt sagen würde. Diese Teams haben während der Umstellung auf Windows 98 nie damit gerechnet, dass irgendwann mal so ein paar Data Scientists ankommen würden und nach ihren Daten fragen könnten. Und so sind jede Menge isolierter *Datensilos* entstanden, welche die optimistischen KI-Ambitionen von Unternehmen schnell auf eine harte Realität treffen lassen. Die Daten in diesen Datensilos sind häufig inkonsistent und nicht standardisiert. Ihre Qualität für Data Science und Machine Learning völlig ungeeignet. Für die erhoffte datengetriebene Entscheidungshilfe sind die Daten zu unzuverlässig. Und die Auflösung dieser Datensilos erweist sich als Mammutaufgabe.

Was gegen Datensilos zumindest langfristig hilft, lässt sich wohl am besten unter dem Begriff *Data Governance* zusammenfassen. Data Governance sind Richtlinien und Prozeduren, die unter anderem eine Verbesserung der Qualität, Konsistenz, Standardisierung, Vollständigkeit, Verfügbarkeit und Sammlung von Daten zum Ziel haben. Dies kann, im Sinne des Data-Mesh-Konzepts, mit zentralen Kontrollinstanzen und dezentraler Verantwortung bei den Datenerzeugern umgesetzt werden. In jedem Fall ist ein essenzieller erster Schritt, ein allgegenwärtiges Bewusstsein für Datenqualität zu schaffen, diese zu priorisieren und für Verbesserungen zu motivieren. Dabei ist es völlig in Ordnung, klein anzufangen. Es kann sinnvoll sein, sich zunächst darum zu bemühen, zumindest alle neuen Datenschemata den Richtlinien der Data Governance zu unterstellen, statt die gesamte Datenhistorie umzukrempeln. Zu

revolutionäre Akte scheitern hingegen meist. Insbesondere bei der Datenerzeugung können leicht umsetzbare Schritte schon einen großen Effekt auf die Datenqualität haben. Beispiele dafür wären Eingabevalidierungen bei Formularen, die Vermeidung von Freitextfeldern für kategorische Daten und ein Verbot der Zweckentfremdung von Datenspalten. Eine klar verständliche, aktuell gehaltene und allgemein zugängliche Dokumentation von Datenschemata erleichtern es den Datenspezialisten und KI-Ingenieuren enorm, zu verstehen, um was für Daten es sich handelt, welche Werte in welchen Spalten zu erwarten sind und wer für die Daten eigentlich verantwortlich ist. Aller Anfang ist schwer, aber wer die Früchte von KI und Data Science ernten möchte, wird am Problem der Datenqualität und -verfügbarkeit nicht vorbeikommen.

Die Bullshitter: „Jetzt mit KI"

Der zweite Archetyp sind die Bullshitter. Häufig anzutreffen sind sie im mittleren Management, aber auch unter den Entwicklern. Die Bullshitter nutzen den Hypebegriff „KI" als bloßes Verkaufsargument. Es gibt sie in zwei Variationen.

Der erste Typ sind die Etikettenschwindler. Sie kleben gerne trendige Begriffe auf Produkte und meinen dies sei eine clevere und legitime Marketingstrategie. In der Realität stecken hinter der „KI" in ihren Produkten simple if-then-else-Entscheidungen oder klassische Sortieralgorithmen. So wurde zum Beispiel ein Kaffeevollautomat mit „Künstlicher Intelligenz" beworben, weil die am häufigsten genutzten Kaffeeoptionen auf den vorderen Plätzen angeordnet wurden *(siehe Abbildung 2)*.

Ein weiteres Beispiel ist der Fall eines Smartphone-Herstellers, welcher seinen KI-basierten Kamera-Zoom als Optimierungstool für detailreiche Mondfotografie anpries. Dabei wurde eine visuelle Objekterkennung verwendet, die den Mond im Foto erkennen konnte – so weit so gut. Dann hat das Modell jedoch im Grunde aus hochauflösenden Mondfotos erlernte Details, welche von überall auf der Erde nahezu identisch aussehen, künstlich in das ansonsten verschwommene Foto eingefügt.

Beim zweiten Bullshitter-Typ verkommt KI und Maschinelles Lernen zum Selbstzweck. Es geht dabei längst nicht mehr darum, ein geschäftsrelevantes Problem so effektiv wie möglich zu lösen, sondern KI-Algorithmen um jeden Preis zum Teil der Lösung zu machen, auch wenn dafür einfachere und vielleicht sogar bessere Lösungen bewusst ignoriert werden.

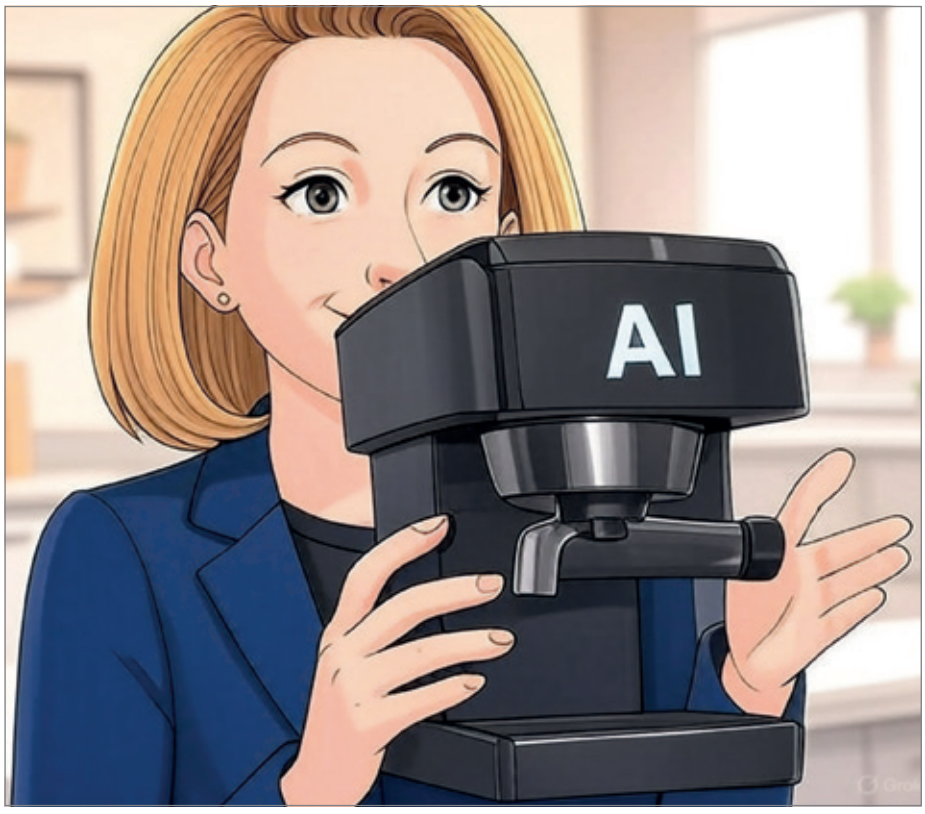


Abbildung 2: Die Bullshitter (Quelle: Grok-Generation)

Was beide Typen gemeinsam haben, ist ein in der Regel sehr rudimentäres Verständnis davon, was KI beziehungsweise ML eigentlich sind. Die meisten von ihnen haben es vermutlich nie über das Introvideo ihrer zwölf angefangenen Udemy-Kurse zum Thema KI hinausgeschafft.

Eine sinnvolle Abhilfe ist es, kritisch zu hinterfragen, ob ein Problem überhaupt dafür geeignet ist, mit KI gelöst zu werden. Hierbei helfen vier Kernfragen *[2]*: (a) Ist komplexe Entscheidungsfindung notwendig? Falls simple Regeln vollkommen ausreichend sind, um ans Ziel zu kommen, gibt es keinen Grund, mit KI-Kanonen auf Spatzen zu schießen. (b) Sind die erforderlichen Daten in ausreichender Quantität und Qualität verfügbar? Viele KI-Lösungen erfordern die Verfügbarkeit von sauberen und für die Domäne nützlichen Daten in hinreichender Menge. (c) Handelt es sich um ein Problem mit hoher beziehungsweise häufiger Arbeitslast? Aufgrund des Entwicklungsaufwands und der Kosten sollten KI-basierte Lösungen nur für häufig wiederkehrende Probleme mit bedeutendem Einfluss auf das Geschäft erwogen werden. (d) Hat eine existierende Lösung Leistungs- oder Wartbarkeitsprobleme? KI sollte insbesondere in Erwägung gezogen werden, wenn noch keine Lösung existiert, eine existierende Lösung unzureichende Ergebnisse liefert oder schlecht skalierbar ist.

Mit diesen Fragen im Hinterkopf, ist es viel einfacher, die Bullshitter zu entlarven und mehr sinnvolle Innovation mit adäquatem Einsatz von KI-Algorithmen zu erreichen.

Die Vogonen: Brandstifter, verkleidet als Feuerwehr

Vogonen sind die hochbürokratische Alienspezies aus dem Buch *„Per Anhalter durch die Galaxis"* von Douglas Adams. Sie tauchen in unserem Kontext in zwei Versionen auf: Die internen Vogonen, anzutreffen in Betriebsräten, Datenschutz-, Personal- oder Rechtsabteilungen und die externen Vogonen, anzutreffen in Politik und öffentlicher Verwaltung *(siehe Abbildung 3)*.

Die internen Vogonen haben nicht selten sehr irrationale Vorbehalte gegen den Einsatz von KI und neigen auch gerne mal zu einer maximal restriktiven Regelauslegung. Dabei sind sie typischerweise überzeugt, gar keine andere Wahl zu haben und verweisen dabei allzu gerne auf die externen Vogonen. Während es hier glücklicherweise viele Ausnahmen gibt, trifft man dennoch immer wieder auf Exemplare mit sehr hohem Mitsprachebedürfnis, welches nicht notwendigerweise auch durch ein hohes Interesse an einer tatsächlichen Lösungsfindung ausbalanciert wird. Insbesondere diese Exemplare zeichnen sich leider auch durch mangelhaftes technisches Grundlagenwissen, flache Lernkurven und ein insgesamt schlechtes Verständnis davon aus, wer im Unternehmen eigentlich das Geld verdient. In der Praxis befinden sich die Vogonen häufig in einem permanenten Tauziehen mit den Data Scientists und KI-Ingenieuren um das metaphorische Gold des Unternehmens: Die Daten.

Wie machen wir nun aus den internen Vogonen zuverlässige Partner? Es ist enorm wichtig, zunächst gegenseitiges Vertrauen zu etablieren. Unabdingbar ist hierfür ein hohes Datenschutz- und IT-Sicherheitsbewusstsein bei den Entwicklern. Wenn die Ingenieure transparent aufzeigen, dass sie Datenschutz und Sicherheit als grundlegenden Baustein ihrer Tätigkeiten betrachten, kann dies bereits viele berechtigte Sorgen beseitigen. Wichtig ist zudem auch die Vermeidung von unnötigen Sonderstellungen durch Begriffe wie KI. Hier muss klargestellt werden, dass sich die meisten Datenverarbeitungsschritte bei Methodiken des maschinellen Lernens in ihrem Wesen kaum bis gar nicht von herkömmlicher Softwareentwicklung unterscheiden. Aufklärungsarbeit und proaktive Kommunikation sorgen für Transparenz und können Vorbehalte effektiv abbauen. Dafür ist es wichtig, dass beide Seiten dieselbe Sprache sprechen, was leider nicht immer der Fall ist. Es sollte jedoch auch klar gemacht werden, dass unnötige Bürokratie und sogenanntes Red Tape verheerende Hindernisse auf dem Weg zum KI-Erfolg sein können, und soweit wie möglich vermindert werden müssen. Es würde insbesondere deutschen Unternehmen nicht schaden, Regulierungen so liberal wie möglich auszulegen und mehr aktive Lobbyarbeit gegen weitere Bevormundung zu betreiben.

Abbildung 3: Die Vogonen (Quelle: Grok-Generation)

Und das bringt uns zu dem externen Vogonen.

In Deutschland werden die Stimmen gegen überbordende Bürokratie zunehmend lauter *[3]*. Auf nahezu allen Ebenen ist ein klarer Trend zu immer mehr Auflagen, komplexeren Gesetzen und Berichtspflichten erkennbar. Die Politik legt ein fundamentales Misstrauen in jegliche Eigenverantwortung von Unternehmern an den Tag. Die politischen Vogonen maßen sich zudem immer öfter an, den Markt besser navigieren zu können als die Unternehmer und haben eine absolut irrationale Neigung zur Planwirtschaft . Verbote und Subventionen sind ihre Heilmittel für alles und es ist ihnen keine Regel zu viel, keine Verordnung zu lang und keine staatliche Kontrolle bevormundend genug.

Diese Form von kostspieliger Bürokratie und irrgeführter Regulierung zerstört offensichtlich Innovationskraft – nicht nur bei KI. Die Vorschriftenflut schafft insbesondere bei den in Deutschland rückgratbildenden KMUs große Verunsicherung und Blockadeerscheinungen. Skalierungseffekte sind hierbei maßgeblich, weil Big Tech neue Regulierungen im Zweifel verhältnismäßig einfach mit einer Legion von Anwälten bewältigen kann. Wenn dann jedoch sogar Konzerne wie Meta eine Nutzung ihrer neuesten Llama-Sprachmodelle in Europa in ihrer Lizenz untersagt bekommen oder Apple KI-basierte Funktionen (Apple Intelligence) vom europäischen Markt ausschließt, ist das kein Grund zur Freude. Schlussendlich brauchen wir mehr wettbewerbsfähige Unternehmen, die sich in freier Marktwirtschaft entfalten können, und keine planwirtschaftlich dressierten Subventionsempfänger – mehr Hayek, weniger Habeck.

Die Gatekeeper: Risikoaversion, statt Innovation

Die Gatekeeper zeichnen sich durch eine ausgeprägte Abneigung von jeglichen unternehmerischen Risiken aus, auch wenn der daraus resultierende Mangel an Innovation über kurz oder lang einen unausweichlichen Niedergang in die Bedeutungslosigkeit für sie bedeutet. Sie scheuen dabei unter anderem die Kosten und den Aufwand für die Einführung neu-

Abbildung 4: Die Gatekeeper (Quelle: Grok-Generation)

er Technologien. Sie leiden jedoch auch an einem Mangel an technischem Sachverstand und haben schlicht nicht die notwendige Führungskompetenz, um ein Team für die Entwicklung von High-Tech-Produkten zu führen *(siehe Abbildung 4)*.

Das wollen sie sich jedoch nicht unbedingt selbst eingestehen und schon gar nicht für andere sichtbar machen. Gewissermaßen resignieren die Gatekeeper aber auch einfach vor den Vogonen und übernehmen nur zu gerne deren Bedenken. Sie sorgen sich zurecht um Datensouveränität, Geschäftsgeheimnisse und schauen resigniert der nächsten Regulierungswelle entgegen.

Auch hier gilt es zuallererst, die unternehmerische Risikoaversion als Blockade für Innovation beim Namen zu nennen und Wege zu finden, um sie zu vermeiden. Das Ziel sollte ein Schritt in Richtung rationaler Innovationsfähigkeit sein. Die Betonung liegt hier auf „rational", statt radikal, denn natürlich sollte ein von Hightech getriebener Innovationsdrang nicht das Kerngeschäft untergraben. Des Weiteren besteht bei Innovationsprojekten auch die Gefahr der Sunk-Cost-Fallacy oder dem Setzen unrealistischer Ziele, wie „Wir sollten einfach unser eigenes LLM trainieren".

Die Brents: Manager mit Torschlusspanik

Die Brents sind benannt nach David Brent, der massiv unterqualifizierten Führungskraft aus der britischen TV-Serie *The Office* (Bernd Stromberg aus der gleichnamigen deutschen TV-Serie oder der Pointy-Haired Boss aus den Dilbert-Comics sind vergleichbare Figuren). Brents sind meistens Teil des Managements ohne ingenieurstechnischen Hintergrund. Sie haben typischerweise einen „Neffen", der ganz tolle Dinge mit KI-Apps auf seinem Telefon machen kann und sind daher überzeugt, dass die Einführung von dieser „KI" im Unternehmen schon nicht so schwer sein kann *(siehe Abbildung 5)*. Sie unterschätzen dabei vollkommen die Hürden von KI-Nutzung und Entwicklung. Des Weiteren haben sie seit der medialen Omnipräsenz von ChatGPT Angst, den KI-Zug völlig zu verpassen – haben sie doch zuvor bereits unter anderem den Digitalisierungs-, Big Data-, und Data-Mesh-Zug verpasst. Nun vernachlässigen sie in ihrem Übereifer wichtige Grundlagen wie Datenverfügbarkeit und -qualität.Die beste Gegenmaßnahme gegen die Torschlusspanik der Brents ist ein systematischer Aufbau von Datenkompetenz („Data Literacy") im gesamten Unternehmen. Zusätzlich kann die Aufmerksamkeit auf einfache Erfolge, also die niedrig hängenden Früchte, verschoben werden. Diese kleinen Erfolge bauen Erfahrung auf und sorgen für Motivation und Momentum. Es sollte den Brents auch verdeutlicht werden, dass sie sich nicht zu sehr von den Bullshittern unter den Wettbewerbern beirren lassen müssen. Stattdessen sollten sie sich auf eine langfristige strategische Ausrichtung für eine datengetriebene Entscheidungsfindung fokussieren.

Die Skeptiker: KI-kritische Pessimisten

Die Skeptiker sind der letzte Archetyp *(siehe Abbildung 6)*. Sie sind häufig ein glückliches Mitglied eines Kim-Teams und enge Vertraute der Gatekeeper. Ihnen fehlt praktisch jegliches Vertrauen in KI, was sich nicht zuletzt in einem Mangel an Erfahrung und Sachverständnis begründet. Sie unterschlagen auch gerne den Umstand, dass beispielsweise auch Menschen Fehler machen oder „halluzinieren" wie ein KI-Sprachmodel Sie warten geduldig auf die *richtige* KI und sind überzeugt, dass jeglicher Einsatz bisheriger Modelle Zeitverschwendung sei, solange diese nicht alle Teilprobleme eigenständig und mit absoluter Perfektion selbst lösen kann. Zu ihrer Verteidigung muss man jedoch auch sagen, dass sie über die Jahre schon jede Menge Bullshitter kommen und gehen gesehen haben, die ihnen schon zuvor das Blaue vom Himmel versprachen.

Beim Umgang mit Skeptikern ist es hilfreich, offen und mit Demut zu kommunizieren, was man sich von einer neuen KI-Initiative verspricht. Dabei ist es nützlich, die häufig nicht völlig unberechtigten Bedenken der Skeptiker proaktiv zu adressieren. Oft ist es auch ratsam die KI explizit als *Entscheidungshilfe* und eben nicht als *Entscheider* darzustellen. Es kann außerdem sinnvoll sein, die KI-Aktivitäten zunächst auf Teams von KI-Befürwortern zu konzentrieren und anhand erster Erfolge auch Vorbehalte bei den Skeptikern abzubauen. Darüber hinaus kann das Schaffen von Weiterbildungsmöglichkeiten auch bei Skeptikern die KI-Berührungsängste mit der Zeit mindern.

Zusammenfassung

Wir benötigen eine solide Datenstrategie mit sauberen, vollständigen, zugänglichen und gut dokumentierten Daten als Fundament. Die Schaffung eines allgemeinen Bewusstseins und eine geteilte Verantwortung für Datenqualität kann diese Grundlage schaffen.

Aller Anfang ist schwer, aber wir können mit kleinen erfolgreichen Vorzeigeprojekten weitere Motivation durch Momentum erzeugen. Hierfür ist es sinnvoll, zuerst mit den KI-Befürwortern zu kooperieren.

Und wir brauchen Mut zu Innovation, denn Risikoaversion ist keine langfristige Strategie. Ein stärkerer Austausch zwischen Managern, Ingenieuren, Data Scientists, Datenschutzbeauftragten und Fachabteilungen ist dabei der erste Schritt zum KI-Erfolg.

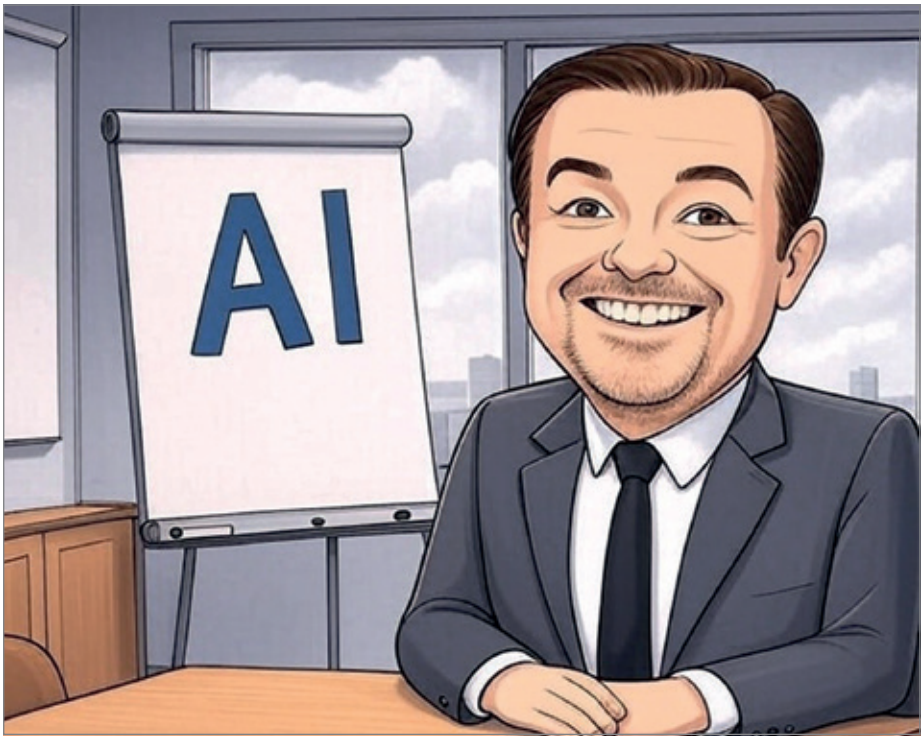


Abbildung 5: Die Brents (Quelle: Grok-Generation)

Quellen

[1] James Vincent und Jon Porter (2023), Samsung caught faking zoom photos of the Moon, The Verge, *https://www.theverge.com/2023/3/13/23637401/samsung-fake-moon-photos-ai-galaxy-s21-s23-ultra*

[2] Kavita Ganesan (2022), The Business Case for AI, Opinosis Analytics Publishing

[3] René Höltschi und Nikolai Thelitz (2024), Anatomie der deutschen Bürokratie: Wie ist so furchtbar geworden, was einst als Fortschritt galt?, NZZ, Berlin, *https://www.nzz.ch/wirtschaft/buerokratie-in-deutschland-keiner-will-sie-dennoch-waechst-sie-ld.1850515*

Über den Autor

Arne Wellnitz studierte Informatik und kann auf jahrelange Erfahrung im Bereich KI und Data Science zurückblicken. Er war unter anderem im Finanzbereich bei einer Vermögensverwaltung, bei einem großen Sportartikelhersteller und in der Automobilindustrie tätig. Aktuell ist er als Senior AI Engineer und AI Business Coordinator in der Paketlogistik angestellt. Seine Freizeit ist von Sport, Lesen, Softwareentwicklung, Kochen und Chilizucht geprägt. Zudem engagiert er sich als Organisator und Host beim Data Science & AI Meetup Nürnberg.

Abbildung 6: Die Skeptiker (Quelle: Grok-Generation)

Arne Wellnitz
awelldev@gmail.com

Herausforderungen bei der Generierung von SQL-Statements mithilfe von LLMs

Patrik Graf, merlin.zwo InfoDesign

Die Generierung von SQL-Statements durch Large Language Models (LLMs) wie GPT-4 von OpenAI bietet viele Möglichkeiten, birgt jedoch auch Herausforderungen. Während LLMs durch ihr umfangreiches Wissen prinzipiell in der Lage sind, komplexe SQL-Abfragen zu erstellen, gibt es typische Probleme wie Halluzinationen, syntaktische Fehler oder die Nutzung nichtexistierender Tabellen und Spalten. In diesem Artikel teile ich meine Erfahrungen, beleuchte die häufigsten Herausforderungen und zeige Lösungen sowie Best Practices auf, die sich in der Praxis bewährt haben.

Typische Herausforderungen

1. Halluzinationen und fehlerhafte Ergebnisse

LLMs neigen dazu, Informationen zu erfinden, da sie auf statistischen Mustern in der natürlichen Sprache basieren und keine direkte Anbindung an eine spezifische Datenbank haben. Dadurch erstellen sie oft scheinbar plausible, aber falsche SQL-Statements zum Beispiel mit nichtexistierenden Tabellen oder Spalten *(siehe Listing 1)*. Das Modell versucht, Lücken zu füllen, indem es basierend auf Wahrscheinlichkeiten Begriffe ergänzt, die zwar in ähnlichen Kontexten vorkommen, aber nicht zwangsläufig in der realen Datenbank existieren. Ein Beispiel ist die Verwendung von Spaltennamen, die plausibel klingen, aber nicht in der zugrunde liegenden Datenbank existieren.

2. Inkonsistente oder syntaktisch fehlerhafte SQL-Abfragen

Modelle wie GPT-3.5-Turbo haben häufig Probleme mit der korrekten SQL-Syntax. Dies liegt vor allem daran, dass LLMs nicht über ein explizites Verständnis von Grammatikregeln verfügen. Dadurch entstehen oft Fehler in der Struktur der

```
SELECT customer_name,
       order_total
  FROM sales_data
 WHERE order_date > '01.01.2024';
```

Listing 1: Die Tabelle sales_data existiert nicht, oder die Spalte customer_name ist falsch

```
SELECT product_name,
       SUM(sales_amount
  FROM orders
 WHERE order_date BETWEEN '01.01.2023' AND '31.12.2023';
```

Listing 2: Fehlende schließende Klammer in der SUM()-Funktion

```
{
    "answer": "Hier ist Ihre SQL-Abfrage.",
    "sql": "SELECT SUM(sales_amount) FROM orders WHERE
order_date > '01.01.2024';"
}
```

Listing 3: Beispiel für eine strukturierte Ausgabe im JSON-Format

SQL-Abfragen. Die erzeugten Abfragen enthalten manchmal falsche Klammern oder unvollständige JOIN-Bedingungen *(siehe Listing 2)*. Ein weiterer Grund für diese Fehler ist, dass SQL eine sehr strukturierte und formale Sprache ist, während LLMs auf natürlichen Sprachmustern basieren. Ohne explizite Regelwerke für SQL-Syntax können sie Syntaxfehler wie fehlende oder falsch platzierte Zeichen generieren. Zudem haben kleinere Modelle wie GPT-3.5-Turbo ein eingeschränktes Kontextfenster, wodurch sie die Gesamtstruktur längerer SQL-Abfragen nicht immer korrekt erfassen, und dadurch unvollständige oder fehlerhafte Statements produzieren.

3. Fehlende Kontextsensitivität

LLMs haben in der Regel keine Kenntnis über die spezifische Datenbankstruktur eines Unternehmens. Zudem fehlt ihnen die direkte Anbindung an diese Datenbanken. Sie können nur durch den bereitgestellten Kontext mit den nötigen Informationen zur Datenbankstruktur versorgt werden, denn ohne detaillierte Tabelleninformationen kann das Modell keine maßgeschneiderten Abfragen generieren. Hierbei spielen DDL-Statements eine entscheidende Rolle. Durch das Bereitstellen von CREATE TABLE-Anweisungen oder DESCRIBE TABLE-Ergebnissen erhält das Modell eine präzisere Vorstellung der vorhandenen Strukturen. Diese zusätzlichen Informationen ermöglichen es, realistische und syntaktisch korrekte SQL-Abfragen zu generieren, die auf den tatsächlich vorhandenen Strukturen basieren. Ebenso wichtig sind Tabellen- und Spaltenkommentare, da sie wertvolle Metadaten über die Bedeutung der einzelnen Tabellen und Felder liefern. Sie helfen dem Modell, den semantischen Kontext besser zu verstehen. Wenn diese Informationen bereitgestellt werden, kann das LLM gezieltere und fachlich korrektere Abfragen generieren, da es die geschäftliche Bedeutung der Datenbankstrukturen berücksichtigt.

Strukturierte Ausgabeformate zur Weiterverarbeitung

Um ein LLM besser in bestehende Workflows zu integrieren, bietet sich die Ergebnisrückgabe im JSON-Format an, da es eine einfache Weiterverarbeitung ermöglicht *(siehe Listing 3)*. Natürlich sind auch andere strukturierte Formate als Rückgabe möglich, wie beispielsweise XML, jedoch bietet hier JSON entscheidende Vorteile:

- Einfach lesbar
- Geringer struktureller Overhead
- Typenunterstützung (z. B. bool, string, etc.)
- Sehr gute Integration in moderne Web-Technologien

Ein Beispiel für einen effektiven Prompt

Ein klar definierter Prompt verbessert die Qualität der generierten SQL-Statements erheblich. Das folgende Beispiel *(siehe Listing 4)* gibt dem LLM detaillierte Anweisungen, um fehlerfreie und strukturierte SQL-Abfragen zu erstellen:

Ein solcher detaillierter Prompt hilft, die generierten SQL-Abfragen zu verbessern und häufige Fehler zu vermeiden. Im Folgenden gehe ich nochmal auf den Zweck der wichtigsten Punkte im Prompt ein, um zu verdeutlichen, was damit im Detail bezweckt wird:

- Strukturierte JSON-Antwort
  Dies ermöglicht die strukturierte Weiterverarbeitung durch ein nachgelagertes System.
- Erstellung eines praxisnahen SQL-Statements
  Das LLM soll vorhandene Spalten und Tabellen nutzen und keine Elemente einfügen. Dadurch wird sichergestellt, dass das generierte SQL hinsichtlich der Datenobjekte tatsächlich ausführbar ist.
- Mandanteneinschränkung
  Falls Tabellen eine Mandantenspalte enthalten, wird mit dem Wert des APEX-Feldes „GLOBAL_MANDANT“ gefiltert. Dies stellt sicher, dass nur die Daten des aktuellen Mandanten abgefragt werden, was für Multi-Tenant-Systeme wichtig ist.
- Benutzerspezifische Abfragen
  Falls sich ein Benutzer auf sich selbst bezieht, wird seine ID durch die APEX-Felder „GLOBAL_M_NR“ oder „APP_USER“ ermittelt, wodurch personalisierte Abfragen möglich werden – ohne direkte Eingabe von Benutzerinformationen.
- Sortierregel: NULLS LAST
  Stellt sicher, dass NULL-Werte am Ende der Sortierung stehen, was bei unseren Anwendungsfällen eine bessere Benutzererfahrung bot.
- NVL für NULL-Werte in numerischen und Datumsspalten
  Verhindert, dass NULL-Werte ungewollt zu Problemen in Berechnungen oder Aggregationen führen.
- Datum als Variable $[AIL_HEUTE]
  Da das LLM das heutige Datum nicht kennt, wird dieses dynamisch in den Prompt eingefügt. Dies verbessert die Generierung von zeitabhängigen SQL-Abfragen.
- Deutsche Zahlen- und Datumsformate
  Alle Datums- und Zahlenwerte sollen mit deutschen Formatmasken ausgegeben werden.
- Definition des Geschäftsjahres durch $[MANDANT_GJ_VON_BIS]
  Geschäftsjahre werden durch eine Variable dynamisch eingefügt, was mandantenspezifische Geschäftsjahresdefinitionen ermöglicht.
- Ausgabe ausschließlich als JSON
  Dadurch wird sichergestellt, dass keine zusätzlichen Texte oder Markup-Tags in der Ausgabe enthalten sind. Bei „älteren“ Modellen sollte dies nicht fehlen.

Unterschiede zwischen den Modellen

Ein Vergleich zwischen verschiedenen LLMs zeigt signifikante Unterschiede in der Qualität der generierten SQL-Statements. Beispielsweise lieferte GPT-4o eine deutlich präzisere Abfrage für die Umsatzermittlung als GPT-3.5-Turbo, welches oft syntaktische Fehler enthielt. Modelle mit größerem Kontextfenster (z. B. GPT-4-Turbo) tendieren dazu, genauere und vollständigere Abfragen zu generieren, insbesondere wenn komplexe Tabellenbeziehungen involviert sind.

Im Folgenden sehen wir einige generierte SQLs zur Frage „Wie war der Umsatz im letzten Geschäftsjahr?“, jeweils mit unterschiedlichen Modellen von OpenAI *(siehe Listings 5-8)*. Als Datenbasis diente unser hauseigenes ERP-System.

```
Dein Job ist es, die Anfrage des Benutzers zu analysieren und darauf folgende Schritte auszuführen:
1. Erstelle ein JSON-Objekt mit den Attributen "answer" und "sql".
2. Generiere ein SQL-Statement basierend auf dem gegebenen Kontext, um die gewünschten Informationen zu extrahieren. Nutze dein Allgemeinwissen um Informationen im SQL zu ergänzen, wenn diese nicht im gegebenen Kontext auffindbar sind. Verwende dabei nur existierende Spalten aus den vorgegebenen Tabellen. Du darfst keine hypothetischen Spalten oder Tabellen verwenden. Achte darauf, dass das generierte SQL praxisnah ist und sich auf real existierende Daten bezieht. Falls mehrere Tabellen erforderlich sind, integriere diese sinnvoll.
3. Erstelle für jede Ausgabespalte einen gut lesbaren Alias. Umschließe den Alias in doppelten Anführungszeichen.
4. Wenn Mandantenspalten in den integrierten Tabellen vorhanden sind, schränke per Funktionsaufruf v('GLOBAL_MANDANT') auf einen Mandanten ein.
5. Nutze als Ausgabespalten niemals Primärschlüssel oder Fremdschlüssel.
6. Der Principal Name von Mitarbeitern ist immer ".....", alles in Kleinbuchstaben.
7. Wenn der Benutzer Daten über sich selbst wünscht, kann seine MAMI_ID über den Funktionsaufruf v('GLOBAL_MNR') und seine MAMI_LDAP_UID über den Funktionsaufruf v('APP_USER') ermittelt werden.
8. Benutze beim Sortieren immer NULLS LAST.
9. Benutze bei numerischen Spalten und bei Datumsspalten, die NULLABLE sind, immer die Funktion NVL.
10. Benutze für relative Zeitangaben immer das aktuelle Datum. Das aktuelle Datum ist $[AIL_HEUTE] im Format DD.MM.YYYY
11. Finde und filtere Zeilen in einer Tabelle, deren Bezeichnungsspalten eine Ähnlichkeit von mindestens 85% mit einem gegebenen Suchbegriff aufweisen. Nutze die UTL_MATCH.JARO_WINKLER_SIMILARITY-Funktion, um die genaue Übereinstimmung unter Berücksichtigung von Transpositionen und Vorsilbenverstärkung zu analysieren. Gib nur die Zeilen zurück, deren Ähnlichkeitswerte für Bezeichnungs- und Namensspalten >= 85% liegen. Namensspalten für Personen, wie Vorname, Nachname oder Kombinationen daraus, sollen auf normale Art gefiltert werden.
12. Benutze für Abfragen mit Geodaten immer die Funktionen von Oracle Spatial.
13. Schreibe eine Antwort im Stil eines Gespräches unter guten Kollegen in das Attribut "answer". Informationen, die du nicht kennst, werden durch das generierte SQL-Statement ermittelt. Halte deine Antwort allgemein. Präsentiere in einfachen Worten, ohne technische Details, das Ergebnis. Erwähne die folgenden Begriffe nicht in deiner Antwort: SQL, Datenbank.
14. Alle Datums- und Zahlenformate sind in deutsch und die Formatmasken sind Oracle-SQL Formatmasken.
15. Bei Anfragen zu Geschäftsjahren gilt ein Geschäftsjahr vom $[MANDANT_GJ_VON_BIS] des Folgejahres.
16. Deine Ausgabe darf nur das JSON-Objekt enthalten.

Kontext, Oracle SQL:
%%BUSINESS_CONTEXT_SQL%%
```

Listing 4: Ein ausführlicher Prompt zum Generieren von lauffähigem SQL

```
SELECT NVL(SUM(repo_wert_rabattiert), 0) AS "Umsatz"
  FROM erp.rng_pos rp
  JOIN erp.rng_kopf rk ON rp.repo_reko_id = rk.reko_id
 WHERE rk.reko_gj = EXTRACT(YEAR FROM ADD_MONTHS(SYSDATE, -3);
```

Listing 5: GPT-3.5-Turbo mit fehlerhafter SQL-Syntax (am Ende fehlt eine Klammer). Zudem ist das SQL fachlich falsch

```
SELECT SUM(NVL(REPO_WERT_RABATTIERT, 0)) AS "Umsatz"
  FROM ERP.RNG_POS
  JOIN ERP.RNG_KOPF ON RNG_POS.REPO_REKO_ID = RNG_KOPF.REKO_ID
 WHERE REKO_GJ = TO_NUMBER(TO_CHAR(ADD_MONTHS(SYSDATE, -6),
'YY')) -- FALSCH!
   AND REKO_MAND_ID = v('GLOBAL_MANDANT')
```

Listing 6: GPT-4-Turbo mit ausführbarem SQL, jedoch fachlich falsch

Lessons Learned: Best Practices

- Präzise Prompts formulieren
  Ein klar definierter Prompt mit detaillierten Anweisungen reduziert Halluzinationen deutlich.
- Einsatz von JSON als Rückgabeformat
  Standardisierte Strukturen erleichtern die Weiterverarbeitung und Integration in bestehende Systeme.
- Kontinuierliche Optimierung
  Durch iterative Anpassung der Prompts wird die Qualität der SQL-Generierung immer weiter verbessert. Neue Modelle müssen gut darauf getestet werden, ob ihre „SQL-Fähigkeiten“ in der Praxis Verbesserungen ermöglichen.

Variable	Bedeutung
$[AIL_HEUTE]	Aktuelles Datum im Format DD.MM.YYYY.
$[MANDANT_GJ_VON_BIS]	Zeitraum für das Geschäftsjahr des Mandanten.
%%BUSINESS_CONTEXT_SQL%%	DDL Statements (CREATE TABLE etc.)

Tabelle: Bedeutung der Variablen im Prompt

```
SELECT SUM(NVL(REPO_WERT_RABATTIERT, 0)) AS "Umsatz"
  FROM ERP.RNG_POS
  JOIN ERP.RNG_KOPF ON RNG_POS.REPO_REKO_ID = RNG_KOPF.REKO_ID
 WHERE RNG_KOPF.REKO_DATUM BETWEEN TO_DATE('01.07.2023', 'DD.MM.YYYY')
  AND TO_DATE('30.06.2024', 'DD.MM.YYYY') -- KORREKT!
   AND RNG_KOPF.REKO_MAND_ID = v('GLOBAL_MANDANT')
```

Listing 7: GPT-4o mit ausführbarem und fachlich richtigem SQL

```
SELECT NVL(SUM(REPO_WERT), 0) AS "Gesamtumsatz" -- FALSCH!
  FROM ERP.RNG_POS RP
  JOIN ERP.RNG_KOPF RK ON RP.REPO_REKO_ID = RK.REKO_ID
 WHERE RK.REKO_GJ = EXTRACT(YEAR FROM SYSDATE) - 1 -- FALSCH!
   AND RK.REKO_DATUM >= TO_DATE('01.07.' || (EXTRACT(YEAR FROM SYSDATE)
- 1), 'DD.MM.YYYY')
 AND RK.REKO_DATUM <= TO_DATE('30.06.' || EXTRACT(YEAR FROM SYSDATE),
'DD.MM.YYYY')
   AND RK.REKO_MAND_ID = v('GLOBAL_MANDANT')
```

Listing 8: GPT-4o-Turbo mit ausführbarem SQL, jedoch fachlich falsch

Fazit und Ausblick

LLMs haben durchaus das Potenzial, SQL-Statements effizient zu generieren, aber sie sind noch nicht auf menschlichem Niveau. Durch präzise Prompts, strukturierte Ausgabeformate wie JSON und eine kontinuierliche Optimierung des bereitgestellten Kontexts lassen sich aber auch schon heute viele Herausforderungen bewältigen. Diese Technologie ist jedoch so schnelllebig, dass dieser Artikel wahrscheinlich schon wieder veraltet ist, bis sie Ihn zu lesen bekommen. Die Modelle werden immer besser im Bewältigen von Entwicklungsaufgaben; in naher Zukunft könnten vielleicht sogar auf SQL spezialisierte Modelle, mit direkten Anbindungsmöglichkeiten an Datenbanken, dies alles hier vollkommen überflüssig machen. Die Zukunft bleibt auf jeden Fall spannend.

Über den Autor

Patrik Graf ist Head of Innovation bei merlin.zwo InfoDesign GmbH & Co. KG und beschäftigt sich intensiv mit der Integration von KI in datenbankgestützten Systemen. Er ist regelmäßiger Sprecher auf Fachkonferenzen und teilt seine Erfahrungen in Artikeln und Vorträgen.

Patrik Graf
patrik.graf@merlin-zwo.de

2025
DOAG
Konferenz + Ausstellung

Die ORACLE Anwenderkonferenz

Nürnberg | 18. – 21. Nov.

anwenderkonferenz.doag.org

EU AI Act: Chance für vertrauenswürdige KI oder Bremsklotz für Unternehmen?

Benedikt Backhaus, selbstständiger KI-Berater und Dozent

Der EU AI Act ist das weltweit erste umfassende KI Gesetz. Er soll Vertrauen schaffen und Missbrauch verhindern, ohne Innovation abzuwürgen. Doch im geopolitischen Wettlauf mit den USA und China, bei hohen Compliance-Kosten und noch offenen Fragen droht er zugleich zum Bürokratiemonster zu werden. Der Artikel ordnet die geopolitische Dimension, die Grundregeln, die Folgen und Handlungsoptionen für Unternehmen ein – und zeigt, wo der AI Act nachjustiert werden muss, damit „Trusted AI made in Europe" wirklich ein Wettbewerbsvorteil wird.

KI-Politik ist Machtpolitik

Die Auseinandersetzung um KI ist längst geopolitisch: Es geht um technologische Vorherrschaft, Werte und Macht. Spätestens seit der Rede von US-Vizepräsident J.D. Vance beim AI Action Summit in Paris ist das völlig klar, er warnte vor „übermäßiger Regulierung", die Wettbewerb und Innovation behindere. In der Tradition der Tech-Industrie des Silicon Valley favorisieren die USA einen „hands-off" Ansatz mit dem offen erklärten Ziel der Dominanz der USA. Kooperation und gemeinsame Regeln? Not so much. Möchte man die US-KI-Politik auf eine Formel bringen, so erscheint mir das ursprüngliche Motto von Facebook-Gründer Mark Zuckerberg am treffendsten: „Move fast and break things."

Demgegenüber verfolgt die Volksrepublik China einen fast gegenteiligen Weg. KI gilt dort als Chefsache, eingebettet in Fünf-Jahres-Pläne und mit massi-

ver staatlicher Förderung. Die Kommunistische Partei reguliert KI strikt nach eigenen Vorstellungen: Algorithmen müssen registriert werden, Zensur und Kontrolle (zum Beispiel bei generativen KI-Inhalten) sind an der Tagesordnung. Das Staatsinteresse rangiert – ganz im Sinne der traditionellen Kultur – über individueller Freiheit und KI wird gezielt zur Überwachung (Gesichtserkennung, Social Scoring) eingesetzt. Hinzu kommen im Land mit 1,4 Milliarden Einwohnern gigantische Datenmengen und – dank großflächig angelegter Bildungsprogramme – eine Vielzahl an Programmierern.

In dieser bipolaren Welt versucht die Europäische Union einen eigenen Weg zu finden – weder Wildwest-Kapitalismus noch Orwellscher Überwachungsstaat. Der AI Act ist ein zentrales Element dieser Strategie, soll er doch zugleich die Wahrung der Europäischen Grund- und Menschenrechte ermöglichen und die Wettbewerbsfähigkeit europäischer Firmen stärken. Der AI Act ist somit ein Werkzeug der Wertediplomatie: Er soll einen globalen Standard setzen, an dem sich andere Staaten orientieren, ähnlich wie die DSGVO im Datenschutz. So versucht die EU, trotz geringerer Tech-Marktmacht, als „regulatorische Supermacht" Einfluss auszuüben. Die Vision: „vertrauenswürdige KI", die mit europäischen Werten (Privatsphäre, Menschenrechte und vielem mehr) vereinbar ist.

Was regelt der AI Act konkret?

KI-Systeme, die nicht den EU-Vorgaben entsprechen, dürfen in Europa künftig gar nicht erst in Verkehr gebracht werden. Für Anbieter bedeutet das: Keine Konformität, kein Markt. Wer den EU-Markt (über 440 Mio. Verbraucher) bedienen will, muss also den AI Act einhalten. Das gilt auch für ausländische Player, denn der AI Act gilt immer, sobald das Ergebnis eines KI-Systems in der EU angewandt wird.

Im Kern folgt die Verordnung einem risikobasierten Ansatz: KI-Systeme werden je nach Gefahrenpotenzial in Kategorien eingestuft, mit abgestuften Pflichten *(siehe Abbildung 1)*. Dieser Ansatz soll dafür sorgen, dass strenge Auflagen nur dort greifen, wo KI wirklich Schaden anrichten kann, und triviale Anwendungen nicht unnötig belastet werden. Die vier Risikostufen nach AI Act sind:

1. **Unakzeptables Risiko – Verbotene KI:** KI-Systeme, die gegen grundlegende Werte und Rechte verstoßen, sind vollständig verboten. Dazu zählen zum Beispiel Manipulationstechniken, die Menschen unbewusst beeinflussen (subliminale Beeinflussung) oder ausnutzende Algorithmen, die etwa Kinder oder vulnerable Gruppen schädigen könnten. Ebenfalls verboten ist ein allgemeines Social Scoring durch Behörden (nach dem Vorbild Chinas) sowie die massenhafte Echtzeit-Biometrie im öffentlichen Raum zu Überwachungszwecken. Letzteres ist nur mit engen Ausnahmen möglich, um beispielsweise akute terroristische Bedrohungen abzuwehren. Diese Verbote traten bereits 6 Monate nach Inkrafttreten, also im Februar 2025, EU-weit in Kraft – ein deutliches Signal der EU, gewisse KI-Einsätze gar nicht erst zu tolerieren.
2. **Hohes Risiko – Streng regulierte KI:** Diese Kategorie umfasst Anwendungen, die ein erhebliches Risiko für Gesundheit, Sicherheit oder Grundrechte darstellen können. Hierunter fallen zwei Gruppen: (a) KI-Systeme, die als Sicherheitskomponente in Produkten dienen (z. B. im Auto, Medizingerät etc.), und (b) KI-Systeme in bestimmten sensiblen Bereichen wie Bildung, Personalwesen (z. B. CV-Screening), Kreditvergabe, Migration, Justiz oder Strafverfolgung. Für diese Hochrisiko-KI gelten besonders strikte Auflagen in Kapitel III des Gesetzes. Anbieter müssen ein umfassendes Compliance-Paket schnüren: sorgfältiges Risikomanagement, hochwertige Trainingsdaten (keine Verzerrungen), ausführliche technische Dokumentation und Protokollierung, Transparenz gegenüber Nutzern, menschliche Aufsicht über das System und Robustheits-/Cybersecurity-Maßnahmen. Bevor ein Hochrisiko-System auf den Markt kommt, muss es eine Konformitätsbewertung durchlaufen und im Erfolgsfall die CE-Kennzeichnung erhalten – ähnlich wie man es von Maschinen oder Spielzeug kennt (der AI Act lehnt sich hier an das bewährte New Legislative Framework an). Hersteller erklären damit, dass ihre KI alle Anforderungen erfüllt. Wichtig: Die Liste der Hochrisiko-Anwendungen (Anhang III) kann die EU-Kommission künftig anpassen, wenn neue Gefahrbereiche auftauchen. Dies sorgt für Flexibilität – nährt aber auch bei Unternehmen die Sorge vor immer längeren Listen und unsicherer Zukunft.
3. **Begrenztes Risiko – Transparenzpflichten:** Für einige nicht hochriskante KI-Systeme schreibt der AI Act trotzdem Transparenz vor. Zum Beispiel müssen Chatbots den Nutzer darauf hinweisen, dass er es mit einer KI zu tun hat (und nicht mit einem Menschen) – damit niemand unbeabsichtigt mit einer Maschine kommuniziert. Auch KI-Generatoren für synthetische Medien (Deepfakes) müssen ihre künstliche Natur kenntlich machen, damit etwa KI-Bilder als solche erkennbar sind. Systeme zur emotionalen Analyse oder biometrischen Kategorisierung sollen ebenfalls Transparenzhinweise geben. Solche Anwendungen dürfen also genutzt werden, erfordern aber gewisse Hinweispflichten, um informierte Entscheidungen zu ermöglichen. Weitere Auflagen gibt es nicht, insbesondere keine Zulassungsverfahren – das Risiko gilt als beherrschbar durch Offenlegung.
4. **Minimales Risiko – Frei nutzbare KI:** Alle anderen KI-Systeme, die in keine der obigen Kategorien fallen, gelten als geringes oder minimales Risiko. Für sie schreibt der AI Act keine unmittelbaren verpflichtenden Anforderungen vor. Hier vertraut der EU-Gesetzgeber auf bewährte Gesetze (z. B. Produkthaftung, allgemeine Sicherheitsanforderungen) und freiwillige Verhaltenskodizes. Die überwiegende Mehrheit heutiger KI-Anwendungen – von Spamfiltern über Empfehlungsalgorithmen bis zu simplen Automatismen – fällt in diese Stufe und bleibt durch den AI Act weitgehend unreguliert. Allerdings gibt es eine Querschnittspflicht: Grundlagenkompetenz in KI. Unternehmen müssen nämlich dafür sorgen, dass ihre Mitarbeiter über ausreichend KI-Know-how verfügen, wenn KI-Systeme eingesetzt werden – selbst bei mini-

Abbildung 1: Eigene Darstellung auf Grundlage der EU-Kommission (Quelle: Benedikt Backhaus)

malem Risiko. Damit will die EU allgemein ein Bewusstsein für den verantwortlichen KI-Einsatz schaffen.

Zusätzlich hat der AI Act noch Sonderregeln für „Generelle KI-Modelle" eingefügt (oft als GPAI oder Foundation Models bezeichnet). Diese wurden erst im Laufe des Gesetzgebungsverfahrens ergänzt, da sie zu Beginn der Debatte um den AI Act noch gar nicht vorhanden waren und die Brisanz der Regulierung erst mit dem Aufkommen von ChatGPT deutlich wurde. Solche grundlegenden Modelle, die für breite Zwecke trainiert sind, müssen gewisse Transparenz- und Sicherheitsvorkehrungen erfüllen, insbesondere wenn sie leistungsstark sind. Offen verfügbare Open-Source-Modelle genießen etwas gelindere Pflichten, um Innovation nicht zu ersticken – allerdings nur, solange sie nicht als „Generelle KI-Modelle" eingestuft sind. Insgesamt wurde hier Neuland betreten, um mit der rasanten Entwicklung von Allzweck-KI Schritt zu halten.

Der AI Act ist mit empfindlichen Bußgeldern bewehrt – teils noch höher als bei der DSGVO. Maximal sind €35 Mio. oder 7% des weltweiten Jahresumsatzes als Strafe möglich (je nachdem, was höher ist). Zum Vergleich: Bei Datenschutz-Verstößen sind es 4%. Schon allein dieses Drohpotential macht KI-Compliance zum Chef-Thema. Ein Regelverstoß – etwa die Inverkehrbringung einer verbotenen KI oder das Ignorieren von Auflagen – kann existenzielle Konsequenzen haben, gerade für Global Player. Zudem droht der Reputationsverlust, wenn ein Unternehmen wegen KI-Verstößen in die Schlagzeilen gerät.

Besonders relevant für Unternehmen ist zudem Artikel 4 des AI Act: Er sieht vor, dass Mitarbeitende ausreichend KI-Kompetenz zur Nutzung von KI-Systemen verfügen müssen:

> *Die Anbieter und Betreiber von KI-Systemen ergreifen Maßnahmen, um nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen, wobei ihre technischen Kenntnisse, ihre Erfahrung, ihre Ausbildung und Schulung und der Kontext, in dem die KI-Systeme eingesetzt werden sollen, sowie die Personen oder Personengruppen, bei denen die KI-Systeme eingesetzt werden sollen, zu berücksichtigen sind.*

Wenngleich keine Details genannt sind, wie diese Kompetenz erworben und nachgewiesen werden kann, dürfte der AI Act damit in der Praxis zu zahlreichen Schulungsmaßnahmen und Upskilling-Programmen beitragen, sofern KI compliant eingesetzt werden soll.

Der AI Act ist somit ein Mix aus Produktrecht (Sicherheit, CE-Kennzeichnung) und Grundrechtsschutz. Er verlangt zum Beispiel von Hochrisiko-Anbietern eine Grundrechts-Folgenabschätzung – ähnlich einer Datenschutz-Folgenabschätzung – um Einflüsse auf beispielsweise Meinungsfreiheit oder Gleichbehandlung abzuschätzen. Diese Kopplung von technischen Sicherheitsstandards mit ethisch-rechtlichen Aspekten ist ein Novum. Entsprechend fehlen bisher Erfahrungswerte und Standards, wie man Grundrechte konkret in KI-Produktprüfungen integriert.

Ferner gelten bisherige Rechtsnormen weiter (z. B. DSGVO). Ein prägendes Beispiel ist der SyRI-Fall in den Niederlanden: SyRI war ein staatliches KI-System zur Aufdeckung von Sozialbetrug, das verschiedene Datenbanken verknüpfte und Risikoprofile armer Bürger erstellte. 2020 stoppte ein Gericht in Den Haag SyRI mit der Begründung, es verletze die

Privatsphäre- und Menschenrechtsgarantien der Bürger. Es war das erste Mal, dass ein Gericht einen automatisierten Algorithmus auf Basis von Grundrechten kassierte – noch ohne spezifisches KI-Gesetz, sondern gestützt auf die Europäische Menschenrechtskonvention (Art. 8, Recht auf Privatsphäre).

AI Act führt zu Unsicherheit – und hemmt unser Tempo

Klare Regeln ermöglichen es Firmen, ihre Investitionen zu planen – etwa welche KI Projekte angesichts kommender Auflagen realistisch sind. Genau hier liegt das Problem des AI Act: in der Umsetzung. Viele Begriffe im AI Act sind nicht klar genug definiert und insbesondere die Einstufung von „Hochrisiko-KI-Systemen" ist für eine ganze Reihe an Systemen offen. Selbst Gabriele Mazzini, einer der Architekten des AI Act, äußerte Bedenken, die Regulierung habe die Messlatte „vielleicht zu hoch gelegt".

Hinzu kommt der Aufwand für die AI-Act-Compliance: Wer KI einsetzt, muss künftig interne Prüfschritte etablieren – etwa Risikobewertungen, Dokumentationen, Mitarbeiter-Schulungen. Insbesondere für Hochrisiko-KI können die Compliance-Kosten erheblich sein. Eine Studie der EU-Kommission schätzte 2021, dass ein mittelständisches Unternehmen (50 MA) für ein einziges Hochrisiko-KI-Produkt €159.000 bis €319.000 aufwenden müsste, um alle Anforderungen zu erfüllen (z. B. externe Audits, Qualitätstests der Datensätze, Implementierung von Monitoring-Systemen). Dieser Bürokratieaufwand trifft große Konzerne weniger, für KMU jedoch ist er eine Herausforderung.

Damit könnte das genaue Gegenteil von dem passieren, was sich die EU vom AI Act und ihrer Vorreiterrolle bei der KI-Regulierung erhofft hatte. Die Sorge ist, dass am Ende nur große US-Tech-Konzerne die Ressourcen haben, um die Auflagen zu stemmen, während europäische Start-ups ins Hintertreffen geraten. Tatsächlich sehen sich junge Unternehmen besonderen Herausforderungen gegenüber: hohe Compliance-Kosten, unklare Begriffe, fehlende Standards – all das kann für ein Start-up abschreckend wirken. Zumal in einem Bereich, in dem im außereuropäischen Ausland eine bessere Venture-Capital-Ausstattung, weniger Bürokratie und eine unternehmerfreundlichere Kultur warten.

Ebenso müssen Unternehmen im Prozess der Umsetzung mitgenommen und informiert werden, damit der AI Act Wirkung entfaltet: Viele Unternehmen sind nämlich derzeit noch unsicher, ob und wie der AI Act sie betrifft. Einer Umfrage unter 778 Entscheidern im deutschsprachigen Raum zufolge wussten im Juli 2024 erst 32% über die Inhalte des AI Act Bescheid.

Die offenen Interpretationen, die hohen Compliance-Kosten und die mangelnde Awareness führen dazu, dass manche Firmen lieber auf Nummer sicher gehen und KI-Anwendungen vorerst pausieren. Hier müssen nach und nach harmonisierte Normen erarbeitet werden, damit Unternehmen genau wissen, wie sie die Anforderungen erfüllen können.

Digitale Souveränität ist in der EU nicht in Sicht

Betrachten wir den aktuellen Stand der EU-Digital-Politik im Kontext des AI Act, müssen wir trotz markiger Forderungen von Spitzenpolitikern feststellen, dass wir von „digitaler Souveränität" meilenweit entfernt sind.

Im Bereich der privaten Investitionen flossen zwischen 2018 und 2023 circa €32,5 Mrd. in EU-KI-Startups, verglichen mit über €120 Mrd. in US-KI-Firmen. Und das war noch vor dem „Generative AI Hype" und den Ankündigungen der Trump-Administration, $500 Mrd. für KI-Infrastruktur im Rahmen des „Stargate Project" zu mobilisieren. Zwar versucht die EU mit Programmen wie Horizon Europe und Digital Europe gegenzusteuern. Doch das entspricht nur einem kleinen Bruchteil der US-Ausgaben, selbst wenn man die Ankündigung weiterer €50 Mrd. aus öffentlichen Mitteln der EU hinzurechnet. Eine weitere Herausforderung: Die Infrastruktur – Hochleistungsrechenzentren, Cloud-Plattformen, KI-Chips – wird in Europa fast vollständig von US-Konzernen gestellt (Amazon AWS, Microsoft Azure, NVIDIA GPUs), und das lässt sich nicht von heute auf morgen ändern.

Auch bei zentralen KI-Innovationen liegen die USA vorn: 2024 wurden 40 bedeutende neue KI-Modelle in den USA entwickelt, in China 15 und in Europa nur 3. Auch in der Modell-Performance holt China rasant auf und führt mittlerweile beim KI-Forschungsoutput (Publikationen, Patente). Zwar wird sich dieses Gefälle angesichts der jüngst geführten Zollstreitigkeiten zwischen den USA und China und der Erhebung von 145% Zoll auf chinesische Importe weiter verschärfen – allerdings eher zugunsten der USA und nicht zugunsten der EU.

Damit steht die EU von zwei Seiten unter Druck: Einerseits muss dringend das Tempo erhöht werden, um mit den rasanten Entwicklungen im KI-Bereich mithalten zu können und so im Wettbewerb nicht den Anschluss zu verlieren. Andererseits wäre eine komplette Aufhebung der KI-Regulierung unglaubwürdig und dürfte nicht zu mehr Rechtssicherheit führen, sondern eher zu größerer Verunsicherung und Vertrauensverlust.

Die Hoffnung auf Abschaffung ist also unbegründet: Kurzfristig wird es keine komplette Neuauflage des AI Act geben – dazu ist er zu frisch. Aber innerhalb des bestehenden Rahmens sind Nachbesserungen denkbar. Zum Beispiel könnten die noch ausstehenden harmonisierten Normen bewusst praxisnah gestaltet werden, um Unternehmen die Umsetzung zu erleichtern. Die EU-Kommission kann zudem von ihrer Befugnis Gebrauch machen, Leitlinien zu erlassen, etwa was genau als „Stand der Technik" gilt oder wie KI-Anbieter den Grundrechtsnachweis führen können. Ebenso könnte ein starkes „KI-Büro" verbindliche Auskünfte und Interpretationen geben und Unsicherheiten schneller ausräumen. Auch die Umsetzung der Förderprogramme wird eine Rolle spielen: Wenn die EU ausreichend Anreize und Finanzierung für vertrauenswürdige KI-Innovationen bietet, mildert das die Sorge, Europa verliere durch den AI Act an Tempo.

Bis dahin bleibt Unternehmen, die KI in der EU einsetzen, nur ein Weg: sich an die Regeln des AI Act bestmöglich zu halten, sich regelmäßig über neue Entwicklungen zu informieren und die Nutzung oder Entwicklung von KI-Systemen weiter voranzutreiben. Die USA und China werden nicht auf uns warten.

BONUS CHECKLISTE: Fünf Schritte zur Compliance mit dem EU AI Act

1. **KI Bestandsaufnahme**
   Welche KI Systeme – intern wie extern – nutzen oder entwickeln wir aktuell, und in welchem Geschäftskontext werden sie eingesetzt?
   Eine vollständige Bestandsaufnahme schafft Transparenz über alle Modelle, Datenpipelines und Drittanbieter-Services im Unternehmen. Dabei sollten vom Eigentümer, Zweck, Trainingsdatenquellen und Schnittstellen dokumentiert werden, um spätere Risikobewertungen lückenlos zu ermöglichen. Die Inventarliste ist kein Einmalprojekt, sondern muss als lebendes Verzeichnis regelmäßig aktualisiert und versioniert werden

2. **Bewertung (insbesondere „verboten"/„High Risk")**
   In welche EU-AI-Act-Risikokategorie (verboten, Hochrisiko, geringes Risiko, minimal) fällt jedes identifizierte System – und warum?
   Anhand der Annex-III-Kriterien und der Verbotsliste wird jedes System klassifiziert: zum Beispiel biometrische Echtzeit-Überwachung (verboten) oder KI-gestützte Kreditwürdigkeitsprüfung (High Risk). Für Hochrisiko-Systeme sind zusätzlich Zweckbindung, Datenqualität, Erklärbarkeit und Governance streng zu prüfen. Dokumentierte Entscheidungspfade erleichtern spätere Audits und mindern Haftungsrisiken

3. **Compliance Maßnahmen**
   Welche organisatorischen, technischen und vertraglichen Kontrollen müssen wir einführen, damit jedes KI-System die EU-AI-Act-Pflichten nachweislich erfüllt?
   Für High Risk Systeme verlangt der AI Act unter anderem Risikomanagement, Daten- und Modell-Governance, menschliche Aufsicht, Robustheits-Tests sowie CE-Konformitätserklärung. Praktisch heißt das: klare Verantwortlichkeiten, SOPs für Modelländerungen, Monitorings mit Alarmschwellen und Lieferantenverträge mit Audit-Rechten. Ein internes KI-Governance-Board steuert Prioritäten und eskaliert Regelverstöße.

4. **Schulung und Sensibilisierung**
   Wie stellen wir sicher, dass alle relevanten Mitarbeitenden die Pflichten, Risiken und Grenzen des EU AI Acts verstehen und in ihrer täglichen Arbeit berücksichtigen?
   Ein rollenbasiertes Trainingsprogramm vermittelt Entwicklern, Fachbereichen und Führungskräften praxisnahe Leitlinien – vom Datenlabeling bis zur Risikokommunikation an den Endnutzer. Regelmäßige Sessions, E-Learning-Module und Fallstudien verankern das Wissen und fördern eine „Responsible AI-Kultur". Erfolgskriterien sind Zertifizierungsraten und bei technischen Systemen zum Beispiel nachweisbar weniger Compliance Incidents

5. **Dokumentation und Berichterstattung**
   Welche Nachweise müssen wir in welcher Form vorhalten, um Behörden, Kunden und internen Auditoren die Konformität unserer KI-Systeme belegen zu können?
   Gefordert sind technische Dokumentation (Modellarchitektur, Trainings & Testdaten, Performance-Metriken), Gebrauchsanweisungen, Transparenzberichte und gegebenenfalls Einträge in die EU-Datenbank für Hochrisiko-KI. Eine zentrale Knowledge Base mit Versionierung erleichtert externe Audits und reduziert Reaktionszeiten auf behördliche Anfragen. Automatisierte Reporting Pipelines stellen sicher, dass Änderungen an Daten oder Modellen sofort in die Compliance-Dossiers einfließen.

Benedikt Backhaus
benedikt.backhaus@gmail.com

Über den Autor

Benedikt Backhaus ist selbstständiger KI-Berater und -Dozent mit Fokus auf den Mittelstand. Er unterstützt Unternehmen dabei, Künstliche Intelligenz strategisch und praxisnah im Arbeitsalltag zu verankern – mit einem klaren Blick jenseits von Hype und Technologie-Überforderung. Als Dozent für Plattformen wie Heise, DECAID Academy und mytalents.ai hat er bereits über 30 Videoformate zur KI-Anwendung in Unternehmen umgesetzt. Darüber hinaus schult er in Workshops und Live-Webinaren Mitarbeitende in der berufsbezogenen Anwendung von KI-Tools wie ChatGPT. Auf LinkedIn teilt er regelmäßig Impulse, Use Cases und Tipps zum produktiven KI-Einsatz. Mehr unter: *linkedin.com/in/benediktbackhaus*

SUPER SAVER BIS 30.09.2025

JavaLand

10. - 12. MÄRZ 2026

im EUROPA PARK® in Rust

#Javaland | www.javaland.eu

Präsentiert von: iJUG Verbund | heise medien | DOAG

Veranstalter: JavaLand

Warum KI-Projekte scheitern – und was wir dagegen unternehmen können

Alexander C. S. Hendorf, opotoc

KI-Projekte scheitern oft nicht an der Technologie, sondern an menschlichen und organisatorischen Hürden. In der Praxis zeigt sich immer wieder, dass die größten Stolpersteine nicht in Algorithmen oder Rechenleistung liegen, sondern in den Strukturen und Gewohnheiten von Unternehmen. Dieser Artikel beleuchtet die häufigsten Fallstricke und zeigt, wie man sie vermeiden kann, um KI erfolgreich im Unternehmen einzusetzen.

Stellen wir uns eine typische Szene vor: Herr Meier, Abteilungsleiter in einem Fachbereich, sieht auf einer Konferenz eine beeindruckende KI-Demo. Begeistert malt er sich aus, wie diese Technologie seine Arbeitsprozesse revolutionieren könnte. Voller Elan kehrt er zurück und fordert von der ohnehin überlasteten IT-Abteilung nun ebenfalls, „etwas mit KI“ umzusetzen. Was folgt, ist allzu oft ein steiniger Weg – und nicht selten das Scheitern des ambitionierten KI-Projekts.

Warum scheitern so viele KI-Initiativen, obwohl die zugrunde liegende Technologie immer leistungsfähiger wird? Studien beziffern die Quote der Fehlschläge auf 60–80 %, manche Schätzungen sprechen gar von über 85 %. In meinen über 15 Jahren Erfahrung mit KI-Projekten habe ich eines gelernt: Technische Hürden sind selten der Hauptgrund für Misserfolg. Vielmehr stolpern KI-Projekte überraschend oft über organisatorische und menschliche Faktoren. Ich formuliere es provokant:

„KI-Projekte scheitern immer an Menschen und Organisationen – nie an der Technik.“

Besonders in fünf Bereichen hakt es aus meiner Sicht: Erstens die übertriebene Fixierung auf aktuelle Hypes und „Shiny Objects“; zweitens die unterschätzte Bedeutung einer soliden Datenbasis; drittens das fehlende Verständnis wichtiger Stakeholder (Stichwort Halbwissen); viertens grundlegende Versäumnisse bei der Softwarequalität und im Engineering; und schließlich das Fehlen einer gelebten Innovationskultur. Im Folgenden beleuchte ich diese Problemfelder im Detail und zeige, warum gerade diese Punkte über Erfolg oder Misserfolg eines KI-Projekts entscheiden. Denn es reicht nicht, ein technisch ausgeklügeltes Modell zu bauen – wenn das Fundament nicht stimmt, bricht das ganze Konstrukt früher oder später in sich zusammen.

Wichtig dabei: Es ist selten ein einzelner Auslöser, der ein KI-Projekt scheitern lässt. Vielmehr kommen oft viele kleinere Hindernisse zusammen. Was mit berechtigter Euphorie beginnt, wandelt sich schrittweise in Frustration – bis das Projekt schließlich entnervt aufgegeben wird. Doch das muss nicht sein. Mit der richtigen Schwerpunktsetzung, klarer Strategie, ehrlicher Aufklärung, professionellem Engineering und einer offenen Innovationskultur können aus KI-Initiativen nachhaltige Erfolge werden.

Fixierung auf Hypes und „Shiny Tools“

In vielen Unternehmen herrscht die Vorstellung, man könne sich Künstliche Intelligenz einfach „einkaufen“. Kaum erscheint ein neues Tool mit KI-Funktionen auf dem Markt, wird es begeistert angeschafft – in der Hoffnung, dadurch automatisch einen Wettbewerbsvorteil zu erzielen. Ein CIO eines mittelständischen Unternehmens sagte einmal zu mir: „KI? Ja, das müssen wir jetzt auch kaufen.“ Diese Aussage bringt das Shiny-Object-Syndrome auf den Punkt: Der Glaube, dass der Erwerb eines fertigen KI-Produkts ohne weitere Anpassungen schon zum Erfolg führt, ist weit verbreitet.

Doch dieser Ansatz greift zu kurz. Moderne KI-Lösungen sind kein Selbstzweck, sondern sollen einen konkreten geschäftlichen Nutzen bringen. Wer lediglich dem neuesten Hype hinterherläuft, ohne ein klares Problem zu definieren, wird schnell enttäuscht. Allzu oft wird Technologie überschätzt und die eigentliche Aufgabenstellung vernachlässigt. Nach dem Motto „Wir probieren mal dieses neue Tool aus“, starten viele Projekte ohne langfristigen Plan oder Integration in die bestehenden Prozesse – und enden in Frust und ineffizientem Ressourceneinsatz. Laut einer McKinsey-Studie scheitern rund 70 % der KI-Projekte bereits in der Pilotphase, wenn ohne klare Strategie und Prozessanbindung experimentiert wird.

Warum bietet der bloße Softwarekauf keine Garantie für nachhaltigen Erfolg? Weil ein KI-Projekt mehr braucht als ein glänzendes Tool: Es benötigt Kontext, Daten und Einbettung. Ein isoliert eingeführtes KI-System mag technisch laufen, fügt sich aber oft nicht in die Abläufe ein oder wird von den Mitarbeitenden nicht akzeptiert. Der kurzfristige Glanz verfliegt schnell, wenn der praktische Nutzen ausbleibt. Die einzige sinnvolle Reihenfolge ist daher: erst das Geschäftsproblem und das Ziel definieren, dann die passende Technologie auswählen – nicht umgekehrt. Nur so lässt sich vermeiden, dass man mit Kanonen auf Spatzen schießt, indem man die neueste Highend-Lösung für ein eigentlich viel kleineres Problem einsetzt.

Die unterschätzte Rolle der Daten

Daten sind der Treibstoff jeder KI-Anwendung. Doch allzu oft wird die Datenqualität und -verfügbarkeit erst beachtet, wenn das KI-Modell schon entwickelt ist – und dann ist es meist zu spät. Schlechte oder unzureichende Daten zählen zu den Hauptgründen, warum KI-Projekte scheitern oder im schlimmsten Fall falsche Ergebnisse liefern, die lange unbemerkt bleiben. Es gilt der alte Grundsatz: Garbage in, garbage out. Ein KI-System kann nur so gut sein wie die Daten, mit denen man es füttert.

In der Praxis kämpfen Unternehmen typischerweise mit einer ganzen Reihe von Datenproblemen:

- Daten in Silos: Wichtige Informationen liegen verstreut in separaten Systemen und Abteilungen. Für bereichsübergreifende KI-Projekte sind diese Barrieren fatal – das Vorhaben scheitert, weil Daten nicht zusammengeführt werden können.
- Mangelnde Zugänglichkeit: Selbst wenn Daten vorhanden sind, sind sie oft nicht ohne Weiteres nutzbar. Häufig liegen sie in proprietären Formaten oder abgeschotteten Fachanwendungen ohne Schnittstellen (APIs) vor, sodass eine automatisierte Verarbeitung erschwert wird.
- Inkonsistenz und fehlende Dokumentation: Viele Datensätze leiden unter uneinheitlichen Formaten und unklaren Definitionen. Feldnamen und Codes sind nicht sauber dokumentiert, historische Daten weisen Lücken oder Widersprüche auf. Ein verlässliches Modell lässt sich darauf kaum trainieren.

Diese Defizite führen zu teils absurden Situationen. So stießen wir in einem Projekt auf hunderte archivierte Dateien, die zwar relevante Informationen enthielten, jedoch passwortgeschützt und jahrzehntealt waren. Die Passwörter kannte niemand mehr – die wenigen Mitarbeiter von damals waren längst im Ruhestand und Dokumentationen gab es nicht. Am Ende mussten wir tatsächlich einen Passwort-Cracker einsetzen, um an unsere eigenen Unternehmensdaten zu gelangen. Dieser Vorfall steht sinnbildlich für gewachsene IT-Strukturen, in denen nicht technische Grenzen, sondern fehlendes Wissensmanagement den Zugang zum „Unternehmensgedächtnis“ blockieren.

Die Lösung liegt auf der Hand: Ohne eine nachhaltige Datenstrategie wird ein KI-Projekt kaum fliegen. Bevor man mit dem Modellieren beginnt, muss die Datenbasis aufgeräumt und vereinheitlicht werden. Ebenso wichtig ist ein durchdachtes Data-Governance-Konzept: Es muss klar geregelt sein, wem die Daten „gehören“ und wer für ihre Qualität verantwortlich ist. Die Komplexität des Datenmanagements wird in vielen Häusern unterschätzt – diese Grundlagenarbeit ist mühselig, aber unverzichtbar. Dabei sollte man jedoch dem Reflex widerstehen, gleich die komplette Datenlandschaft perfekt zu bereinigen. Besser ist es, mit einem klar umrissenen, qualitativ hochwertigen Teilbestand zu starten. Versäumnisse aus Jahrzehnten lassen sich nicht über Nacht nachholen; wichtiger ist, überhaupt einen belastbaren Anfang zu machen.

Stakeholder: Halbwissen und Kommunikationslücken

KI-Projekte bringen fast immer mehrere Abteilungen an einen Tisch – von der IT über die Fachbereiche bis zur Geschäftsleitung. Unterschiedliche Perspektiven sind zwar wichtig, doch hier lauern erhebliche menschliche und organisatorische Stolpersteine. Häufig werden KI-Initiativen von Entscheidungsträgern gesteuert, die selbst nur wenig technisches Know-how besitzen. Gerade in höheren Managementebenen mangelt es oft an praktischer IT- und KI-Erfahrung. Das führt dazu, dass wichtige Entscheider zwar enorme Erwartungen an die KI-Lösung hegen, die technische Umsetzung aber nur abstrakt nachvollziehen können.

Typisch ist die Kluft zwischen Fachbereich und IT. Die Fachabteilung hat eine vage Idee – und wirft diese dann der IT „über den Zaun" in der Hoffnung, dass dort schon etwas Passendes programmiert wird. Auf beiden Seiten ist vielfach nur Halbwissen über KI vorhanden, was Missverständnisse provoziert. Ich erinnere mich an ein Projekt, bei dem der Fachbereich unbedingt ein „KI-Modell" haben wollte, aber eigentlich nur eine simple Regel-Engine benötigte. Solche falschen Vorstellungen führen leicht dazu, dass am falschen Problem gearbeitet oder unnötig komplizierte Lösungen entwickelt werden.

Erschwerend kommt hinzu, dass KI derzeit ein Modethema ist, zu dem fast jeder eine Meinung hat. Interdisziplinäre Zusammenarbeit ist zwar wertvoll, aber zu viele Köche verderben den Brei: Wenn ein KI-Projekt zu breit aufgestellt ist und jeder Beteiligte mitreden will, wird es unübersichtlich und langsam. Das erhöht das Risiko des Scheiterns erheblich.

Ein weiteres verbreitetes Problem ist die überzogene Erwartungshaltung gepaart mit mangelnder Geduld. Bleiben schnelle Erfolge aus, macht sich rasch Ernüchterung breit. Vielen Unternehmen fehlen hier eine gesunde Fehlerkultur und das Verständnis, dass KI-Modelle iterativ verbessert werden müssen. Ich habe erlebt, dass mancher Entscheider nach wenigen Wochen schon greifbare Wunderresultate sehen wollte – dabei benötigt ein brauchbares KI-Modell oft viele Monate kontinuierlicher Arbeit und Experimente.

Umso wichtiger ist es, Ziele realistisch zu formulieren und von Anfang an klar zu kommunizieren, welche Veränderungen und welcher Zeitrahmen mit der Einführung einer KI-Lösung verbunden sind. Alle Beteiligten müssen an Bord geholt und nötigenfalls weitergebildet werden. So wird aus Halbwissen echtes Verständnis. Nur dann kann die Einführung gelingen und das Projekt die notwendige Unterstützung im Unternehmen finden.

Organisationskultur: Angst vor Veränderung statt Innovationsfreude

Eine der größten Hürden für erfolgreiche KI-Projekte ist die Unternehmenskultur selbst. In manchen Firmen gilt implizit das Motto: *„Wer nichts entscheidet, macht auch nichts falsch."* Entscheidungen werden eher vermieden als ermöglicht. Ein Nein gilt als risikoarm, ein Ja als potenzielle Last, weil man Verantwortung übernehmen muss. Diese Haltung bremst jede Innovation aus. Hier ist vor allem die Führung gefragt: Sie muss Mitarbeitern den Rücken stärken und einen Rahmen schaffen, in dem mutige Entscheidungen belohnt statt bestraft werden.

Zugleich braucht es eine Kultur, die Neugier fördert und Fehler toleriert. KI erfordert mehr als nur eine Datenkultur – sie braucht eine echte Innovationskultur. Datengetriebenes Arbeiten (also Entscheidungen auf Basis von Fakten) ist ein guter Anfang, aber für KI-Projekte reichen Zahlen allein nicht. Ohne Fehlerkultur, Vertrauen und Risikobereitschaft kommt ein so neues Vorhaben schnell zum Erliegen. Wer Angst hat, wegen eines Fehlschlags abgestraft zu werden, wird lieber gar nichts ausprobieren.

Tatsächlich bringen KI-Initiativen oft Unbequemes ans Licht: Wissenslücken, ineffiziente Prozesse und Silodenken werden plötzlich sichtbar. Ein KI-Projekt wirkt wie ein Scheinwerfer, der in einen über Jahre hinweg gewachsenen Organisations-Dschungel leuchtet, in dem kaum jemand den vollständigen Überblick hat. Unterschiedliche Arbeitsweisen in den Abteilungen, fehlende oder widersprüchliche Prozesse – all das kommt nun zutage und muss hinterfragt werden. Das sorgt unvermeidlich für Reibung mit Bereichen, die ihre gewohnten Abläufe ungern ändern möchten.

Hier entscheidet sich, ob ein Unternehmen wirklich aus seinen KI-Piloten lernen will. Die Einführung einer KI-Lösung bietet die Chance, verkrustete Strukturen aufzubrechen und durch Standardisierung und Vereinfachung wieder mehr Überblick und Effizienz zu schaffen. Das erfordert jedoch die Bereitschaft, die sprichwörtliche Extrameile zu gehen – eine Investition, die sich langfristig auszahlt. Gerade am Anfang gilt: *keep it simple*. Statt sich in komplexen Abhängigkeiten zu verlieren, sollte man erste Erfolge mit überschaubaren Lösungen erzielen und darauf aufbauen.

Nicht zu unterschätzen ist auch der Faktor Angst und Besitzstandswahrung. Wissen wird in vielen Unternehmen gehortet, weil es als Jobgarantie gilt – wer ein Spezialwissen monopolisiert, fühlt sich unentbehrlich. Dieser soziale Ballast kann ein KI-Projekt enorm ausbremsen. In meiner Erfahrung kamen Projekte oft nur deshalb voran, weil es gelang, Vertrauen aufzubauen – nicht, weil alle Beteiligten die Technik sofort verstanden hätten. Hier ist wieder das Management gefragt: Es muss einen sicheren Raum schaffen, in dem Probleme offen angesprochen werden dürfen, ohne Angst vor Konsequenzen. Nur wenn sich die Menschen sicher fühlen, werden sie Fehler eingestehen, Hindernisse benennen und aktiv an echten Lösungen mitarbeiten.

Softwarequalität und interne Kompetenzen

Ein KI-Projekt besteht nicht nur aus dem Training eines Modells – es ist auch ein ausgewachsenes Software-Projekt. Nach meiner Erfahrung scheitern KI-Vorhaben selten an der KI-Technologie selbst, sondern viel häufiger an handwerklichen Defiziten im Software-Engineering. Unternehmen, die KI erfolgreich einsetzen wollen, brauchen daher eine solide Softwarebasis und müssen „Python-ready" sein. Das heißt: Entwickler mit den richtigen Fähigkeiten und eine Organisation, die moderne Entwicklungspraktiken beherrscht.

Python hat sich als dominierende Programmiersprache für Data Science und Machine Learning etabliert, aber die

Sprache allein genügt nicht. Zu einer Python-ready-Organisation gehört die Beherrschung zeitgemäßer Methoden und Werkzeuge der Softwareentwicklung. Wesentliche Bestandteile sind zum Beispiel:

- Sauberer Entwicklungs-Workflow: Klare Prozesse für das Schreiben, Versionieren und Deployen von Code.
- Tests: Unit-Tests und automatisierte Integrationstests, um die Funktionsfähigkeit bei jeder Änderung sicherzustellen.
- Code-Qualitätssicherung: Maßnahmen wie Code Reviews und Linter, um konsistenten, wartbaren Code zu gewährleisten.
- CI/CD-Pipelines: Automatisierte Build- und Deployment-Prozesse, um Änderungen schnell und zuverlässig in Produktion zu bringen.

Diese Punkte werden in vielen Unternehmen noch als „nice to have" abgetan, sind aber in Wahrheit essenziell für nachhaltigen Erfolg. KI-Systeme sind komplex und müssen regelmäßig angepasst oder neu trainiert werden – ohne professionelle Softwarestruktur gerät das schnell außer Kontrolle. Wer die Softwarequalität vernachlässigt, riskiert steigenden Wartungsaufwand, zunehmende Fehler und Frustration sowie letztlich steigende Kosten. Zudem schreckt eine chaotische Codebasis gute Leute ab: Top-Entwickler haben kaum Lust, in einem Umfeld ohne Standards und Tests zu arbeiten. Umgekehrt schafft eine konsequente Qualitätssicherung die Grundlage dafür, dass KI-Lösungen auch nach der Prototyp-Phase stabil weiterentwickelt und skaliert werden können. Das erhöht die langfristige Wettbewerbsfähigkeit erheblich.

Ein oft übersehener Erfolgsfaktor ist der Aufbau interner Kompetenzen. Viele Unternehmen versäumen es, das nötige Know-how aufzubauen, um KI-Projekte eigenständig umsetzen und betreiben zu können. Dazu gehören insbesondere Kenntnisse in:

- Python: Die wichtigste Sprache im KI-Umfeld.
- Open-Source-Tools und -Bibliotheken: Ein Großteil des KI-Fortschritts basiert auf frei verfügbaren Frameworks (von TensorFlow bis PyTorch) – man muss wissen, wie man diese effektiv einsetzt.
- Data Management: Umfassendes Wissen im Umgang mit Daten, von der Integration verschiedener Datenquellen bis zur Sicherung der Datenqualität.
- Linux: Das Betriebssystem der Wahl für viele KI-Infrastrukturen und Server.

Wer hier spart und glaubt, man könne sich KI-Erfolg einfach durch Zukauf von Cloud-Services oder externen Lösungen sichern, wird langfristig scheitern. Ein erfolgreiches KI-Projekt braucht interne Expertise und ein tiefes Verständnis der zugrunde liegenden Technologien.

Ich habe gelernt, die vorhandene Umgebung und das Know-how meiner Kunden kritisch zu hinterfragen. In einem Extremfall sahen wir uns gezwungen, einem Kunden den KI-Prototypen als komplett vorkonfigurierten Server zu übergeben, weil die chronisch unterbesetzte IT-Abteilung nicht in der Lage war, einen simplen Python-Dienst zum Laufen zu bringen. Dieser Fall mag besonders drastisch sein, aber er zeigt: Man muss realistisch einschätzen, welches Wissen im eigenen Haus vorhanden ist, was erst aufgebaut oder von extern hinzugeholt werden muss – und wie lange das dauert. Letztlich bildet fundiertes Software-Engineering gepaart mit geschulten Mitarbeitern das Rückgrat eines jeden KI-Projekts. Ohne dieses Rückgrat nützt der cleverste Algorithmus wenig.

Fazit

Zusammenfassend zeigt sich: KI-Projekte scheitern in der Regel nicht an fehlenden Algorithmen oder Rechenpower. Viel entscheidender sind organisatorische Weichenstellungen und realistische Erwartungen. Die weit verbreitete Illusion, man könne den Erfolg einfach durch den Einsatz des neuesten „Wundertools" erzwingen, führt in die Irre. Statt blind Technik zu kaufen, sollten Unternehmen den Fokus auf eine nachhaltige KI-Strategie legen, die eng mit den Geschäftsproblemen und -zielen verzahnt ist.

Ein zentrales Learning aus der Praxis ist die Bedeutung der Grundlagen: Datenqualität schlägt Modellkomplexität. KI-Projekte sollten von der Datenbasis ausgehend gedacht werden – Investitionen in saubere, gut integrierte Daten zahlen sich langfristig weit mehr aus als vorschnelle Experimente auf wackliger Informationsgrundlage. Ebenso wichtig sind die Menschen und Prozesse. Interdisziplinäre Zusammenarbeit, kontinuierliche Weiterbildung und eine Kultur des Ausprobierens ohne Angst vorm Scheitern sind Erfolgsfaktoren, die kein Algorithmus ersetzen kann. Führungskräfte müssen hier als Vorbilder agieren: Offenheit für neue Technologien zeigen, aber auch deren Grenzen realistisch einschätzen und kommunizieren.

Letztlich gilt: Wer KI nachhaltig erfolgreich einführen will, muss sowohl die technologischen als auch die organisatorischen und kulturellen Hausaufgaben machen. Nur wenn ein stabiles Fundament aus Daten, Kompetenzen und einer innovationsfreundlichen Kultur gelegt ist, lassen sich die vielbeschworenen KI-Potenziale tatsächlich heben – und zwar nicht nur in einer kurzfristigen Demo, sondern dauerhaft im produktiven Einsatz.

Über den Autor

Alexander C. S. Hendorf ist unabhängiger Berater mit über 20 Jahren Erfahrung in den Bereichen Digitalisierung, Daten und Künstlicher Intelligenz. Er unterstützt Organisationen – besonders in regulierten, von Altlasten geprägten Branchen – dabei, Daten- und KI-getriebene Strategien umzusetzen, die Geschäftserfolge sichern und nachhaltige Veränderungen bewirken. Als COO eines Musikunternehmens leitete er dessen digitale Transformation. Heute engagiert er sich mit Pioneers Hub und im Python Softwareverband als Vorstand für Open Source und inklusive in Deep-Tech-Communities und setzt sich für zukunftsfähige Innovation ein.

Alexander C. S. Hendorf
hendorf@opotoc.com

Erste Schritte mit Transparent Data Encryption (TDE) – Teil 2

Meris Bihorac, DBConcepts

Aufgrund der zunehmenden digitalen Bedrohungen, die die Datensicherheit gefährden, ist das Verständnis und die Umsetzung robuster Schutzstrategien von entscheidender Bedeutung. In diesem Beitrag werden Hardware-Sicherheitsmodule (HSM) vorgestellt – hochentwickelte Geräte zur Verwaltung und Sicherung kryptografischer Schlüssel. Im Anschluss daran konzentrieren wir uns auf die praktische Anwendung der transparenten Datenverschlüsselung (TDE) und beschreiben die Konfiguration und Sicherheitstests einer minimalen TDE-Umgebung. Zusätzlich werden wir die Auswirkungen von TDE auf den Backup-Vorgang betrachten und die Lizenzanforderungen für den Einsatz solcher Verschlüsselungstechnologien erläutern.

Hardwarebasierte Lösungen:

Ein Hardware-Sicherheitsmodul (HSM) ist ein spezielles Hardwaregerät, das für die sichere Speicherung, Verwaltung und Verwendung von kryptografischen Schlüsseln und Geheimnissen entwickelt wurde. Die Verwendung eines HSM für die Verwaltung von Wallet-Einträgen oder Verschlüsselungsschlüsseln bietet eine zusätzliche Sicherheitsebene, da diese in der Regel getrennt von dem Servicierten System betrieben wird.

Beispiele:

- Thales – Luna HSM (On Premise)
- Entrust – nShield HSM (On Premise)
- AWS – AWS CloudHSM (Cloud)
- Google – Google Cloud HSM (Cloud) [1]

Best Practices für Wallet:

Das Wallet ist ein wesentlicher Bestandteil des Sicherheitssystems und muss sorgfältig behandelt werden, um den Schutz der Daten durchgängig zu gewährleisten. Es ist unerlässlich, regelmäßige Backups des Wallets sowohl lokal als auch extern an sicheren Orten aufzubewahren. Unmittelbar nach der Erstellung muss der Master-Key gesichert werden, und dies sollte jedes Mal wiederholt werden, wenn Änderungen vorgenommen werden.

Ownership and Permissions

Zum weiteren Schutz des Wallets wird empfohlen, den Besitzer auf den Benutzer Oracle zu übertragen und genaue Berechtigungen festzulegen *(siehe Listing 1)*.

Immutable Wallet Entries

Die Sicherstellung der Unveränderbarkeit von Wallet-Einträgen verhindert versehentliches Löschen oder Ändern und erhöht die Sicherheit. Die folgenden Befehle machen die Wallet-Dateien wie ewallet.p12 und cwallet.sso nicht löschbar *(siehe Listing 2)*.

Weitere mögliche Sicherheitsoptionen für das Wallet:

- Read-Only Mountpoint
- Betriebssystem-Härtung
  - Anwendung des Prinzips der geringsten Privilegien
  - Verschlüsselung auf Betriebssystem-ebene verwenden (zum Beispiel BitLocker, LUKS)
- Verwendung eines Hardware-Sicherheitsmoduls (HSM)

Unterstützte Verschlüsselungsalgorithmen für Transparent Data Encryption

(siehe Tabelle 1)

Datenbankverschlüsselung Konvertierungen

(siehe Tabelle 2)

Vergleich zwischen Tablespace und Column Encryption

Tablespace Encryption

- Verschlüsselung des gesamten Tablespace (alle Datenfiles)
- Es kann gewählt werden, welche Tablespaces verschlüsselt werden und welche nicht.
- Tablespace Key (derselbe Key, der von allen Datenfiles in einem Tablespace verwendet wird)
- Standard-Algorithmus: AES 128
- Die Daten werden entschlüsselt, wenn sie aus den Datenfiles in den Buffer-Cache oder über einen „Direct Path Read“ gelesen werden.

Column Encryption

- Verschlüsselung einzelner Spalten innerhalb einer Tabelle
- Eigener Table Key pro Tabelle (derselbe Key, der von allen verschlüsselten Spalten in einer Tabelle verwendet wird)
- Standard-Algorithmus: AES 192
- Daten werden entschlüsselt, wenn sie in die PGA verschoben werden. [4]

Einrichten der Tablespace Encryption

1. Konfigurieren Sie den Speicherort für das Wallet in der Datei sqlnet.ora *(siehe Listing 3)*.
2. Erstellen Sie das Wallet und aktivieren Sie es *(siehe Listing 4)*.
3. Erstellen Sie den Master Key *(siehe Listing 5)*.
4. Erstellen Sie einen verschlüsselten Tablespace *(siehe Listing 6)*.
5. Erstellen Sie eine Tabelle im verschlüsselten und eine weitere Tabelle im unverschlüsselten Tablespace *(siehe Listing 7 und 8)*.
6. Auslesen des Datenfiles auf Betriebssystemebene *(siehe Listing 9)*.

```
chown -R oracle:oinstall /path/to/wallet
chmod 700 /path/to/wallet
chmod -R 600 /path/to/wallet/*
```

Listing 1: Berechtigungen einstellen

```
chattr +i ewallet.p12
chattr +i cwallet.sso
```

Listing 2: Attribut für eine Datei setzen

```
ENCRYPTION_WALLET_LOCATION =
  (SOURCE =
      (METHOD = FILE)
    (METHOD_DATA =
            DIRECTORY = /home/oracle/wallet)
      )
  )
```

Listing 3: sql.net Konfiguration des Speicherorts

```
mkdir -p /home/oracle/wallet

ADMINISTER KEY MANAGEMENT CREATE KEYSTORE '/home/oracle/wallet' IDENTIFIED BY "SecurePassword123!";

ADMINISTER KEY MANAGEMENT SET KEYSTORE OPEN IDENTIFIED BY "SecurePassword123!";


set pages 400 lines 200
col wrl_parameter form a50
select wrl_type,wrl_parameter,status,wallet_type,wallet_order from v$encryption_wallet;

WRL_TYPE   WRL_PARAMETER            STATUS               WALLET_TYPE   WALLET_OR
---------- ------------------------ -------------------- ------------- ---------
FILE       /home/oracle/wallet/     OPEN_NO_MASTER_KEY   PASSWORD      SINGLE
```

Listing 4: Erstellung des Wallets

```
ADMINISTER KEY MANAGEMENT SET ENCRYPTION KEY IDENTIFIED BY "SecurePassword123!" WITH BACKUP;

select wrl_type,wrl_parameter,status,wallet_type,wallet_order from v$encryption_wallet;
WRL_TYPE    WRL_PARAMETER               STATUS      WALLET_TYPE  WALLET_OR
----------- --------------------------- ----------- ------------ ---------
FILE        /home/oracle/wallet/        OPEN        PASSWORD     SINGLE


SQL> select key_use,keystore_type,key_id from v$encryption_keys;
KEY_USE    KEYSTORE_TYPE     KEY_ID
---------- ----------------- ------------------------------------------------------
TDE        SOFTWARE KEYSTORE AaUnq+4MpU9UvzWUmGcuop4AAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Listing 5: Master Key erstellen

```
CREATE BIGFILE TABLESPACE bank_tbs
DATAFILE '/u01/app/oracle/oradata/TESTDB01/bank_tbs01.dbf' SIZE 10G
AUTOEXTEND ON NEXT 1G  MAXSIZE 30G
ENCRYPTION USING 'AES256' DEFAULT STORAGE(ENCRYPT);

SELECT t.name AS tablespace_name, e.ENCRYPTIONALG
FROM v$tablespace t
JOIN v$encrypted_tablespaces e
ON t.ts# = e.ts#;

TABLESPACE_NAME                ENCRYPT
------------------------------ -------
BANK_TBS                       AES256
```

Listing 6: Verschlüsselten Tablespace anlegen

Verschlüsselte Tabelle

```
CREATE TABLE BANK_USER.customers (
    customer_id NUMBER GENERATED BY DEFAULT AS IDENTITY PRIMARY KEY,
    name VARCHAR2(100),
    dob DATE,
    address VARCHAR2(200),
    phone_number VARCHAR2(15),
    email VARCHAR2(100),
    credit_card_number VARCHAR2(16) NOT NULL,
    card_expiry_date DATE NOT NULL,
    created_at DATE DEFAULT SYSDATE
) TABLESPACE BANK_TBS;
```

Listing 7: Tabelle im verschlüsselten Tablespace erstellen

Wie oben ersichtlich, haben wir zwei Tabellen erstellt, eine im verschlüsselten und eine im „normalen" Tablespace, und in beide Tabellen „vertrauliche" Daten eingefügt

Auf Betriebssystemebene können wir mit Utility Strings klar erkennen, welcher Tablespace verschlüsselt ist und welcher nicht *(siehe Listing 10)*.

Datapump mit TDE

Sie haben sich gefragt, ob die verschlüsselten Daten sicher sind, wenn sie mit dem Oracle Data Pump Export normal exportiert werden? Die Antwort lautet **NEIN!** Einen Beispiel-Export zeigt *Listing 11*.

Wie Sie bereits erkennen können, sind die Daten jetzt in einer unverschlüsselten Dump-Datei gespeichert worden.

Beim Import in eine andere Datenbank wird festgestellt, dass der Tablespace aufgrund des fehlenden Master Keys nicht erstellt werden kann und die Daten daher auch nicht importiert werden können *(siehe Listing 12)*.

Wenn Sie den normalen Tablespace manuell erstellen, so können Sie die Da-

Unverschlüsselte Tabelle

```
CREATE TABLE BANK_USER.customers_not_enc (
    customer_id NUMBER GENERATED BY DEFAULT AS IDENTITY PRIMARY KEY,
    name VARCHAR2(100),
    dob DATE,
    address VARCHAR2(200),
    phone_number VARCHAR2(15),
    email VARCHAR2(100),
    credit_card_number VARCHAR2(16) NOT NULL,
    card_expiry_date DATE NOT NULL,
    created_at DATE DEFAULT SYSDATE
) TABLESPACE USERS;
```

Listing 8: Tabelle im „normalen“ (unverschlüsselten) Tablespace erstellen

```
INSERT INTO BANK_USER.customers_not_enc  (name, credit_card_number, card_expiry_date) VALUES (Secret Not En-
crypted', '4111111111111111', TO_DATE('2028-12-31', 'YYYY-MM-DD'));

INSERT INTO BANK_USER.customers  (name, credit_card_number, card_expiry_date) VALUES (' Secret Encrypted',
'4111111111111111', TO_DATE('2028-12-31', 'YYYY-MM-DD'));
```

Listing 9: Daten in die Tabellen einfügen

```
[oracle@TESTDB01]$ strings users01.dbf | grep Secret
Secret Not Encrypted
[oracle@TESTDB01]$ strings bank_tbs01.dbf | grep Secret
[oracle@TESTDB01]$
```

Listing 10: Auslesen der Datafiles mit Hilfe von Utility Strings

```
Export: Release 19.0.0.0.0 - Production on Mon Dec 9 17:40:59 2024
Version 19.25.0.0.0

Copyright (c) 1982, 2019, Oracle and/or its affiliates.  All rights reserved.

Connected to: Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Starting "SYS"."SYS_EXPORT_SCHEMA_01":  "/******** AS SYSDBA" schemas=BANK_USER directory=datapump_dir dump-
file=bank_user_%U.dmp logfile=bank_user_export.log parallel=4
Processing object type SCHEMA_EXPORT/TABLE/INDEX/STATISTICS/INDEX_STATISTICS
Processing object type SCHEMA_EXPORT/TABLE/STATISTICS/TABLE_STATISTICS
Processing object type SCHEMA_EXPORT/USER
Processing object type SCHEMA_EXPORT/SYSTEM_GRANT
Processing object type SCHEMA_EXPORT/ROLE_GRANT
Processing object type SCHEMA_EXPORT/DEFAULT_ROLE
Processing object type SCHEMA_EXPORT/TABLESPACE_QUOTA
Processing object type SCHEMA_EXPORT/PRE_SCHEMA/PROCACT_SCHEMA
Processing object type SCHEMA_EXPORT/TABLE/TABLE_DATA
Processing object type SCHEMA_EXPORT/STATISTICS/MARKER
Processing object type SCHEMA_EXPORT/TABLE/TABLE
Processing object type SCHEMA_EXPORT/TABLE/IDENTITY_COLUMN
Processing object type SCHEMA_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
. . exported "BANK_USER"."CUSTOMERS"                      12.03 KB      30 rows
. . exported "BANK_USER"."CUSTOMERS_NOT_ENC"              8.531 KB       1 rows
ORA-39173: Encrypted data has been stored unencrypted in dump file set.
Master table "SYS"."SYS_EXPORT_SCHEMA_01" successfully loaded/unloaded
******************************************************************************
Dump file set for SYS.SYS_EXPORT_SCHEMA_01 is:
  /home/oracle/bank_user_01.dmp
  /home/oracle/bank_user_02.dmp
  /home/oracle/bank_user_03.dmp
Job "SYS"."SYS_EXPORT_SCHEMA_01" successfully completed at Mon Dec 9 17:41:21 2024 elapsed 0 00:00:19
```

Listing 11: Schema-Export

```
Connected to: Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Mastertabelle "SYS"."SYS_IMPORT_FULL_01" erfolgreich geladen/entladen
"SYS"."SYS_IMPORT_FULL_01":  "/******** AS SYSDBA" full=y directory=EXP_DIR dumpfile=bank_user_%U.dmp logfile=-
bank_user_import3.log parallel=4  wird gestartet
Objekttyp DATABASE_EXPORT/PRE_SYSTEM_IMPCALLOUT/MARKER wird verarbeitet
Objekttyp DATABASE_EXPORT/PRE_INSTANCE_IMPCALLOUT/MARKER wird verarbeitet
Objekttyp DATABASE_EXPORT/TABLESPACE wird verarbeitet
ORA-31684: Objekttyp TABLESPACE:"UNDOTBS1" ist bereits vorhanden
ORA-31684: Objekttyp TABLESPACE:"TEMP" ist bereits vorhanden
ORA-31684: Objekttyp TABLESPACE:"USERS" ist bereits vorhanden
ORA-39083: Objekttyp TABLESPACE:"BANK_TBS" konnte nicht erstellt werden, Fehler:
ORA-28361: Master-Schlussel noch nicht festgelegt
```

Listing 12: Schema-Import

```
CREATE SMALLFILE TABLESPACE bank_tbs
    DATAFILE  SIZE 2G
    AUTOEXTEND ON NEXT 256M MAXSIZE UNLIMITED;
Copyright (c) 1982, 2019, Oracle and/or its affiliates.  All rights reserved.

Connected to: Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Mastertabelle "SYS"."SYS_IMPORT_FULL_01" erfolgreich geladen/entladen
"SYS"."SYS_IMPORT_FULL_01":  "/******** AS SYSDBA" full=y directory=EXP_DIR dumpfile=bank_user_%U.dmp logfile=-
bank_user_import.log parallel=4  wird gestartet
Objekttyp SCHEMA_EXPORT/USER wird verarbeitet
Objekttyp SCHEMA_EXPORT/DEFAULT_ROLE wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLESPACE_QUOTA wird verarbeitet
Objekttyp SCHEMA_EXPORT/PRE_SCHEMA/PROCACT_SCHEMA wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLE/TABLE wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLE/TABLE_DATA wird verarbeitet
. . "BANK_USER"."CUSTOMERS_ENC"                 8.531 KB       1 Zeilen importiert
. . "BANK_USER"."CUSTOMERS_NOT_ENC"             8.546 KB       1 Zeilen importiert
Objekttyp SCHEMA_EXPORT/TABLE/IDENTITY_COLUMN wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLE/CONSTRAINT/CONSTRAINT wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLE/INDEX/STATISTICS/INDEX_STATISTICS wird verarbeitet
Objekttyp SCHEMA_EXPORT/TABLE/STATISTICS/TABLE_STATISTICS wird verarbeitet
Objekttyp SCHEMA_EXPORT/STATISTICS/MARKER wird verarbeitet
```

Listing 13: Erstellen eines normalen Tablespace und erneutes Importieren

ten problemlos importieren, da die Daten ursprünglich im obigen Beispiel Export ohne Verschlüsselung exportiert wurden.

Erstellen Sie einen normalen Tablespace und importieren Sie die Daten erneut *(siehe Listing 13)*.

Sie sehen, dass sowohl die verschlüsselten Daten aus der Tabelle CUSTOMERS_ENC, als auch die unverschlüsselten aus CUSTOMERS_NOT_ENC, problemlos importiert wurden.

Außerdem ist es möglich, direkt über das Betriebssystem auf die Daten aus den Dumps zuzugreifen. Wenn Sie das Utility Strings unter Linux verwenden, um die Dumps zu analysieren, so können Sie sowohl die verschlüsselten als auch die unverschlüsselten Einträge eindeutig identifizieren. Dieses Verhalten ist jedoch darauf zurückzuführen, dass die Verschlüsselungsoption für den Export mittels Datapump nicht verwendet wurde. Auf dem System, auf dem der verschlüsselte Tablespace ursprünglich eingerichtet wurde, bleibt das verschlüsselte Geheimnis unlesbar, wie im ersten Beispiel oben demonstriert *(siehe Listing 14)*.

Zusammenfassend kann festgehalten werden, dass Transparent Data Encryption Ihre Daten auf den Filesystemen und den Speichersystemen darunter hervorragend schützt. Für die Daten ins Frontend – zu den Applikationen – ist die Verschlüsselung transparent und muss daher durch ergänzende Methoden wie etwa Netzwerkverschlüsselung, Zugriffschutz und anderes gewährleistet werden.

Auswirkungen von TDE auf das Backup

Die Auswirkung der TDE auf die Sicherung muss berücksichtigt werden, sie ist abhängig vom Sicherungsmechanismus und kann zu Problemen führen. Nachfolgend wird ein kurzer Überblick über den Sicherheitsmechanismus gegeben:

Algorithmus	Schlüssel-Größe	Parameter-Name
Advanced Encryption Standard (AES)	• 128 bits (default for tablespace encryption) • 192 bits (default for column encryption) • 256 bits	• AES128 • AES192 • AES256
ARIA	• 128 bits • 192 bits • 256 bits	• ARIA128 • ARIA192 • ARIA256
GOST	• 256 bits	• GOST256
SEED	• 128 bits	• SEED128
Triple Encryption Standard (DES)	• 168 bits	• 3DES168

Tabelle 1

Funktionalität	Offline-Konvertierung	Online-Konvertierung
Release mit minimaler Konvertierbarkeit	Oracle Database 11g release 2 (11.2)	Oracle Database 12c release 2 (12.2) and later
Unterstützte Algorithmen	Alle symmetrischen Verschlüsselungsalgorithmen, die TDE unterstützt.	Alle symmetrischen Verschlüsselungsalgorithmen, die TDE unterstützt.
Wann kann die Konvertierung durchgeführt werden?	Wenn der Tablespace offline ist oder sich die Datenbank in der Mount-Phase befindet.	Wenn der Tablespace online ist und die Datenbank im Lese-/Schreibmodus geöffnet ist.
Wird für den Umbau zusätzlicher Platz benötigt?	Nein	Ja
Oracle Data Guard Konvertierungsrichtlinien	Sowohl der Primär- als auch der Standby-Server müssen manuell umgestellt werden. Zuerst den Standby konvertieren und dann umschalten, um die Ausfallzeit zu minimieren.	Nach der Konvertierung des Primärsystems erfolgt die Konvertierung des Standby-Systems automatisch. Eine Online-Konvertierung kann nicht direkt auf dem Standby-System durchgeführt werden.
Können Verschlüsselungsschlüssel erneut verschlüsselt werden?	Nein, aber nachdem der Tablespace verschlüsselt wurde, können Sie die Online-Konvertierung verwenden, um erneut einen Schlüssel zu erstellen, allerdings in Kompatibilität mit Oracle Database 12c Release 2 (12.2).	Ja
Können Verschlüsselungsoperationen parallel durchgeführt werden?	Parallele Verschlüsselungskonvertierungen auf der Ebene der Datendateien können mit mehreren Benutzersitzungen durchgeführt werden.	Parallel laufende Verschlüsselungskonvertierungen auf Tablespace-Ebene können mit mehreren laufenden Benutzersitzungen durchgeführt werden.

Tabelle 2

Komprimierung

Durch Komprimierung wird die Größe von Datendateien verringert, indem Algorithmen zur Verdichtung der Daten angewendet werden, wodurch der benötigte Speicherplatz reduziert und die Datenübertragung beschleunigt wird.

Bei der Komprimierung von verschlüsselten Daten kann die Komprimierungsrate deutlich verschlechtert werden. Daher ist das im Backupzyklus unbedingt zu beachten. Speziell bei großen Datenmengen kann die Umstellung auf die Verschlüsselung zu Speicherplatzproblemen im Backup-Backend führen.

```
[oracle@TESTDB02 export]$ ll
total 444
-rw-r-----. 1 oracle oinstall  20480 Dec  9 18:03 bank_user_01.dmp
-rw-r-----. 1 oracle oinstall  32768 Dec  9 18:03 bank_user_02.dmp
-rw-r-----. 1 oracle oinstall 389120 Dec  9 18:03 bank_user_03.dmp
-rw-r-----. 1 oracle oinstall  12288 Dec  9 18:03 bank_user_04.dmp
[oracle@TESTDB02 export]$ strings * | grep Secret
Secret Encrypted
Secret Not Encrypted
```

Listing 14: Auslesen der Daten-Dumps mit dem Strings Utility

Deduplizierung

Die Deduplizierung reduziert den Speicherbedarf, in dem wiederholende Daten während des Sicherungsprozesses identifiziert und referenziert werden. Es wird nur eine einzige Kopie der Daten gespeichert, wobei nachfolgende idente Daten durch Verweise auf die gespeicherten Daten ersetzt werden, was die Speichereffizienz deutlich verbessern kann.

Anfangs kann aufgrund der Umstellung auf eine Verschlüsselung die Datenmenge, die nicht dedupliziert werden kann, markant ansteigen. Weitere Backups im Nachgang können aber durchaus von einer Deduplizierung profitieren.

Desweiteren gilt es zu beachten, dass die Deduplizierungsrate von Daten innerhalb eines Backups nur noch bedingt funktional ist (ähnlich wie bei der Komprimierung).

Thin Provisioning

Beim Thin Provisioning wird der Speicherplatz dynamisch auf der Grundlage des tatsächlichen Datenbedarfs zugewiesen, anstatt große Mengen an physischem Speicher im Voraus zu reservieren. Diese Methode verbessert die Speichernutzung, indem sie den Speicherplatz nach Bedarf bereitstellt und eine Überbelegung und Verschwendung verhindert.

Thin Provisioning bleibt von Transparent Data Encryption unbeeinflusst. *[4]*

TDE-Lizenzierung

Die Transparente Datenverschlüsselung ist Teil des Pakets Advanced Securtiy Option und ist nur für die Enterprise Edition (EE) verfügbar. *[5]*

Bei Oracle Database Services ist die transparente Datenverschlüsselung standardmäßig integriert sowie aktiv und erfordert keine separate Lizenzierung. Dies gilt auch, wenn Sie Ihre eigene Lizenz mitbringen (BYOL). Dadurch ist TDE ohne zusätzliche Kosten in verschiedenen Oracle-Datenbankumgebungen, einschließlich Cloud- und On-Premise-Konfigurationen, leicht zugänglich.

Im Rahmen von Cloud at Customer (C@C) ist TDE auch ohne gesonderte Lizenzgebühren erhältlich, selbst im BYOL-Modell. Dieser Ansatz von Oracle stellt sicher, dass Sicherheitsfunktionen wie TDE leicht verfügbar sind, was ihre Verwendung zum Schutz sensibler Daten über verschiedene Bereitstellungsmodelle hinweg fördert. *[6]*

Quellen

[1] E. Davies, „hardware security module (HSM)," [Online]. Available: *https://www.techtarget.com/searchsecurity/definition/hardware-security-module-HSM.*

[2] Oracle, „Advanced Security Guide," Oracle, [Online]. Available: *https://docs.oracle.com/en/database/oracle/oracle-database/21/asoag/introduction-to-transparent-data-encryption.html#GUID-0E59AD1E-514A-4F93-9A1F-E5683A535A89.*

[3] Oracle, „Advanced Security Guide," Oracle, [Online]. Available: *https://docs.oracle.com/en/database/oracle/oracle-database/21/asoag/encryption-conversions-tablespaces-and-databases1.html#GUID-79072DFC-4DC6-4920-9B2A-D649123D15C7.*

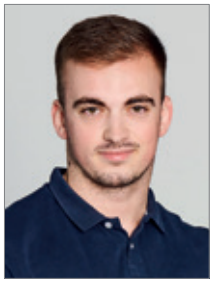
[4] „Back to basics with Transparent Data Encryption (TDE)," Oracle, [Online]. Available: *https://asktom.oracle.com/ords/r/tech/catalog/session-landing-page?p2_event_id=339826926174340904386935319045761301740.*

[5] Oracle, „Database Licensing Information," [Online]. Available: *https://docs.oracle.com/cd/E11882_01/license.112/e47877/options.htm#DBLIC143.*

[6] Oracle, „Database Licensing Information User Manual," [Online]. Available: *https://docs.oracle.com/en/database/oracle/oracle-database/19/dblic/Licensing-Information.html#GUID-B6113390-9586-46D7-9008-DCC9EDA45AB4.*

Über den Autor

Mein Name ist Meris Bihorac und ich bin ein 24-jähriger Oracle DBA aus Wien. Während meiner Ausbildung habe ich mich zunächst mehr auf die Entwicklung von Datenbanken konzentriert. Als ich jedoch die umfangreichen Features von Oracle entdeckte, wie Real Application Clusters (RAC), Data Guard und Hochverfügbarkeitslösungen, verlagerte sich mein Interesse deutlich in Richtung Administration. Diese neu entdeckte Faszination hat sich seitdem zu einer tiefen Leidenschaft für die Optimierung und Verwaltung von Datenbankumgebungen entwickelt. Daneben reizte mich schon in jungen Jahren die Herausforderung, nach Bugs zu suchen, insbesondere nach solchen, die die Web-Sicherheit betreffen, wie SQL-Injections. In meiner Freizeit habe ich als Bug-Hunter gearbeitet und dabei verschiedene bedeutende Schwachstellen aufgedeckt, darunter eine große Sicherheitslücke, von der über 200.000 Benutzer betroffen waren. Diese Erfahrungen haben meine berufliche Laufbahn nachhaltig geprägt und mein Engagement für hervorragende Leistungen im Bereich der Datenbankverwaltung weiter vorangetrieben.



Meris Bihorac
meris.bihorac@dbconcepts.com

Programmieren nach Daten – Teil 1

Jürgen Sieben, ConDes

Dies ist eine kleine Reihe mit Anwendungsbeispielen, die zeigen, wie der Einsatz von SQL in der Programmierung der Datenbank nicht nur erhebliche Codemengen einsparen, sondern zudem auch die Effizienz und Funktionalität Ihrer Anwendungen erhöhen kann. Hier geht es nicht um abgefahrene Spezialfunktionen, sondern um eine Erweiterung des Blickwinkels in der Programmierung mit PL/SQL.

Delta-View

Delta-View ist ein Beispiel aus einem Projekt, die Daten sind vereinfacht und stark reduziert, das Problem ist aber real. In einer Anwendung werden Berechtigungen für Mitarbeiter erteilt, bis zu einem gegebenen Maximalbetrag Geld an externe Unternehmen anzuweisen. Da es im Projekt zum Teil um bis zu dreistellige Millionenbeträge ging, dürfte nachvollziehbar sein, dass dieser Bereich der Anwendung unter erhöhter Aufmerksamkeit bezüglich der Sicherheit stand. Ein Teil der Berechtigungserteilung ist an die Zugehörigkeit des Mitarbeiters zu einer Abteilung gebunden, da für gewisse Abteilungen andere Maximalbeträge gepflegt werden als für andere. Daher ist es erforderlich, das aktuelle Organigramm der Unternehmung zu kennen.

Dieses Organigramm wird in einem externen Programm (SAP) gepflegt und über eine Schnittstelle täglich an unsere Anwendung übermittelt. Im Gegensatz zu unserer Anwendung pflegt SAP allerdings keine historische Übersicht, sondern mel-

```
SQL> select best_id, best_parent_id, best_valid_from, best_valid_until
  2    from bestand;

BEST_ID BEST_PARENT_ID BEST_NAME          BEST_VALID_FROM BEST_VALID_UNTIL
------- -------------- ------------------ --------------- ----------------
    123                Abteilung A        12.05.1985      31.12.2999
    234            123 Unterabteilung A1  14.06.2000      31.12.2999
    345                Abteilung B        13.04.1990      31.12.2999
    456            345 Unterabteilung B1  13.04.1990      31.12.2999
    567            345 Unterabteilung B2  06.05.2000      31.12.2999

SQL> select sap_id, sap_parent_id, sap_name
  2    from sap;

SAP_ID SAP_PARENT_ID SAP_NAME
------ ------------- -----------------
   234           345 Unterabteilung B2
   345               Abteilung B
   456           345 Unterabteilung B1
   567           345 Unterabteilung B3
   678           345 Unterabteilung B4
```

Listing 1: Beispielsdaten BESTAND

det nur den aktuell gültigen Stand. Unsere Anwendung muss jedoch nachzeichnen können, warum ein Mitarbeiter vor drei Monaten eine Berechtigung erteilt bekommen hat. Daher müssen die Daten des Organigramms in unserer Anwendung historisch abgelegt werden. Da einige Änderungen am Organigramm weitere Konsequenzen nach sich zogen (ein Team kann zum Beispiel einer anderen Abteilung zugeordnet werden, wodurch die an alle dort arbeitenden Mitarbeiter erteilten Berechtigungen entzogen werden müssen), reichte eine einfache SQL-Anweisung nicht aus, es musste Logik programmiert werden. Zudem musste jede Änderung an Berechtigungen protokolliert werden.

Das Unternehmen ist groß und weit verzweigt, der Umfang der Organigrammdaten ist zwar nicht riesig, aber auch nicht trivial: etwa 6.000 Zeilen werden durch SAP täglich gemeldet und müssen abgeglichen werden. Dabei müssen folgende Szenarien unterschieden werden:

- Welche Daten wurden neu erstellt, geändert oder gelöscht?
- Sind Änderungen aufgetreten, die das Entziehen von Berechtigungen erfordern?

Bespieldaten

Um Ihnen das Prinzip zu erläutern, verwende ich sehr vereinfachte Daten. *Listing 1* zeigt eine Tabelle BESTAND und eine Quelltabelle SAP, die das aktuelle Organigramm enthält. Wir erkennen eine einfache Hierarchie über die Spalten ORG_ID und ORG_PARENT_ID sowie ein fachliches Gültigkeitsband, das für die spätere Historisierung verwendet wird.

Strukturell sind beide Tabellen bis auf das fehlende Gültigkeitsband in SAP gleich, was aber keine Voraussetzung für das folgende Verfahren ist, sondern lediglich der Vereinfachung dient. Wir erkennen, dass die Daten nicht gleich sind, aber schon bei diesen 5 Zeilen ist es nicht trivial, zu erkennen, welche Änderungen genau aufgetreten sind und welche Konsequenzen daraus zu ziehen sind: Welche Zeilen sind neu, welche nicht mehr vorhanden, welche Art von Änderung hat es gegeben? Vielleicht ist nur der Name einer Organisationseinheit geändert worden, was trivial wäre, vielleicht ist eine Einheit aber auch umgezogen?

Es scheint keine Alternative zu geben als die, über eine der beiden Relationen Zeile für Zeile zu iterieren, sich die jeweils zugehörige Zeile der anderen Relation anzusehen und auf Änderungen zu untersuchen. Da wir nicht sehen können, welche Zeile in der jeweils anderen Relation in der eigenen Relation nicht mehr vorhanden ist, müssen wir anschließend noch eine Gegenprüfung von der anderen Relation aus durchführen. Es kommt hinzu, dass aufgrund der Natur der Daten nicht davon auszugehen ist, dass viele Änderungen auftreten, immerhin handelt es sich um das Organigramm einer Unternehmung, da werden von einem Tag zum anderen nur wenige Zeilen geändert, wenn überhaupt eine. Ist vor diesem Hintergrund der Aufwand, alle 6.000 Zeilen einzeln zu prüfen, sinnvoll?

Lösungsansatz

Zur Lösung dieses Problems werde ich einige simple Tricks verwenden. Die Kunst ist hier, auf diese Tricks zu kommen, nicht die technische Raffinesse der Anweisung.

Erforderliche DML-Aktion erkennen

Kümmern wir uns zunächst um das Problem, zu erkennen, ob eine insert-, update- oder delete-Anweisung auf den Bestand erforderlich ist, und noch nicht darum, ob

```
SQL> select best_id, best_name, sap_id, sap_name
  2    from bestand
  3    full join sap
  4      on best_id = sap_id;

BEST_ID BEST_NAME         SAP_ID SAP_NAME
------- ----------------- ------ -----------------
    234 Unterabteilung A1    234 Unterabteilung B2
    345 Abteilung B          345 Abteilung B
    456 Unterabteilung B1    456 Unterabteilung B1
    567 Unterabteilung B2    567 Unterabteilung B3
                             678 Unterabteilung B4
    123 Abteilung A
```

Listing 2: Full Join zwischen den Relationen

```
SQL> select coalesce(sap_id, best_id) best_id,
  2         sap_parent_id best_parent_id, sap_name best_name,
  3         case when best_id is null then 'I'
  4              when sap_id is null then 'D'
  5              else 'U' end best_dml_action
  6    from bestand
  7    full join sap
  8      on best_id = sap_id;

BEST_ID BEST_PARENT_ID BEST_NAME         BEST_DML_ACTION
------- -------------- ----------------- ---------------
    234            345 Unterabteilung B2 U
    345                Abteilung B       U
    456            345 Unterabteilung B1 U
    567            345 Unterabteilung B3 U
    678            345 Unterabteilung B4 I
    123                                  D
```

Listing 3: Auswertung und Ermittlung des DML-Flags

```
SQL> with data (target_id, source_id) as (
  2         values ('A', 'A'),
  3                ('A', null),
  4                (null, 'A'),
  5                ('A', 'B'),
  6                (null, null))
  7  select target_id, source_id,
  8         decode(target_id, source_id, 0, 1) has_changed
  9    from data;

TARGET_ID SOURCE_ID HAS_CHANGED
--------- --------- -----------
A         A                   0
A         null                1
null      A                   1
A         B                   1
null      null                0
```

Listing 4: Einsatz der DECODE-Funktion zum Attributvergleich

alle Attribute unverändert sind. Der Trick ist unglaublich einfach, denn wir verwenden einen full Join. Dieser Join-Typ zeigt uns alle Zeilen beider beteiligter Relationen sowie die Matches zwischen den Relationen.

Listing 2 zeigt das Ergebnis; nun haben wir 6 Zeilen anstatt der 5 Zeilen der Quelltabellen. Von besonderem Interesse sind die Primärschlüsselwerte BEST/SAP_ID der beiden Relationen, da sie NULL sind, falls der Join sie »nicht sieht«. Was bedeutet das aber? Wenn in BESTAND der Primärschlüssel fehlt, ist diese Zeile nur in SAP vorhanden, mithin müssen wir diese Zeile in BESTAND einfügen. Umgekehrt verhält es sich, wenn in SAP der Primärschlüssel fehlt, dann muss in BESTAND die Zeile terminiert werden.

Machen wir uns dieses Wissen zunutze, indem wir ein Flag einführen, das diesen Sachverhalt für uns notiert *(siehe Listing 3)*. Ich werde gleichzeitig von den Inhalten der Tabelle BESTAND Abschied nehmen, da uns diese, bis auf den Primärschlüssel beim Löschen einer Zeile aus dieser Tabelle, nicht mehr interessiert; der aktuelle Stand ist relevant und der steht in SAP.

Das Ergebnis bis hierhin wäre, dass alle Zeilen ein Update erfahren. Da jede Änderung an den Tabellen protokolliert wird, wäre dies nicht sehr sinnvoll, denn im Protokoll stünden jeden Tag 6.000 Einträge, selbst wenn die Werte an sich nicht geändert wurden. Daher müssen wir ein Verfahren etablieren, mit dem wir erkennen können, ob sich Attribute der beteiligten Zeilen geändert haben. Nur wenn tatsächlich eine Änderung vorliegt, wird eine update-Anweisung ausgeführt, ansonsten passiert nichts.

Zeilen vergleichen

Es existieren viele mögliche Wege, Unterschiede zwischen Zeilenattributen zu erkennen. Ein möglicher Weg besteht in der Pflege eines Attributes, das anzeigt, wann eine Zeile zuletzt geändert wurde. Da ein solches Attribut nicht immer vorhanden oder korrekt gepflegt ist, haben sich allgemeinere Verfahren etabliert, wie beispielsweise ein Vergleich über einen Hashcode. Hashcodes haben eigene Probleme, die zu falsch positiven, gleichen Hashcodes führen können, etwa null-Werte in einer Konkatenation von Spaltenwerten, die man zwar lösen kann,

```
SQL> select coalesce(sap_id, best_id) best_id,
  2         sap_parent_id best_parent_id, sap_name best_name,
  3         case when best_id is null then 'I'
  4              when sap_id is null then 'D'
  5              else 'U' end best_dml_action
  6    from bestand
  7    full join sap
  8      on best_id = sap_id
  9   where decode(best_id, sap_id, 0, 1) +
 10         decode(best_parent_id, sap_parent_id, 0, 1) +
 11         decode(best_name, sap_name, 0, 1) > 0;

BEST_ID BEST_PARENT_ID BEST_NAME         BEST_DML_ACTION
------- -------------- ----------------- ---------------
    234            345 Unterabteilung B2 U
    567            345 Unterabteilung B3 U
    678            345 Unterabteilung B4 I
    123                                  D
```

Listing 5: Aussonderung unveränderter Zeilen

```
SQL> select coalesce(sap_id, best_id) best_id,
  2         sap_parent_id best_parent_id, best_name,
  3         case when best_id is null then 'I'
  4              when sap_id is null then 'D'
  5              else 'U' end best_dml_action,
  7         decode(best_parent_id, sap_parent_id, 0, 1) best_umzug_flg
  7    from bestand
  8    full join sap
  9      on best_id = sap_id
 10   where decode(best_id, sap_id, 0, 1) +
 11         decode(best_parent_id, sap_parent_id, 0, 1) +
 12         decode(best_name, sap_name, 0, 1) > 0

BEST_ID BEST_PARENT_ID BEST_NAME         BEST_DML_ACTION BEST_UMZUG_
FLG
------- -------------- ----------------- --------------- -----------
---
    234            345 Unterabteilung B2 U               TRUE
    567            345 Unterabteilung B3 U               FALSE
    678            345 Unterabteilung B4 I               TRUE
    123                                  D               FALSE
```

Listing 6: Endgültige Abfrage der Delta-View

doch führt dies zu einer recht aufwändigen Berechnung der Codes.

Ich möchte daher in *Listing 4* eine alternative Lösung vorschlagen, die vielleicht etwas unkonventionell wirkt, aber funktioniert und insbesondere für überschaubare Attributmengen durchaus praktikabel ist. Die Idee besteht in der Verwendung der decode-Funktion, die in diesem Zusammenhang durch ihren pragmatischen Umgang mit dem null-Wert besticht, denn dieser wird im Vergleich als gleich einem anderen null-Wert angesehen und als ungleich einem not null-Wert, was wir hier brauchen. Die decode-Funktion kann zudem mit den meisten Datentypen der Datenbank umgehen und ist dadurch flexibel einsetzbar.

Filterung auf geänderte Zeilen

Mit diesem Rüstzeug sind wir in der Lage, unsere bestehende Abfrage durch einen Filter zu erweitern, der unveränderte Zeilen aussteuert. Hierzu werden wir alle Attribute mit ihrem Partner aus der anderen Relation in je einer decode-Funktion vergleichen und aufsummieren. Unveränderte Zeilen haben eine Summe von 0, sodass uns nur Zeilen interessieren, deren Summe darüber hinausgeht *(siehe Listing 5)*.

Natürlich wird es mühsam, diese decode-Anweisungen einzeln hinzuschreiben, wenn 100 Attribute verglichen werden müssen. Dies wäre ein idealer Einsatzzweck für ein SQL-Makro, das uns diesen Teil der Abfrage dynamisch aus den Spalten der beteiligten Tabellen berechnet und in die select-Anweisung einfügt, bevor diese geparst wird. Bei unterschiedlichen Spaltenstrukturen oder -namen könnte eine View auf eine der Tabellen für eine Vereinheitlichung sorgen, bevor das SQL-Makro aufgerufen wird.

Besondere Änderungen erkennen

Nun sind wir beinahe fertig, denn mit diesen Mitteln haben wir nur noch die Zeilen als Ergebnis vorliegen, die entweder neu oder nicht mehr vorhanden sind oder deren Werte sich geändert haben. Allerdings könnte es wichtigere und unwichtigere Änderungen geben, die unterschieden werden müssen. Nehmen wir für unser Beispiel an, ein Umzug einer Organisationseinheit innerhalb des Unternehmens sei wichtiger als eine Namensänderung. Wie erkennen wir solche Änderungen? Wir wiederholen einfach den Trick mit der decode-Funktion, nur nutzen wir sie, um ein Flag zu berechnen, das in der select-Klausel der Abfrage angezeigt und so für die weitere Verarbeitung zugänglich gemacht wird. *Listing 6* zeigt die fertige Abfrage, die noch, um ihrem Namen gerecht zu werden, als View in der Datenbank abgelegt werden kann.

Über diese Ergebnismenge wird nun ein trivialer Loop programmiert, der lediglich einfache Methoden zum Logging und zur Speicherung der geänderten Daten ausführt. Fachliche Logik benötigt der Code nicht mehr, denn diese ist bereits in der View enthalten.

Und was ist mit MERGE?

Falls ihre Problemstellung vollständig in SQL umsetzbar wäre und zum Beispiel kein explizites Logging aller Änderungen erfordert, lassen sich die beiden Ansätze, merge-Anweisung und Delta-View, kombinieren. Wie dies funktioniert, zeigt *Listing 7*. Beachten Sie, dass das Gültigkeitsband in der bisherigen Form nicht mehr verwendet werden kann, weil historische Speicherung dort ohne externen Code nicht funktioniert. Alternativ wäre eine

```
SQL> merge into bestand t
  2  using (select coalesce(sap_id, best_id) best_id,
  3                sap_parent_id best_parent_id,
  4                sap_name best_name,
  5                coalesce(best_valid_from, trunc(sysdate)) best_
valid_from,
  6                case when best_id is null then 'I'
  7                     when sap_id is null then 'D'
  8                     else 'U' end best_dml_action
  9           from bestand
 10           full join sap
 11             on best_id = sap_id) s
 12     on (t.best_id = s.best_id)
 13   when matched then update set
 14        t.best_parent_id = s.best_parent_id,
 15        t.best_name = s.best_name
 16        where
 17          decode(t.best_parent_id, s.best_parent_id, 0, 1) +
 18          decode(t.best_name, s.best_name, 0, 1) +
 19          decode(t.best_valid_from, s.best_valid_from, 0, 1) > 0
 20        delete where best_dml_action = 'D'
 21   when not matched then insert (
 22          best_id, best_parent_id, best_name, best_valid_from)
 23        values (
 24          s.best_id, s.best_parent_id, s.best_name, s.best_valid_
from);
4 Zeilen zusammengeführt.
```

Listing 7: MERGE-Anweisung mit Delta-View

historisierende Speicherung nur mit einem Startdatum möglich. Beachten Sie ebenso, dass im Matched-Zweig eine Auswertung des DML-Flags dafür sorgt, dass nicht mehr benötigte Zeilen direkt gelöscht werden.

Eine einfache merge-Anweisung ohne die Delta-View könnte zu löschende Daten nicht erkennen. Der Vorteil, die merge-Anweisung durch diesen Filter zu erweitern, erscheint nicht so offensichtlich, schließlich bleibt es bei einer merge-Anweisung, die über die gesamte Ergebnismenge iterieren muss. Bei vielen Zeilen ist der Unterschied jedoch relevant, denn eine update-Anweisung wird den alten Wert der Zeile in das Redo Log schreiben, unabhängig davon, ob sich die Werte wirklich geändert haben. Da dieser Aufwand überwiegend unnötig ist, ist der Filter empfehlenswert, denn er reduziert die Datenmenge in der Redo-Log-Datei und damit generell die Last auf dem Datenbankserver.

Bewertung

Zurück zum Projekt. Natürlich haben wir diese Strategie als Erfolg gefeiert, denn anstatt 6.000 Zeilen zweimal Zeile für Zeile zu vergleichen (was, nebenbei, noch eine erhebliche Menge Code erfordert), ist eine einfache Abfrage gegen eine View möglich geworden. Dies bietet zudem den Vorteil, dass sie, wenn sie nach der Verarbeitung ein zweites Mal aufgerufen wird, einfach keine Zeilen liefern wird, weil keine Unterschiede vorhanden sind.

Nachdem meine Kollegin eine Nacht über diese Lösung geschlafen hatte, kam sie am nächsten Tag auf mich zu und sagte, dass sie diese Lösung keinesfalls implementieren werde, das sei schlicht zu gefährlich. Nanu? Warum das? „Was“, so argumentierte meine Kollegin, „passiert, wenn SAP an einem Tag keine oder grob falsche Daten liefert?“ Ein cleverer Einwand! Die Delta-View reagiert in diesem Fall so, dass alle Zeilen der Bestandstabelle als unterschiedlich gekennzeichnet würden. Liefe das automatisierte Skript am nächsten Morgen, wäre die Folge, dass allen Mitarbeitern ihre Berechtigungen entzogen würden. Das wären im Fall meines Kunden über 16.000 Mitarbeiter. Verantwortlich wäre meine Kollegin gemacht worden und nicht etwa die eigentliche Quelle des Problems, der unvollständige oder fehlerhafte Import aus SAP.

Wie lässt sich dieses Problem vermeiden? Man könnte automatisierte Plausibilitätsprüfungen einbauen, um abzuschätzen, ob die Zahl der Korrekturen realistisch ist, aber letztlich haben wir uns für einen anderen Weg entschieden: Wir haben dem Kunden im Haus eine Administrationsoberfläche angeboten, in der für den aktuellen Tag eingesehen werden kann, welche Änderungen geplant sind. Dort kann dann die Synchronisierung zwischen den Tabellen angestoßen werden. Im Hinterkopf haben wir die Option, dies in einen halbautomatischen Prozess umzubauen. Die eigentliche Nachricht ist jedoch eine andere: Diese Option hatten wir nur, weil es eine Delta-View gab. Hätte es diese nicht gegeben, wäre die einzige Möglichkeit gewesen, den Code auszuführen und anschließend zu sehen, was passiert ist. Auch in diesem Fall wären allen Mitarbeitern die Berechtigungen entzogen worden, wenn falsche Daten von SAP übermittelt worden wären, doch diesmal ohne die Möglichkeit des Eingreifens. Als Fazit kann man also sagen, dass die Delta-View diesen Prozess nicht nur ungleich einfacher, sondern auch deutlich sicherer gemacht hat – und dies bei signifikant reduzierter Codemenge.

Unabhängig von diesem konkreten Einsatzfall ist ein anderer Punkt wichtig: Die Daten kontrollieren, was zu tun ist, nicht ein diskret programmierter Code. Wenn ich, neben dem Titel der Rubrik, noch ein anderes Zitat nutzen darf, dann vielleicht: „Fragen Sie nicht, was Sie mit den Daten tun können, sondern was die Daten für Sie tun können.“

Jürgen Sieben
j.sieben@condes.de

Privilegienkapselung – eine einfache Methode zur Entlastung des DBAs

Matthias Mann

Oft wiederkehrende Vorgänge, welche administrative Privilegien erfordern, können in Prozeduren gekapselt werden und ausgewählte Benutzer können darauf berechtigt werden. Dies ist eine einfache Maßnahme zur Absicherung der Datenbankinstanz und verringert das Arbeitsvolumen des DBAs.

Ein einfaches Beispiel ist das erzwungene Beenden von Datenbanksessions.

Es eignet sich für solche Umgebungen, in denen die Datenbank nur die Schemas für eine Applikation enthält und es die Funktion eines Applikationsadminstrators gibt.

Als erstes legen wir einen Datenbankuser an und statten diesen mit den benötigten Privilegien aus. Der Account ist ein "Schema-Only" account, man kann sich also nicht direkt mit diesem anmelden *(siehe Listing 1 und 2)*.

Nun definieren wir eine Prozedur, welche die administrative Tätigkeit ausführt, in unserem Beispiel das erzwungene Beenden einer Datenbank-Session *(siehe Listing 3)*.

Optional kann man das Recht zum Ausführen der Prozedur an eine lokale oder globale Datenbankrolle vergeben *(siehe Listing 4)*.

Im ersten Fall kann die Rolle durch den DBA an die lokalen Benutzer vergeben werden, die diese Administrationstätigkeit ausführen sollen *(siehe Listing 5)*.

Im Falle einer global authentifizierten Rolle wird im Unternehmensdirectory eine zugehörige Enterprise-Rolle angelegt. Diese kann durch den Directory-Administrator an die gewünschten Directory-Benutzer vergeben werden kann.

Der Applikations-Administrator kann nun unter kontrollierten Bedingungen Sessions beenden und braucht nicht mehr die Hilfe des DBAs zu beanspruchen *(siehe Listing 6)*.

```
SQL> -- create a service account to  host procedures
SQL> create user service no authentication;
SQL> alter user service account lock;
```

Listing 1

```
SQL> -- grant required privileges
SQL> grant create procedure to service;
SQL> grant alter system to service;
SQL> grant select on gv_$session to service;
```

Listing 2

```
-- create procedure
SQL> create procedure service.kill_session
      ( pn_sid number
       ,pn_serial number
       ,pn_instid number)
     as
       lv_user varchar2(30);
       v_statement varchar2(255);
     begin
       select username into lv_user from sys.gv_\$session
       where sid = pn_sid and serial# = pn_serial and inst_id = pn_instid;
       if lv_user is not null and lv_user not in ('sys','system')
         then
        v_statement := 'alter system kill session  '''||pn_sid||','||pn_serial||',@'||pn_instid||''' immediate';
         dbms_output.put_line(v_statement);
         execute immediate v_statement;
       else
         raise_application_error(-20000,'attempt to kill protected system session has been blocked.');
       end if;
     end;
     /
```

Listing 3

```
SQL> grant execute on service.kill_session to service_role;
```

Listing 4

```
SQL> grant service_role to appl_mgr;
```

Listing 5

```
APPL_MGR> -- get session data which should be ended
APPL_MGR> select sid, serial#, inst_id from gv$session where …
APPL_MGR> exec service.kill_session(<sid>,<serial#>,<inst_id>);
```

Listing 6

BEST OF DOAG ONLINE

Eine Auswahl der besten DOAG News vom 9. Mai bis 17. Juli 2025

Low-Code mit Hochdruck: Wie Oracle APEX die Zukunft der App-Entwicklung verändert

Seit über 20 Jahren arbeitet Simon Hunt, Keynote Speaker der APEX connect 2025, mit Oracle APEX, entwickelt Anwendungen für hochsichere Bereiche und treibt Innovationen voran. Im Interview geht's um KI, Cloud, Sicherheit – und eine Achterbahn.

Multicloud-Strategien mit Oracle: Chancen, Herausforderungen und Trends

Im DOAG.tv-Interview spricht Johannes Michler von Promatis mit Jessica Steger über Multicloud-Strategien, die Rolle von Oracle Cloud Infrastructure und aktuelle Entwicklungen im Cloud-Markt.

Hochverfügbarkeit neu gedacht: Oracle Database 23ai im Praxiseinsatz

Mit Oracle Database 23ai verspricht Oracle mehr Automatisierung, kürzere Ausfallzeiten und KI-gestützte Optimierung. Was steckt wirklich dahinter? Einblicke aus erster Hand.

DOAG Datenbank Kolumne: Oracle-Datenbank-Benutzer-Profile (für Passwörter)

Die Möglichkeit mit Hilfe von Benutzer Profilen das Passwort-Management in der Oracle-Datenbank abzubilden, gibt es schon "ewig". Ich kann nicht mehr sicher sagen, ob die Passwort-Einstellungen schon in Oracle 7 vorhanden waren, oder doch erst mit Oracle 8 gekommen sind.

Emotionale Maschinen: Wie KI unsere Entscheidungen (nicht) beeinflusst

Künstliche Intelligenz wirkt menschlicher denn je. Doch wie reagieren Konsumenten auf sprechende Bots und virtuelle Influencer? Ein Blick hinter die Fassade der KI-Verführung.

KI im Arbeitsalltag: Vom Entwickler zum Dialogpartner

Künstliche Intelligenz verändert die Arbeit in der IT – das erlebt Wolf G. Beckmann täglich. Im Interview mit Thomas Günther erklärt der Experte, warum KI kein Ersatz, sondern eine wertvolle Unterstützung ist – und was Unternehmen sowie Berufseinsteiger beachten sollten.

Wir begrüßen unsere neuen Mitglieder

Natürliche Mitglieder:

- Ken McCarthy
- Kewal Bhansali

Korporative Mitglieder:

- SplashBI, Repräsentantin: Maria Vikatou
- adorsys GmbH & Co.KG, Repräsentantin: Katrin Tschiersch
- LGT Financial Services AG, Repräsentant: Manfred Grasser
- TeamBank AG, Repräsentant: Sascha Wiegandt

Termine

September 09

12.09.2025
Oracle Zero Downtime Migration (ZDM): Das Werkzeug zur Migration in die Oracle Cloud oder auf Exadata
DB WebSession mit Marc Wagner
Online

16. - 17.09.2025
German Low-Code Day 2025
Herstellerneutrale Kongressmesse für programmierfreie Software-Entwicklung für IT-Führungskräfte in Wirtschaft und Verwaltung.
Eine Veranstaltung von Low-Code Association e.V. in Kooperation mit DOAG.
Hannover

24.09.2025
DOAG DBTalk: Zukunft des DBA
DBTalk mit Christian Trieb
Online

Oktober 10

08 - 09.10.2025
Expertenseminar: Praxisworkshop Oracle Database Appliance
Berliner Expertenseminar mit Florian Barth
Berlin

10.10.2025
Migration von RAC auf individueller Hardware zu Data Guard auf ODA
DB WebSession mit Ralf Appelbaum
Online

22.10.2025
DOAG DBTalk: Datenbank-Security-Check-Tools in der Praxis
DBTalk mit Bruno Cirone und Christian Pfundtner
Online

23.10.2025
Regionaltreffen Dresden
DBA ade? Der Weg vom Datenbankspezialisten zum Cloud-Architekten
Dresden

November 11

17. - 18.11.2025
European NetSuite User Days 2025
Die European NetSuite User Days werden in diesem Jahr im Rahmen der DOAG 2025 Konferenz + Ausstellung in Nürnberg stattfinden.
NürnbergConvention Center (NCC Ost)

18. - 19.11.2025
Low-Code Creator 2025
Die Konferenz für Low-Code-Entwicklerinnen und -Entwickler in Zusammenarbeit mit Low-Code Association e.V.
NürnbergConvention Center (NCC Ost)

18.11. - 21.11.2025
DOAG 2025 Konferenz + Ausstellung
Die führende Anwender-Konferenz mit Fokus auf Oracle-Technologien im deutschsprachigen Raum.
NürnbergConvention Center (NCC Ost)

19. - 20.11.2025
KI Navigator 2025
Konferenz zur Anwendung von KI in IT, Wirtschaft und Gesellschaft
NürnbergConvention Center (NCC Ost)

Impressum

Red Stack Magazin wird gemeinsam herausgegeben von DOAG e.V. (Deutschland, Tempelhofer Weg 64, 12347 Berlin, *www.doag.org*), AOUG Austrian Oracle User Group (Österreich, Lassallestraße 7a, 1020 Wien, *www.aoug.at*) und SOUG Swiss Oracle User Group (Schweiz, Dornacherstraße 192, 4053 Basel, *www.soug.ch*).

Red Stack Magazin ist die Community-Publikation für angewandte Informations- und Kommunikationstechnologie (ITK) im Raum Deutschland, Österreich und Schweiz. Es setzt bewusst auf Technologieoffenheit mit Blick auf Anwendung und IT-Innovationen.

Es bildet die Interessensschwerpunkte der Anwenderinnen und Anwender ab – von Cybersicherheit bis Datenschutz, von Datenbank und Development über Data Analytics bis Digitalisierung, von Cloud und Infrastruktur über Künstliche Intelligenz bis Open Source und Soft Skills – vermittelt praktisches Wissen und fördert den Know-how-Transfer und die Netzwerkbildung zwischen den Leserinnen und Lesern.

Die Inhalte des Red Stack Magazin werden von ausschließlich ehrenamtlichen Autorinnen und Autoren eingereicht und von der Redaktion aufbereitet.

Red Stack Magazin wird verlegt von der DOAG Dienstleistungen GmbH, Tempelhofer Weg 64, 12347 Berlin, Deutschland, gesetzlich vertreten durch den Geschäftsführer Fried Saacke, deren Unternehmensgegenstand Vereinsmanagement, Veranstaltungsorganisation und Publishing ist. DOAG e.V. hält 100 Prozent der Stammeinlage der DOAG Dienstleistungen GmbH. DOAG e.V. wird gesetzlich durch den Vorstand vertreten; Vorsitzender: Björn Bröhl

Redaktion:
Sitz: DOAG Dienstleistungen GmbH
(Anschrift s.o.)
ViSdP: Fried Saacke
Redaktionsleitung Red Stack Magazin:
Martin Meyer
Kontakt: redaktion@doag.org

Autorinnen und Autoren dieser Ausgabe
(in alphabetischer Reihenfolge):
Benedikt Backhaus, Meris Bihorac,
Patrik Graf, Christian Harms,
Fabian Heidenstecker, Alexander Hendorf,
Thomas Keßler, Stefan Latuski,
Matthias Mann, Martin Meyer,
Oliver Szymanski, Jürgen Sieben,
Arne Wellnitz, Stefan Winkler

Titel, Gestaltung und Satz:
Diana Tkach
DOAG Dienstleistungen GmbH
(Anschrift s.o.)

Fotonachweis:
Titel: © freepik | *www.freepik.com*
S. 10: © cegoh | *www.pixabay.com*
S. 14: © mikedepue | *www.pixabay.com*
S. 22: © pixabay | *www.pixabay.com*
S. 26: © ungnguyen0905 | *www.pixabay.com*
S. 32: @ Schnapp_schuss | *www.pixabay.com*
S. 38: © Myriams-Fotos | *www.pixabay.com*
S. 44: © Ghinzo | *www.pixabay.com*
S. 50: © sasint | *www.pixabay.com*
S. 54: © BrianPenny | *www.pixabay.com*
S. 61: © SCY |*www.pixabay.com*
S. 66: © andreaschitz | *www.pixabay.com*
S. 69: © freepik | *www.freepik.com*

Anzeigen:
sponsoring@doag.org

Mediadaten und Preise:
www.doag.org/go/mediadaten

Druck:
WIRmachenDRUCK GmbH,
www.wir-machen-druck.de

Inserentenverzeichnis


DIE KONFERENZ FÜR LOW-CODE ENTWICKLER*INNEN
LOW-CODE CREATOR
18. – 19. NOVEMBER 2025
NÜRNBERG CONVENTION CENTER (NCC OST)
LOW-CODE.DOAG.ORG

KI Navigator

DER NAVIGATOR ZUR ANWENDUNG VON KI

→ 19. + 20. NOVEMBER 2025

IN NÜRNBERG

DIE KONFERENZ, DIE ORIENTIERUNG
BEI DER ANWENDUNG VON KI BIETET.

KINAVIGATOR.EU