



Mittelstand-Digital
Zentrum
Fokus Mensch

Digitale Souveränität als Basis für sichere KI-Anwendungen

EIN KURZLEHRBUCH

Gefördert durch:



Bundesministerium
für Wirtschaft
und Energie

Mittelstand-
Digital

aufgrund eines Beschlusses
des Deutschen Bundestages

Inhaltsverzeichnis

Einleitung.....4

BASISWISSEN Transformation der Arbeit: Wie sich die Arbeitswelt verändert.....6

BASISWISSEN Open-Source Large Language Models: Der aktuelle Stand in Wissenschaft und Praxis9

KOMPETENZWISSEN Function-Calling in Sprachmodellen: Souveräne Chancen für den deutschen Mittelstand12

KOMPETENZWISSEN LLM-Auswahl im Mittelstand: Die Tücken automatisch übersetzter Benchmarks14

BASISWISSEN Reasoning Modelle und explainable AI: Was ist das eigentlich?16

KOMPETENZWISSEN Model Context Protocol (MCP) und Open-Source-Sprachmodelle: Die Eröffnung neuer souveräner Wege.....18

BASISWISSEN KI-basierte Agenten-Systeme: Frameworks und Praxis.....21

KOMPETENZWISSEN Agentic RAG: Die nächste Evolutionsstufe für intelligente KI-Systeme.....24

BASISWISSEN EU AI ACT: Potenziale und Hürden.....28

BASISWISSEN Social Engineering: Die unsichtbare Waffe der Cyberkriminellen – und warum KI alles verändert30

BASISWISSEN Phishing erkennen und vermeiden: So schützen Sie sich vor digitalen Ködern32

BASISWISSEN Gefälschte Kundenbewertungen: Warum der richtige Umgang mit ihnen so wichtig ist.....34

BASISWISSEN Kundenbewertungen als Erpressung: Wenn Fake-Bewertungen zum Geschäftsrisiko werden36

Schlusswort.....38

Glossar40

Autorenverzeichnis46

Impressum.....48

Einleitung

Die zunehmende Digitalisierung hat in den letzten Jahren – insbesondere seit der COVID-19-Pandemie – einen Innovationsschub ausgelöst, der weit über die bloße Umstellung auf digitale Kommunikations- und Arbeitsformen hinausgeht. Technologische Entwicklungen wie automatisierte Systeme und intelligente Softwarelösungen haben nicht nur das Privatleben (z. B. Online-Bestellungen, digitale Behördengänge, Telemedizin), sondern auch die Arbeitswelt (z. B. virtuelle Meetings, Remote Work, digitale Beratung) grundlegend verändert.

Ein besonders dynamisches Feld dieser Entwicklung ist die **Generative Künstliche Intelligenz (GenKI)**. Generative KI-Systeme wie unter anderem ChatGPT oder Gemini sind in der Lage, eigenständig neue Inhalte zu erzeugen – etwa Texte, Bilder, Audio oder Softwarecode – und finden zunehmend Anwendung in Unternehmen, etwa beim Verfassen von E-Mails, Marketingtexten oder bei der Ideenfindung. Diese Technologien markieren einen Wendepunkt in der digitalen Transformation: Sie ermöglichen eine nie dagewesene Geschwindigkeit und Effizienz in wenig standardisierbaren Geschäftsprozessen und bergen zugleich disruptives Potenzial für bestehende Arbeitsstrukturen.

Doch mit den Chancen gehen auch Herausforderungen einher. Viele Mitarbeitende wissen nicht, wie diese KI-Systeme funktionieren oder wie es um **Datenschutz** und ethische Aspekte steht – was dazu führt, dass das volle Potenzial oft ungenutzt bleibt. Gleichzeitig nutzen auch Kriminelle **GenKI**, um **Cyberattacken** wie Phishing, **Ransomware** oder **Social Engineering** zu optimieren. Die Angriffe werden dadurch raffinierter und gefährlicher.

Um diese Lücke zwischen Potenzial und Risiko zu schließen, wird seit 2023 das Mittelstand-Digital Zentrum Fokus Mensch als Teil des Mittelstand-Digital Netzwerkes vom Bundesministerium für Wirtschaft und Energie (BMWE) gefördert. Ziel ist es, kleine und mittlere Unternehmen (KMU) praxisnah zu unterstützen – mit Expertenwissen, Demonstrationszentren, Best-Practice-Beispielen und Netzwerken für den Erfahrungsaustausch.

In diesem Kurzlehrbuch wurden dreizehn praxisrelevante Beiträge des Mittelstand-Digital Zentrum Fokus Mensch zusammengestellt, um KMU grundlegendes Wissen und Sensibilisierungsmöglichkeiten zu den Themen **GenKI**, **Digitale Souveränität** und **Cybersicherheit** zu vermitteln. **GenKI** beschreibt Systeme des maschinellen Lernens, die eigenständig neue Inhalte wie Texte, Bilder, Audio oder Softwarecode erzeugen können. Diese Modelle lernen aus großen Datenmengen und generieren kontextgerechte, oft kreative Ergebnisse.

Für KMU ist **GenKI** besonders relevant: Sie kann den Fachkräftemangel abfedern, Routineaufgaben automatisieren, Marketing- und Vertriebsprozesse beschleunigen und interne Ressourcen entlasten. Darüber hinaus ermöglicht sie eine bislang unerreichte Geschwindigkeit bei Ideenfindung, Produktentwicklung und Kundenkommunikation – ohne unmittelbare Investitionen in IT-Infrastruktur. Damit wird **GenKI** zu einem strategischen Werkzeug, um unter steigendem Wettbewerbsdruck handlungsfähig zu bleiben und Innovationen voranzutreiben.

GenKI darf allerdings nicht die **digitale Souveränität** unserer Unternehmen in Deutschland schwächen, beschädigen oder verhindern. Unter **digitaler Souveränität** verstehen wir die Fähigkeit zur selbstbestimmten und sicheren Entwicklung, Bereitstellung und Nutzung digitaler Technologien unter Berücksichtigung der spezifischen Risikokonstellation des jeweiligen Anwendungskontextes. Sie umfasst die Sensibilisierung für Chancen, Risiken, Sicherheit, ethische Aspekte und die Resilienz digitaler Werkzeuge. Das Mittelstand-Digital Zentrum Fokus Mensch bietet hierfür Konzepte und Technologien, die die Integration in Geschäftsprozesse erleichtern – etwa KI-gestützte Prozessoptimierung, den Umgang mit **Fake-Bewertungen** und Social-Engineering-Angriffen sowie Methoden zur Mitarbeiterbindung, Agilität und Kommunikation. Ziel ist es, Unternehmen zukunftsorientiert und widerstandsfähig aufzustellen.

Die Kombination aus Wissen zu **GenKI** und **digitaler Souveränität** kann Unternehmen dazu befähigen, neue Systeme bewusst zu nutzen und Mitarbeitende nicht als Schwachstelle, sondern als aktive Schutzfaktoren gegen **Cyberattacken** zu verstehen. Dafür ist es entscheidend, dass nicht nur IT-Fachkräfte und Führungspersonen die Funktionsweise von KI-Systemen kennen, sondern auch alle Mitarbeitenden verstehen, wie große **Sprachmodelle (Large Language Models, LLMs)** arbeiten und in welchen Situationen sie sinnvoll eingesetzt werden können. Ebenso wichtig ist die Aufklärung über Cyberbedrohungen und Schutzmechanismen.

Wir hoffen, dass diese Zusammenstellung informativer Beiträge Ihnen und Ihren Mitarbeitenden dabei hilft, sich zu den Themen **Künstliche Intelligenz (KI)** und **Cybersicherheit** weiterzubilden – und eine Grundlage für den offenen Austausch innerhalb Ihrer Organisation zu schaffen. Darüber hinaus möchten wir Ihnen die Möglichkeit geben, sich bei anbieterneutralen Orientierungsfragen jederzeit vertrauensvoll an Mittelstand-Digital zu wenden.

Das Mittelstand-Digital Netzwerk bietet mit den *Mittelstand-Digital Zentren* und der Initiative *IT-Sicherheit in der Wirtschaft* umfassende Unterstützung bei der Digitalisierung mit dem Schwerpunkt Künstliche Intelligenz. Kleine und mittlere Unternehmen profitieren von konkreten Praxisbeispielen und passgenauen, anbieterneutralen Angeboten zur Qualifikation und IT-Sicherheit. Das Bundesministerium für Wirtschaft und Energie ermöglicht die kostenfreie Nutzung der Angebote von Mittelstand-Digital. Weitere Informationen finden Sie unter: www.mittelstand-digital.de



Basiswissen

Transformation der Arbeit: Wie sich die Arbeitswelt verändert

Dr. Daryoush Daniel Vaziri

Die Arbeitswelt steht vor einem tiefgreifenden Wandel. **Generative Künstliche Intelligenz (GenKI)**, Automatisierung und **digitale Tools** verändern nicht nur, wie wir arbeiten, sondern auch was wir arbeiten. Tätigkeiten, die früher spezifisches Fachwissen erforderten, werden zunehmend durch **Künstliche Intelligenz (KI)**-gestützte Systeme ergänzt – oder ganz ersetzt. Das betrifft nicht nur industrielle Prozesse, sondern vor allem wissensbasierte Berufe.

Dieser Beitrag beleuchtet, wie sich Berufsbilder ver-

ändern, welche Chancen und Risiken sich daraus ergeben und warum es entscheidend ist, diesen Wandel aktiv mitzugestalten.

Was bedeutet „Transformation der Arbeit“?

Unter der Transformation der Arbeit versteht man die fundamentale Veränderung von Tätigkeiten, Rollen und Anforderungen in der Berufswelt. Diese Veränderungen entstehen durch technologische

Innovationen, neue Geschäftsmodelle und veränderte gesellschaftliche Erwartungen. Insbesondere **GenKI** wirkt wie ein Katalysator: Sie macht komplexe Fähigkeiten skalierbar und reduziert damit Unterschiede im Fähigkeitsniveau zwischen Fachkräften und Nicht-Expert:innen.

Veränderung in Berufsbildern – konkrete Beispiele

■ Programmierung

Low-Code- und No-Code-Plattformen sowie KI-gestützte Programmierassistenten wie GitHub Copilot ermöglichen es mittlerweile auch Nicht-Programmierer:innen, Anwendungen zu entwickeln. Entwickler:innen müssen sich darauf einstellen, stärker konzeptionell zu denken und KI sinnvoll in den Entwicklungsprozess zu integrieren. Reine Umsetzungskompetenz wird weniger gefragt sein.

■ Kundenservice

Chatbots und Voicebots übernehmen Standardanfragen rund um die Uhr. Was bleibt, sind komplexe Anliegen und emotionale Eskalationen – der menschliche Faktor wird dort gebraucht, wo Empathie, Fingerspitzengefühl und individuelle Lösungsfindung erforderlich sind. Mitarbeitende im Kundenservice entwickeln sich so zu „Case Manager:innen“ mit erweitertem Verantwortungsbereich.

■ Kreativberufe

Texte, Bilder, Videos – KI kann heute bereits beeindruckende kreative Ergebnisse liefern. Doch der Mensch bleibt wichtig: als Kurator:in, Ideengeber:in oder Redakteur:in. Kreative Prozesse wandeln sich von der reinen Produktion hin zur Orchestrierung, Bewertung und Ergänzung maschinell erzeugter Inhalte.

■ Controlling und Analyse

Reporting, Datenaufbereitung und Prognosen lassen sich zunehmend automatisieren. KI erkennt Muster schneller als jeder Mensch. Der Fokus im Controlling verlagert sich hin zu strategischer Beratung und zur Übersetzung von Daten in konkrete Handlungsempfehlungen.

Risiken: Jobverlust, Überforderung, Identitätswandel

Die technische Entwicklung bringt auch Schattenseiten mit sich:

- Verlagerung oder Wegfall von Arbeitsplätzen: Aufgaben mit hoher Automatisierbarkeit könnten verschwinden, vor allem dann, wenn Unternehmen den technologischen Wandel primär zur Kostenreduktion einsetzen, anstatt ihn für eine echte Transformation zu nutzen.
- Überforderung: Neue Tools, neue Rollen, neue Anforderungen – viele Beschäftigte fühlen sich bereits heute durch die Geschwindigkeit des Wandels unter Druck gesetzt.
- Identitätswandel: Wer bin ich, wenn meine Kernkompetenz plötzlich von einer Maschine ersetzt werden kann? Diese Frage betrifft nicht nur Einzelne, sondern ganze Berufsbilder.

Chancen: Demokratisierung von Wissen und neue Rollenprofile

Doch die Veränderung eröffnet auch enorme Potenziale:

- Zugang zu Fähigkeiten: KI ermöglicht Menschen ohne spezifische Vorkenntnisse Zugang zu professionellen Tools – etwa zum Programmieren, Schreiben oder Entwerfen.
- Höherwertige Tätigkeiten: Routineaufgaben entfallen. Dadurch kann Zeit für Kreativität, Strategie und Innovation frei werden.
- Neue Berufe und Rollen: **Prompt Engineer**, **KI-Strateg:in**, **Datenethiker:in** – durch den Wandel entstehen auch ganz neue Berufsbilder, die heute noch kaum existieren.

Warum wir die Veränderung mitgestalten müssen

Eines ist klar: Die Arbeitswelt der Zukunft wird nicht mehr so aussehen wie heute. Diese Veränderung wird nicht nebenbei passieren – sie wird fordernd, unbequem und teilweise schmerzhaft sein. Umso wichtiger ist es, sich nicht von ihr überrollen zu lassen.

Unternehmen, Mitarbeitende und Gesellschaft müssen den Wandel aktiv gestalten. Das heißt:

- Frühzeitige Auseinandersetzung mit generativer KI und Automatisierung
- Investitionen in Weiterbildung, Technologiekompetenz und neue Denkweisen
- Diskussionen über ethische, gesellschaftliche und wirtschaftliche Auswirkungen

Fazit: Jetzt souverän handeln

Die Transformation der Arbeit ist nicht mehr abstrakte Zukunftsmusik, sie ist längst Realität. Wer sich nicht mit Potenzialen, Chancen und Risiken beschäftigt, riskiert den Anschluss zu verlieren. Durch zügiges Handeln können die Weichen für eine sinnvolle und menschengerechte Arbeitswelt von morgen gestellt werden.

Gerade **GenKI** bietet viele Möglichkeiten, doch sie fordert uns auch heraus, neu zu denken: über unsere Rollen, unsere Kompetenzen und über den Wert menschlicher Arbeit in einer zunehmend automatisierten Welt.

Die beschriebenen Veränderungen in der Arbeitswelt werden vor allem durch technologische Entwicklungen vorangetrieben. Besonders Sprachmodelle prägen diesen Wandel, da sie den Umgang mit Wissen, Kommunikation und kreativen Prozessen grundlegend verändern. Dabei geht es nicht nur um neue Arbeitsweisen, sondern auch um die Frage, wem diese Technologien gehören und wer über ihre Nutzung bestimmt. Ebenso entscheidend ist jedoch die menschenzentrierte Perspektive: Wie können diese Werkzeuge so gestaltet und eingesetzt werden, dass sie Menschen entlasten, ihre Kreativität fördern und ihnen neue Möglichkeiten zur Entfaltung eröffnen, statt sie lediglich zu ersetzen?

Während **proprietäre KI-Systeme** wie unter anderem ChatGPT oder Gemini den Zugang zu Künstlicher Intelligenz vereinfachen, wächst zugleich eine Bewegung, die auf Offenheit, Transparenz und Unabhängigkeit setzt. **Open-Source LLMs** ermöglichen Unternehmen, eigene Lösungen zu entwickeln und die Kontrolle über Daten und Prozesse zu behalten.

Das folgende Kapitel beleuchtet den aktuellen Stand dieser offenen Sprachmodelle, vergleicht sie mit proprietären Angeboten und zeigt auf, welche Potenziale und Herausforderungen sie für Unternehmen mit sich bringen.



Basiswissen

Open-Source Large Language Models: Der aktuelle Stand in Wissenschaft und Praxis

Dr. Daryoush Daniel Vaziri

Spätestens seitdem Chatbots wie zum Beispiel ChatGPT hat das Thema „**Künstliche Intelligenz**“ die breite Öffentlichkeit erreicht. Viele Unternehmen fragen sich: Wie können wir große Sprachmodelle, sogenannte **Large Language Models (LLMs)**, für unser Geschäft nutzen?

Während große Anbieter wie zum Beispiel OpenAI, Anthropic oder Google ihre Modelle als proprietäre Cloud-Lösungen bereitstellen, gibt es auch eine immer größere Auswahl an **Open-Source LLMs**, die frei verfügbar und anpassbar sind.

Dieser Beitrag beleuchtet, wo **Open-Source LLMs**

heute stehen, wie sie sich von proprietären Angeboten unterscheiden und welche Chancen und Risiken sie für kleine und mittlere Unternehmen (KMU) in Deutschland bieten

Was sind Open-Source LLMs?

Unter **Open-Source LLMs** versteht man große Sprachmodelle, deren Quellcodes und/oder Modellparameter öffentlich zugänglich sind. Das bedeutet: Unternehmen können sie herunterladen, lokal betreiben, anpassen und in eigene Produkte integrieren – oft ohne Lizenzkosten.

Open-Source vs. Proprietär – eine vereinfachte Übersicht		
Kriterium	Open-Source LLM	Proprietäres LLM
Kosten	Keine Lizenzkosten, aber Betriebs- und Wartungskosten	Laufende API- oder Nutzungsgebühren
Datenschutz	Daten können lokal bleiben	Datenverarbeitung meist in der Cloud des Anbieters
Anpassung	Vollständig möglich	Meist eingeschränkt oder nicht möglich
Einstiegshürde	Höher (Technik & Hardware)	Gering (nur API-Anbindung nötig)
Leistungsniveau	Stark steigend, je nach Modell wettbewerbsfähig	Führend bei Spitzenmodellen

Proprietäre Modelle hingegen sind in der Regel nur über eine **API** oder **Cloud-Plattform** nutzbar, und die genauen Funktionsweisen bleiben geheim.

Der aktuelle Stand – Beispiele und Entwicklungen

Vor zwei Jahren dominierten proprietäre Modelle den Markt. Inzwischen hat sich viel getan:

- Meta **LLaMA 3**: Eines der leistungsstärksten **Open-Source-Modelle**, verfügbar in unterschiedlichen Größen, gut dokumentiert, stark in vielen Standardaufgaben.
- Mistral & Mixtral: Modelle aus Europa mit hoher Effizienz, oft günstiger zu betreiben als vergleichbare proprietäre Angebote.
- Falcon, Gemma, OpenHermes: Spezialisiert auf unterschiedliche Einsatzgebiete, von allgemeiner Textverarbeitung bis zu Code-Generierung.
- **Ökosystem-Tools**: Plattformen wie **Hugging Face** bieten fertige Modelle, Werkzeuge und Community-Support.

Kurz gesagt: **Open-Source LLMs** sind nicht mehr nur Forschungsprojekte, sondern in vielen Szenarien produktionsreif.

Vorteile für KMU

Open-Source LLMs bringen einige besondere Stärken mit:

- **Kostenkontrolle**: Keine laufenden **API**-Gebühren pro Anfrage – Betrieb auf eigener Hardware oder günstiger Cloud möglich.
- **Datensouveränität**: Volle Kontrolle darüber, wo und wie Daten verarbeitet werden – ein Vorteil in sensiblen Branchen.
- **Anpassbarkeit**: Modelle lassen sich gezielt auf unternehmensspezifische Daten trainieren oder optimieren.
- **Unabhängigkeit**: Keine Bindung an einen einzelnen Anbieter; geringeres Risiko, dass Preise steigen oder Zugänge eingeschränkt werden.

Nachteile und Herausforderungen

Offene Modelle bringen jedoch auch Hürden mit sich:

- **Technisches Know-how** nötig: Einrichtung, Betrieb und Anpassung erfordern Fachwissen oder einen spezialisierten Dienstleister.
- **Hardwarekosten**: Für große Modelle sind leistungs-

fähige Server oder Cloud-Ressourcen erforderlich.

- **Qualitätsunterschiede**: Nicht jedes Open-Source-Modell ist so leistungsstark oder stabil wie führende proprietäre Modelle.
- **Wartung und Updates**: Sicherheitspatches, Optimierungen und Upgrades liegen in der eigenen Verantwortung.

Warum KMU das Thema jetzt prüfen sollten

Der Markt entwickelt sich rasant. Während proprietäre **Anbieter** ihre Preise und Nutzungsbedingungen diktieren können, wächst das Open-Source-Angebot stetig – auch aus Europa, was datenschutzrechtliche Vorteile bringen kann.

Wer früh testet, kann Erfahrung sammeln, Strategien entwickeln und Wettbewerbsvorteile sichern.

Es ist dabei nicht nötig, sofort alles umzustellen: Pilotprojekte helfen, realistische Einschätzungen zu gewinnen.

Fazit: Open-Source LLMs sind erwachsen geworden

Open-Source-Sprachmodelle sind längst keine Nischenlösung mehr. Sie bieten KMU die Möglichkeit, Kosten zu senken, Daten zu schützen und flexibel zu bleiben, erfordern aber technisches Engagement.

Für viele kleine und mittlere Unternehmen kann eine **Hybrid-Strategie** sinnvoll sein: proprietäre Modelle für schnelle, einfache Anwendungen – **Open-Source LLMs** für datensensible, individuelle Projekte.

Fest steht: Wer sich jetzt mit den Möglichkeiten vertraut macht, kann die Weichen für eine unabhängige und datenschutzfreundlichere KI-Strategie stellen.

Open-Source-Sprachmodelle eröffnen Unternehmen neue Möglichkeiten, **Künstliche Intelligenz** (KI) flexibel, sicher und unabhängig zu nutzen. Doch ihr volles Potenzial entfaltet sich erst dann, wenn sie nicht nur Texte erzeugen, sondern aktiv in bestehende Systeme eingebunden werden.

Mit der sogenannten **Function-Calling**-Funktion beginnt eine neue Entwicklungsstufe: Sprachmodelle werden von reinen Informationssystemen zu

handelnden Akteuren. Sie können Aufgaben ausführen, Daten abrufen oder Prozesse auf kontrollierte und nachvollziehbare Weise anstoßen.

Das nächste Kapitel zeigt, wie **Function-Calling** funktioniert, welche technischen und organisatorischen Voraussetzungen dafür notwendig sind und warum dieser Ansatz insbesondere für den deutschen Mittelstand ein wichtiger Schritt in Richtung souveräner KI-Nutzung ist.



Kompetenzwissen

Function-Calling in Sprachmodellen: Souveräne Chancen für den deutschen Mittelstand

Darius Hennekeuser

Gerade für kleine und mittlere Unternehmen, die auf **digitale Souveränität** besonderen Wert legen, ist der Ansatz interessant. **Function-Calling** ermöglicht, dass Daten lokal verarbeitet werden, Prozesse nachvollziehbar bleiben und dennoch die Effizienzgewinne moderner Künstlicher Intelligenz (KI) genutzt werden. Mit offenen Modellen, die on-premise oder in europäischen **Sovereign Clouds** betrieben werden, lassen sich ungewollte Datenabflüsse verhindern und Vendor Lock-ins vermeiden. Letztes meint vor allem die Abhängigkeit von einem einzelnen KI-Anbieter. Transparenz entsteht durch **Logging** und **Audit Trails**, Austauschbarkeit durch standardisierte Schnittstellen wie **JSON-Schemas** oder **OpenAI-kompatible APIs**. Auf diese Weise behalten Unternehmen die Hoheit über ihre Technologieentscheidungen und stellen **Compliance** mit **Datenschutzgrundverordnung (DSGVO)** und anderen Vorgaben sicher.

Was bedeutet das in der Praxis?

In der Praxis funktioniert das so: Unternehmen definieren zunächst einen klaren Katalog von digitalen Tools, also Funktionen, die das Modell aufrufen darf. Diese sind mit Namen, Beschreibung und Parametern hinterlegt und bilden eine „**Allow List**“, die Missbrauch vorbeugt. Ein **Orchestrator** übernimmt die Aufgabe, die Ausgaben des Modells zu validieren, die Funktionsaufrufe tatsächlich auszuführen und die Ergebnisse wieder zurückzugeben. **Guardrails** stellen sicher, dass sowohl eingehende Nutzereingaben als auch ausgehende Funktionsparameter geprüft und gegen Geschäftsregeln validiert werden. So lassen sich Fehler und Risiken wie **Prompt Injection** abwehren. Gleichzeitig wird durch strukturierte **JSON**-Outputs Stabilität geschaffen: Parameter werden strukturiert verarbeitet (dies ist nur über die das strukturierte **JSON**-Format möglich, nicht

über natürliche Sprache), Validierungen können automatisiert erfolgen, Tests sind einfacher durchzuführen.

Ein Beispiel aus dem Maschinenbau verdeutlicht die Vorteile: Meldet eine Kundin eine Störung, kann der Assistent über ein digitales Tool die Ersatzteilliste der betroffenen Maschine abrufen, Lagerbestände prüfen, einen Servicetermin vorschlagen und schließlich mit Zustimmung der Kundin einen Einsatz planen. Der Prozess läuft schneller, konsistenter und dokumentiert ab – die Reaktionszeit sinkt, die Kundenzufriedenheit steigt.

Damit der Einstieg gelingt, empfiehlt es sich, klein anzufangen. In einem Praxisprojekt reicht es, ein oder zwei eng umrissene Tools zu definieren, die zunächst nur Lesezugriffe erlauben. Erfolgsmaßstäbe sind hier die Genauigkeit und die Zeitersparnis. In einer zweiten Stufe können Schreiboperationen hinzukommen, zunächst noch mit manueller Freigabe. In der Produktionsreife schließlich werden mehrere **digitale Tools** integriert, Rollenrechte eingeführt, **Monitoring** und **Alerts** etabliert und Notfallmechanismen vorgesehen. Auf diese Weise lässt sich Schritt für Schritt Erfahrung sammeln, Risiken kontrollieren und Mehrwert generieren.

Die Anwendungsmöglichkeiten sind vielfältig: Im Kundenservice können Tickets automatisch angelegt oder Garantiefälle geprüft werden, in Einkauf und Logistik lassen sich Lieferstatus und Zolldokumente abfragen, im Vertrieb können Besuchsberichte oder Angebotsentwürfe erstellt werden. Auch Personalabteilungen profitieren von automatisierten Onboarding-Checklisten oder Richtlinienassistenten. In der Produktion wiederum helfen Tool Calls bei Schichtübergaben, Störungsklassifikation oder Ersatzteil-Disposition.

Ob die Modelle dabei on-premise auf eigener Hardware oder in einer europäischen Cloud laufen, hängt von den Anforderungen an Datenhoheit und Flexibilität ab. Wichtig ist, dass eine einheitliche Schnittstelle als „schmale Taille“ zwischen Modellen und digitalen Tools etabliert wird. So lassen sich Modelle austauschen, ohne die Applikationen neu entwickeln zu müssen, und hybride Szenarien bleiben möglich.

Fazit: Function-Calling steigert die Produktivität

Für den Mittelstand bedeutet **Function-Calling** den Schritt von der reinen Interaktion zur echten Prozessautomation mit Sprachmodellen. Mit offenen Technologien, klaren Sicherheits- und **Compliance**-Kon-

trollen und einer schrittweisen Einführung lässt sich Produktivität steigern, ohne die Kontrolle über Daten und Prozesse abzugeben. Wer jetzt erste Pilotprojekte startet, legt den Grundstein für eine souveräne, zukunftsfähige Nutzung von KI im eigenen Unternehmen.

Unternehmen können mit **Function-Calling** nicht nur Informationen abrufen, sondern komplette Prozesse anstoßen und automatisieren. Damit wird deutlich, dass der wirtschaftliche Nutzen von KI maßgeblich davon abhängt, wie präzise und verlässlich die zugrunde liegenden Modelle arbeiten.

Gerade im Mittelstand stellt sich daher zunehmend die Frage, welches Sprachmodell für den eigenen Einsatz am besten geeignet ist. Die Auswahl ist groß und vermeintlich objektive Vergleichstests sollen Orientierung bieten. Doch viele dieser **Benchmarks** sind auf den englischsprachigen Raum ausgerichtet und lassen sich nicht ohne Weiteres auf deutsche Anwendungsfälle übertragen.

Das folgende Kapitel widmet sich genau dieser Herausforderung. Es zeigt, warum **automatisch über-setzte Benchmarks** problematisch sein können, welche Fallstricke bei der Modellauswahl lauern und wie Unternehmen zu fundierten Entscheidungen für den Einsatz von Large Language Models im deutschen Mittelstand gelangen können.



Kompetenzwissen

LLM-Auswahl im Mittelstand: Die Tücken automatisch übersetzter Benchmarks

Darius Hennekeuser

Auf den ersten Blick scheinen automatische Übersetzungstools wie zum Beispiel DeepL, Google Translate oder GPT-4 gute Arbeit zu leisten. Die übersetzten Sätze sind flüssig, die Aufgaben verständlich, doch im Detail ergeben sich gravierende Probleme.

So verschwinden beispielsweise im Original bewusst eingebaute Grammatikfehler, die **Large Language Models** erkennen und korrigieren sollen, nach der Übersetzung spurlos. Damit ist die Aufgabe nutzlos geworden, weil das Modell gar nicht mehr beweisen kann, ob es Fehler identifiziert. Ein anderes Beispiel betrifft Programmieraufgaben: Übersetzungssysteme verändern Schlüsselwörter in Python-Code, wodurch der Code schlicht nicht mehr lauffähig ist.

Für ein Modell, das eigentlich seine Programmierfähigkeiten demonstrieren soll, ist die Aufgabe damit nicht lösbar. Hinzu kommen subtile Bedeutungsveränderungen. Einzelne Begriffe oder Redewendungen verlieren bei der Übersetzung ihren Sinn, sodass ein Test plötzlich unklar oder sogar unsinnig wird. Kurz gesagt: Was auf Englisch ein sauberer **Benchmark** ist, wird im Deutschen schnell zur Stolperfalle.

Für KMU ist das mehr als eine akademische Randnotiz. Denn wenn ein **Benchmark** fehlerhaft übersetzt ist, kann leicht das falsche Modell als „besser“ erscheinen. Wer sich auf solche Testergebnisse verlässt, riskiert teure Fehlentscheidungen. Beispielsweise indem ein Modell gewählt wird, das in engli-

schen Tests stark abschneidet, aber in der deutschen Praxis schwach performt. Besonders heikel wird es in **regulierten Branchen** wie Gesundheit, Recht oder Verwaltung. Dort können falsche Annahmen über die Leistungsfähigkeit eines Modells nicht nur zu finanziellen Verlusten, sondern auch zu **Compliance-Verstößen** führen.

Was also tun?

Internationale **Benchmarks** können durchaus als Orientierung dienen, sollten aber niemals ungeprüft übernommen werden. Automatische Übersetzungen sind ein nützlicher Startpunkt, doch sie müssen durch menschliche Revision ergänzt werden, um die ursprüngliche Funktionalität der Aufgaben zu bewahren. Vor allem sollten Unternehmen ihre Modelle nicht nur anhand fremder **Benchmarks** bewerten, sondern auch mit eigenen Beispielen testen, etwa mit typischen Kundenanfragen, Produkttexten oder internen Dokumenten. So zeigt sich, wie ein Modell im konkreten Arbeitsalltag funktioniert.

Ebenso wichtig ist eine realistische Erwartung an den Prozess. Eine gründliche Evaluierung kostet Zeit, spart aber langfristig Ressourcen, weil Fehlentscheidungen und kostspielige Nachbesserungen vermieden werden. Und nicht zuletzt sollte der Blick auf regulatorische Entwicklungen nicht fehlen: Mit dem **EU AI Act** (wird in Kapitel 10 erläutert) werden in den kommenden Jahren strengere Anforderungen an Nachvollziehbarkeit und Qualität von **Künstliche Intelligenz (KI)**-Systemen eingeführt. Eine saubere, **sprachspezifische Evaluierung** ist damit nicht nur eine Frage der Best Practice, sondern bald eine gesetzliche Pflicht.

Die zentrale Botschaft lautet also: Wer im Mittelstand auf **LLMs** setzt, darf sich nicht mit der bequemen Lösung zufriedengeben. **Automatisch übersetzte Benchmarks** wirken auf den ersten Blick hilfreich, bergen aber erhebliche Risiken. Erst durch eine Kombination aus maschineller Übersetzung, menschlicher Nachbearbeitung und eigenen Praxistests lassen sich fundierte Entscheidungen treffen und damit die richtigen Weichen für einen erfolgreichen und sicheren KI-Einsatz im deutschen Mittelstand stellen.

Fazit: Die Qual der Wahl

Die Auswahl eines geeigneten Sprachmodells ist für mittelständische Unternehmen ein entscheidender Schritt auf dem Weg zu einem verantwortungsvollen und effizienten KI-Einsatz. Doch mit der technischen Entscheidung allein ist es nicht getan. Sobald Modelle produktiv eingesetzt werden, stellt sich eine weitere zentrale Frage: Wie lassen sich ihre Ergebnisse nachvollziehen und bewerten?

Gerade in geschäftskritischen Anwendungen ist Vertrauen in die Funktionsweise von KI-Systemen unerlässlich. Anwenderinnen und Anwender müssen verstehen können, wie ein Modell zu einem bestimmten Ergebnis gelangt ist, um dessen Zuverlässigkeit einschätzen zu können. Hier setzt das Konzept der **Explainable AI (XAI)** an, das Transparenz und Nachvollziehbarkeit in den Mittelpunkt stellt.

Das folgende Kapitel führt in die Methode des **Reasonings** ein – einen Ansatz, der darauf abzielt, die gedanklichen Zwischenschritte von Sprachmodellen sichtbar zu machen. Es zeigt, wie **Reasoning** die Erklärbarkeit von KI-Systemen fördern kann, wo die methodischen Grenzen liegen und welche Rolle dieser Ansatz für eine vertrauenswürdige Nutzung generativer KI spielt.

Explainable AI

Praxis: Prompt-Engineering mit GPT-4
Lernziele: GPT-4 für verschiedene Aufgaben, das System verstanden, die Ergebnisse validiert werden können.
Dieses Dokument ist ein Beispiel für die Erklärbarkeit von KI-Systemen.



Basiswissen

Reasoning Modelle und explainable AI: Was ist das eigentlich?

Benedikt Haag

Reasoning als Methode für Explainable AI (erklärbare Künstliche Intelligenz (KI))

Mit der zunehmenden Verbreitung generativer KI-Systeme gewinnt die Frage nach ihrer Erklärbarkeit an Bedeutung. Anwenderinnen und Anwender stehen häufig vor der Herausforderung, Entscheidungen oder Ergebnisse von KI-Systemen zwar nutzen zu können, deren Entstehung jedoch kaum nachvollziehen zu können. **Explainable AI (XAI)** bezeichnet den Forschungs- und Anwendungsbe- reich, der Methoden entwickelt, um Entscheidungen, Vorhersagen oder generierte Inhalte von KI-Systemen transparent, nachvollziehbar und überprüfbar zu machen. Ziel ist es, Vertrauen zu schaffen, Fehlentscheidungen zu vermeiden und die Zusammenarbeit zwischen Mensch und Maschine zu verbessern.

Ein Ansatz innerhalb dieses Feldes ist das sogenannte **Reasoning**, also die explizite Darstellung der gedanklichen oder rechnerischen Zwischenschritte, die ein Modell auf dem Weg zu einem Ergebnis durchläuft. Unter **Reasoning** versteht man den Prozess, bei dem ein KI-System seine internen Schlussfolgerungen sichtbar macht, etwa

in Form logisch strukturierter Argumentationsketten oder textuell formulierter Teilschritte (sogenannter Chains-of-Thought).

Reasoning als Mittel zur Nachvollziehbarkeit

Aus Anwendersicht liegt der Wert von **Reasoning** in der Möglichkeit, den zuvor undurchsichtigen inneren Ablauf eines Modells zumindest teilweise nachzu- vollziehen. Wenn ein KI-System nicht nur ein Ergebnis präsentiert, sondern auch erläutert, warum es zu diesem Schluss gelangt ist, entsteht eine neue Qualität der Interaktion. Die Nutzerin oder der Nutzer kann die einzelnen Schritte kritisch prüfen, Rück- schlüsse auf mögliche Fehlannahmen ziehen und die Entscheidung in ihren Kontext einordnen. Dies stärkt das Vertrauen in die Systemausgabe und fördert eine reflektierte Nutzung generativer KI.

In praktischen Anwendungsszenarien, beispiels- weise in der technischen Diagnose, der Entschei- dungsunterstützung oder der automatisierten Text- generierung, kann **Reasoning** somit als Werkzeug dienen, um die Nachvollziehbarkeit komplexer KI-Prozesse zu erhöhen. Ein System, das seine Gedankengänge offenlegt, ermöglicht es Anwen-

dern, die Plausibilität einer Antwort zu beurteilen, eigene Hypothesen zu vergleichen oder fehlerhafte Schlussfolgerungen frühzeitig zu erkennen.

Leistungssteigerung und Erklärfunktion

Empirische Befunde zeigen, dass Modelle, die Zwi- schenschritte explizit generieren, bei Aufgaben mit mehreren logischen oder rechnerischen Teilschrit- ten, z.B. in der Mathematik oder bei Argumentati- onsaufgaben, häufig höhere Genauigkeit erzielen. Der sichtbare Denkprozess wirkt hier zugleich als Leistungsmerkmal und als Erklärmechanismus.

Allerdings ist dieser Zusammenhang nicht uni- versell. Noch ist unklar, inwieweit die generierten Gedankenketten tatsächlich den internen Verarbei- tungsprozess abbilden oder lediglich ein nützliches Artefakt darstellen, das richtige Antworten wahr- scheinlicher macht. Sprachmodelle auf Basis von **Transformern** verarbeiten Informationen parallel und verteilt, nicht sequentiell. Die sichtbare **Chain- of-Thought** ist daher meist eine vereinfachte, nar- rative Projektion – eine lineare Darstellung, die nur einen Teil der komplexen internen Rechengänge repräsentiert.

In der Forschung wird deshalb zunehmend zwi- schen zwei Perspektiven unterschieden:

- 1. **Reasoning** als Mittel zur Leistungssteigerung des Modells, und
- 2. **Reasoning** als Instrument, um Einblick in die inneren Mechanismen des Modells zu gewinnen.

Während der performative Nutzen empirisch gut belegt ist, bleibt der Nutzen als Methode für XAI ein offenes Forschungsfeld.

Herausforderungen und Grenzen

Für Anwenderinnen und Anwender ergibt sich dar- aus eine doppelte Herausforderung: Einerseits kann **Reasoning** Transparenz erzeugen, ande- rerseits darf die sichtbare Begründung nicht unkriti- sch als tatsächliche „Denkspur“ der KI verstanden werden. Wenn die Zwischenschritte zwar plausibel wirken, aber keinen realen Bezug zu den internen Modellprozessen haben, entsteht eine Schein- Erklärbarkeit, die Vertrauen zwar erhöht, aber faktisch unbegründet ist.

Um **Reasoning** verlässlich als Bestandteil erklär- barer KI zu nutzen, bedarf es daher ergänzender Maßnahmen: etwa unabhängiger **Verifikationsme- chanismen**, **Unsicherheitsindikatoren**, **Evaluati-**

onsmetriken zur Qualität von Erklärungen und kon- trollierter **Prompting-Strategien**, die die Stabilität der generierten Begründungen erhöhen. Nur durch die Kombination solcher Verfahren kann sicherge- stellt werden, dass Erklärungen nicht nur plausibel, sondern auch prüfbar sind.

Fazit: Reasoning als Baustein erklärbarer KI

Reasoning leistet einen wesentlichen Beitrag zur Erklärbarkeit generativer KI-Systeme, ersetzt diese aber nicht. Die Offenlegung von Zwischenschritten kann Transparenz und Vertrauen fördern, sofern sie methodisch fundiert, validierbar und anwenderori- entiert gestaltet ist.

Für die Praxis bedeutet dies: **Reasoning** sollte nicht als rhetorisches Zusatzmerkmal verstanden wer- den, sondern als Teil eines systematischen Ansatz- zes, der Verständlichkeit, Nachvollziehbarkeit und Verantwortlichkeit in der KI-Nutzung vereint.

Langfristig kann aus der Verbindung von **Reaso- ning** mit ergänzenden XAI-Methoden ein Rahmen entstehen, der es ermöglicht, die Leistungsfähigkeit generativer Modelle zu bewahren und ihre Ergeb- nisse zugleich verlässlich erklärbar zu machen. **Reasoning** markiert damit keinen Endpunkt, son- dern einen entscheidenden Zwischenschritt auf dem Weg zu wirklich erklärbarer, vertrauenswürdiger KI.

Reasoning zeigt, wie generative KI-Systeme nach- vollziehbarer und transparenter gestaltet werden können. Doch neben der Erklärbarkeit rückt zuneh- mend auch die Frage in den Fokus, wie Sprachmo- delle technisch sicher und effizient in bestehende Unternehmensprozesse integriert werden können. Transparenz allein genügt nicht, wenn die Systeme nicht kontrolliert, erweiterbar und datensouverän betrieben werden können.

Hier setzt die Entwicklung standardisierter Schnitt- stellen an. Sie soll ermöglichen, dass Modelle auf strukturierte und geprüfte Weise mit internen Daten- quellen, digitalen Tools oder Anwendungen intera- gieren. Ein besonders vielversprechender Ansatz ist das **Model Context Protocol (MCP)**, das eine gemeinsame Sprache zwischen Sprachmodellen und Unternehmenssystemen schafft.

Das folgende Kapitel beleuchtet, wie **MCP** funk- tioniert, welche Vorteile dieser Ansatz für die **digitale Souveränität** des Mittelstands bietet und wie er zusammen mit **Open-Source-Modellen** den Weg zu einer sicheren und unabhängigen Nutzung von KI ebnen kann.



Kompetenzwissen

Model Context Protocol (MCP) und Open-Source-Sprachmodelle: Die Eröffnung neuer souveräner Wege

Darius Hennekeuser

Warum Schnittstellenkompetenz der Schlüssel zur Souveränität wird

Mit dem Aufkommen generativer KI-Systeme erleben viele Unternehmen, wie leistungsfähig Sprachmodelle in der täglichen Arbeit sein können, vom automatisierten Bericht bis zur Kundenkommunikation. Doch der nächste Schritt besteht nicht mehr nur darin, KI zu nutzen, sondern sie sicher, nachvollziehbar und kontrolliert in eigene Systeme zu integrieren.

Hier setzt das **Model Context Protocol (MCP)** an. **MCP** ist ein offener Standard, der definiert, wie Sprachmodelle mit anderen Anwendungen interagieren dürfen. Anstatt Modelle blindlings auf Daten zugreifen zu lassen, schafft **MCP** einen strukturierten Kommunikationskanal zwischen Mensch, Maschine und Organisation. Für kleine und mittlere Unternehmen (KMU) eröffnet sich damit die Chance, **generative KI** unter eigener Regie zu betreiben und **digital**

digitale Souveränität praktisch umzusetzen.

Funktionsweise: Was MCP leistet

MCP ist im Kern ein Protokoll zur sicheren Integration von Sprachmodellen in Anwendungen. Es stellt standardisierte Nachrichtenformate und **Adapter** bereit, die genau festlegen, welche Aktionen ein Modell ausführen darf.

Hierzu zählen unter anderem:

- eine Datenbankabfrage,
- der Zugriff auf interne Dokumente,
- das Auslösen eines Workflows,
- oder das Erstellen eines Berichts im **ERP**-System.

Die Kommunikation läuft dabei in klar definierten Schritten:

1. Das Sprachmodell formuliert eine Anfrage im **MCP**-Format.
2. Der **Adapter** prüft Berechtigungen und Kontext.
3. Das Zielsystem führt die Anfrage aus und liefert strukturierte Daten zurück.
4. Das Modell interpretiert diese Daten und erstellt eine natürliche Sprachantwort.

So entsteht eine sichere Vermittlungsschicht: Das Modell erhält nur Zugriff auf freigegebene Informationen, und alle Aktionen sind nachvollziehbar protokolliert.

Sicherheitsarchitektur und Governance

Gerade im Mittelstand, wo IT-Ressourcen begrenzt sind, ist Transparenz entscheidend. **MCP** ermöglicht die Umsetzung wichtiger **Governance**-Prinzipien:

- Rollen- und Rechtemanagement: Über definierte **Adapter** lässt sich exakt steuern, welche Daten oder Funktionen zugänglich sind.
- **Auditierbarkeit**: Alle Anfragen werden protokolliert – wichtig für Nachvollziehbarkeit und **Compliance**.
- Datenlokalität: Da die Modelle lokal oder auf europäischen Servern betrieben werden können, bleibt die Datenhoheit beim Unternehmen.
- Trennung von Anwendung und Modell: Sollte ein Modell gewechselt werden (z. B. von Mistral auf **LLaMA 3**), bleiben alle Schnittstellen stabil – ein zentraler Faktor für Investitionssicherheit.

Diese Architektur fördert eine „**Zero-Trust**“-Philosophie: Zugriff wird nicht grundsätzlich gewährt, sondern bewusst und kontextbezogen freigegeben.

Verbindung zu Open-Source und digitaler Souveränität

Open-Source-Sprachmodelle sind die natürliche Ergänzung zu **MCP**. Während **MCP** den sicheren Rahmen vorgibt, liefern **Open-Source-Modelle** die Freiheit, diesen Rahmen nach eigenen Bedürfnissen zu gestalten.

Praxisbeispiele für KMU

1. Maschinenbau:

Ein Hersteller nutzt **MCP**, um sein internes **CAD**-Archiv mit einem Open-Source-Modell zu verbinden. Ingenieure können in natürlicher Sprache Fragen stellen („Welche Gehäusevarianten wurden 2022 in Edelstahl gefertigt?“). Das Modell greift über **MCP** strukturiert auf Metadaten und Dokumentationen zu, ohne sensible Konstruktionsdaten zu exportieren.

2. Handwerk:

Ein Sanitärbetrieb integriert ein Sprachmodell lokal, um Rechnungen, Lieferscheine und Angebote zu verarbeiten. Über **MCP** erhält das Modell Zugriff auf das Buchhaltungssystem, klassifiziert Dokumente automatisch und erstellt Zahlungserinnerungen – alles innerhalb der eigenen IT-Umgebung.

3. Dienstleistungsbranche:

Ein Beratungsunternehmen nutzt **MCP**, um Kundenkommunikation im **CRM**-System zu analysieren. Das Modell erkennt Anliegen, leitet sie an passende Teams weiter und erstellt Vorschläge für personalisierte Antworten. Die Daten bleiben im Unternehmen – ein klarer Vorteil gegenüber Cloud-Chatbots.

4. Bildung und Qualifizierung:

MCP kann auch als Grundlage für interne Lernassistenten dienen. Mitarbeitende erhalten über eine sichere Schnittstelle Zugriff auf Schulungsunterlagen, Prozessbeschreibungen oder Sicherheitsrichtlinien. Fragen wie „Wie melde ich eine Datenschutzverletzung?“ oder „Welche Schritte brauche ich für eine Maschinenwartung?“ werden direkt transparent und nachvollziehbar beantwortet.

Für den Mittelstand bedeutet das:

- Betrieb in eigener Infrastruktur – keine sensiblen Daten in externen Clouds.
- Unabhängigkeit von Lizenzmodellen und Preispolitik großer Anbieter.
- Anpassbarkeit: **Open-Source-Modelle** können gezielt mit unternehmenseigenen Daten weitertrainiert oder feinjustiert werden.
- Transparenz: Der Quellcode und die Trainingsdatenstrukturen sind offen einsehbar – eine wichtige Voraussetzung für Vertrauen und **Compliance**.

Damit entsteht eine technologische Selbstbestimmung, die im Einklang mit europäischen Werten und Datenschutzvorgaben (**DSGVO**, **KI-Verordnung**) steht.

Organisatorische Voraussetzungen

Technologie allein genügt nicht. Damit **MCP** im Unternehmen Wirkung entfalten kann, braucht es:

- Verständnis im Management, dass Standardisierung Investitionen schützt und Abhängigkeiten verringert.
- IT-Kompetenz, um **Adapter** und Berechtigungen zu konfigurieren.
- Einbindung der Mitarbeitenden, damit diese die KI als Unterstützung und nicht als Risiko wahrnehmen.
- Schulung und Sensibilisierung: Nur wenn Nutzer wissen, wie **MCP** und **Sprachmodelle** funktionieren, kann Vertrauen entstehen.

MCP kann hier ein verbindendes Element zwischen IT, Fachbereichen und Geschäftsführung sein – ein Werkzeug, um digitale Selbstbestimmung organisatorisch zu verankern.

MCP im Kontext von Cybersicherheit

Ein oft unterschätzter Aspekt ist die Sicherheitswirkung von **MCP**.

Da Modelle nur über definierte Schnittstellen kommunizieren, wird das Risiko von Datenlecks, ungewollten **API**-Aufrufen oder sogenannter **Prompt Injection** stark reduziert. Unternehmen können so die Kontrolle über Datenflüsse behalten – ein zentraler

Factor angesichts zunehmender KI-gestützter **Cyberattacken**.

Zudem erlaubt **MCP** eine saubere Trennung zwischen vertraulichen Daten und externen Modellen. Wenn KI-Anwendungen mit sensiblen Dokumenten arbeiten, geschieht dies in einer isolierten Umgebung, was einen entscheidenden Vorteil gegenüber der direkten Nutzung öffentlicher Cloud-Dienste darstellt.

Perspektive: MCP als Baustein einer europäischen KI-Infrastruktur

Langfristig könnte **MCP** zum Rückgrat einer europäischen KI-Architektur werden. Wenn öffentliche Verwaltungen, Forschungseinrichtungen und Unternehmen denselben Standard verwenden, lassen sich Anwendungen interoperabel gestalten.

Denkbar sind:

- Branchenübergreifende Wissensnetzwerke, in denen Modelle sicher Daten austauschen.
- Standardisierte Audit-Schnittstellen, die KI-Einsatz dokumentieren.
- Kompatible Schulungssysteme, die Wissen über **MCP**-fähige Tools vermitteln.

Damit wird **MCP** zum Schlüssel für eine vertrauenswürdige, europäische KI-Ökonomie, die Innovation ermöglicht, ohne **Datenschutz** oder Unabhängigkeit zu gefährden.

Fazit: MCP kann einen wichtigen Beitrag zur souveränen KI-Integration leisten

MCP ist weit mehr als ein technischer Standard. Es ist eine Strategie zur Rückgewinnung digitaler Gestaltungsmacht. In Verbindung mit **Open-Source-Modellen** erlaubt es mittelständischen Unternehmen, **Sprachmodelle** unabhängig von den Plattformen globaler **Anbieter** sicher, nachvollziehbar und wirtschaftlich einzusetzen.

Es stärkt die **digitale Souveränität**, reduziert Risiken, schützt Daten und schafft zugleich eine Grundlage für Innovation. Wer heute beginnt, **MCP**-fähige Schnittstellen zu nutzen, investiert nicht nur in Technologie, sondern in **digitale Resilienz**, Wettbewerbsfähigkeit und Selbstbestimmung.

Damit wird **MCP** zum zentralen Baustein einer neuen, souveränen KI-Ära im deutschen Mittelstand.

**Basiswissen**

KI-basierte Agenten-Systeme: Frameworks und Praxis

David Golchinfar

Seit den Durchbrüchen großer Sprachmodelle (**LLMs**) in den letzten Jahren hat sich das Feld der **Künstlichen Intelligenz (KI)** rapide weiterentwickelt. Doch während Systeme wie GPT oder **LLaMA** ursprünglich vor allem für Textgenerierung oder Dialog genutzt wurden, zeichnet sich heute eine neue Richtung ab: agentische Systeme.

Diese verbinden die Fähigkeiten von **LLMs** mit Werkzeugnutzung, Gedächtnis und Entscheidungslogik, sodass KI nicht nur reagiert, sondern auch eigenständig plant, handelt und Ziele verfolgt. Agentische Systeme gelten deshalb als nächster logischer Schritt in der KI-Evolution. Doch wie baut man solche Systeme konkret auf und welche **Frame-**

works stehen zur Verfügung?

Open-Source-Frameworks für KI-Agenten

Für Entwickler:innen, die eigene Agenten entwerfen wollen, gibt es inzwischen eine ganze Reihe an **Open-Source-Frameworks**, die unterschiedliche Ansätze verfolgen.

Im momentanen Entwicklungsstand in Forschung und Praxis ist **LangChain** wohl das bekannteste **Open-Source-Framework**. Es bietet die Möglichkeit, Sprachmodelle mit digitalen Tools zu kombinieren, externe Datenquellen anzubinden und ein

Gedächtnis zu implementieren. Besonders verbreitet ist das sogenannte **ReAct-Pattern**, bei dem der Agent erst überlegt („thought“), dann eine Aktion ausführt („action“) und schließlich das Ergebnis reflektiert. Dadurch lassen sich flexible, nachvollziehbare Workflows abbilden, die ideal für **Chatbots** oder automatisierte Rechercheaufgaben sind.

Ein anderer Ansatz kommt von Microsofts **AutoGen**. Dieses **Framework** wurde speziell entwickelt, um Multi-Agent-Systeme aufzubauen. Hier können mehrere Agenten miteinander interagieren, ihre Rollen klar definieren und gemeinsam eine Aufgabe lösen. **AutoGen** unterstützt dabei sowohl rein autonome Agenten-Teams als auch **hybride Setups**, bei denen Menschen eingreifen können. Besonders stark ist es in Szenarien wie Softwareentwicklung oder Data Science, wo komplexe Aufgaben in Teilprobleme zerlegt werden müssen.

CrewAI verfolgt ein ähnliches Prinzip, allerdings mit einem klareren Fokus auf kollaborative Rollenverteilungen. Ein **CrewAI**-Setup besteht aus verschiedenen Spezialagenten – beispielsweise einem Researcher, einem Planner und einem Executor –, die zusammenarbeiten, als wären sie ein menschliches Team. Dieses Design eignet sich vor allem für Business-Anwendungen wie automatisiertes Projektmanagement oder Content-Produktion, bei denen unterschiedliche Arbeitsschritte orchestriert werden müssen.

Für Entwickler:innen, die schnelle Prototypen bauen wollen, bietet **Smolagents** von **Hugging Face** eine besonders schlanke Lösung. Das **Framework** ist bewusst minimalistisch gehalten und erlaubt es, innerhalb weniger Zeilen lauffähige Agenten zu erstellen. Damit ist es ideal geeignet, um erste Ideen auszuprobieren oder in bestehende Projekte integriert zu werden.

Ein weiterer wichtiger Baustein ist **LlamaIndex**, das darauf spezialisiert ist, Sprachmodelle mit Wissensquellen zu verbinden. Ein **LLM** kann ohne zusätzliche Daten zwar viel, bleibt aber von seinem Trainingsstand begrenzt. **LlamaIndex** ermöglicht es Agenten, auf Unternehmensdokumente, Datenbanken oder **APIs** zuzugreifen. Damit lassen sich leistungsfähige Wissensagenten bauen, die in Organisationen als interne Recherche- oder Analyse-Assistenten eingesetzt werden.

Besonders spannend ist auch **PydanticAI**, das mit strengen Typprüfungen und Validierungen arbeitet. Während viele **Frameworks** den Agenten weitgehend frei operieren lassen, setzt **PydanticAI** klare Strukturen, was die Ausgabenformate und Daten-

qualität angeht. Das macht es vor allem für produktionsreife Anwendungen interessant, in denen verlässliche und überprüfbare Ergebnisse entscheidend sind.

Industrielle Systeme und Praxisbeispiele

Neben den offenen **Frameworks** gibt es inzwischen auch konkrete agentische Systeme, die von Unternehmen entwickelt und teilweise schon produktiv eingesetzt werden.

Ein prominentes Beispiel ist Devin von Cognition AI. Devin wird als „erster KI-Softwareentwickler“ vermarktet. Der Agent kann ohne menschliche Kontrolle eigenständig Tickets aus einem Projektmanagement-System aufnehmen, Code schreiben, Tests durchführen und sogar Deployments anstoßen. Zwar befindet sich Devin noch in der Entwicklung, aber er zeigt, wohin die Reise gehen könnte: KI als vollwertiges Mitglied eines Entwicklerteams.

Ein weiteres interessantes Projekt ist **ChatDev**, das als virtuelle Softwarefirma aufgebaut ist. Hier übernehmen unterschiedliche Agenten verschiedene Rollen wie CEO, CTO, Entwickler oder Tester. Gemeinsam arbeiten sie an der Erstellung von Programmen. Das Projekt ist eine Art Rollenspiel für Agenten, das zeigt, wie kollaborative Multi-Agent-Systeme in Zukunft funktionieren könnten.

Auch in der Industrie finden agentische Systeme bereits Anwendung. Im folgenden nennen wir ein paar Beispiele. **Siemens** setzt sie im Bereich der vorausschauenden Wartung ein, wo Agenten Daten von Sensoren auswerten, Anomalien erkennen und Wartungspläne vorschlagen. **JPMorgan** experimentiert mit Agenten für Finanzanalysen, die eigenständig Marktberichte durchforsten und Investmententscheidungen vorbereiten. Und auch im Kundenservice sind agentische Systeme auf dem Vormarsch, die auf Basis von **LangChain** orchestrierte Multi-Agent-Workflows für **CRM**-Aufgaben ermöglichen.

Forschung und neue Ansätze

Während in der Praxis erste Systeme entstehen, arbeitet die Forschung daran, die Grundlagen agentischer Systeme systematisch zu beschreiben.

Ein Beispiel ist **SwarmAgentic**, ein 2025 vorgestelltes **Framework**, das die Entwicklung von Agenten mit Schwarmintelligenz automatisiert. Hier sollen Agenten nicht nur einzeln programmiert, sondern durch kollektives **Verhalten emergent** gestaltet

Tipps für Entwickler:innen

Wer selbst mit agentischen Systemen experimentieren möchte, sollte zunächst klein anfangen. Ein einzelner Agent mit **LangChain** oder **Smolagents** ist ein guter Einstieg, da er schon grundlegende Fähigkeiten wie Planung und Tool-Nutzung demonstriert.

Im nächsten Schritt lohnt es sich, auf Multi-Agent-Frameworks wie AutoGen oder **CrewAI** umzusteigen, sobald die Aufgaben komplexer werden und verschiedene Rollen sinnvoll erscheinen. Dabei ist entscheidend, die richtige Balance zwischen Autonomie und Kontrolle zu finden: **Human-in-the-Loop** bleibt wichtig, um Fehler rechtzeitig abzufangen.

Für produktionsreife Anwendungen sind **Frameworks** wie **PydanticAI** oder **LlamaIndex** interessant, da sie robuste Ausgabenformate und tiefe Datenintegration ermöglichen. Wichtig ist außerdem, die Leistung der Agenten nicht nur über Textqualität, sondern anhand klarer Benchmarks zu messen, beispielsweise, wie zuverlässig sie ein Ziel tatsächlich erreichen.

werden.

Die sogenannte **Typologieagentik** bietet eine methodische Klassifikation agentischer Systeme anhand von acht Dimensionen – etwa, wie autonom, kognitiv stark oder umweltangepasst ein Agent agiert. Ziel ist es, Standards für die Entwicklung und Bewertung zu schaffen.

Auch Architekturen wie das **ACE-Modell** (Aspirational, Cognitive, Executive) bieten Orientierung. Sie strukturieren Agenten ähnlich wie das OSI-Modell in der Netzwerktechnik in verschiedene Schichten: von der langfristigen Zielsetzung bis hin zur konkreten Task-Ausführung.

Fazit: Agentische Systeme sind die Zukunft

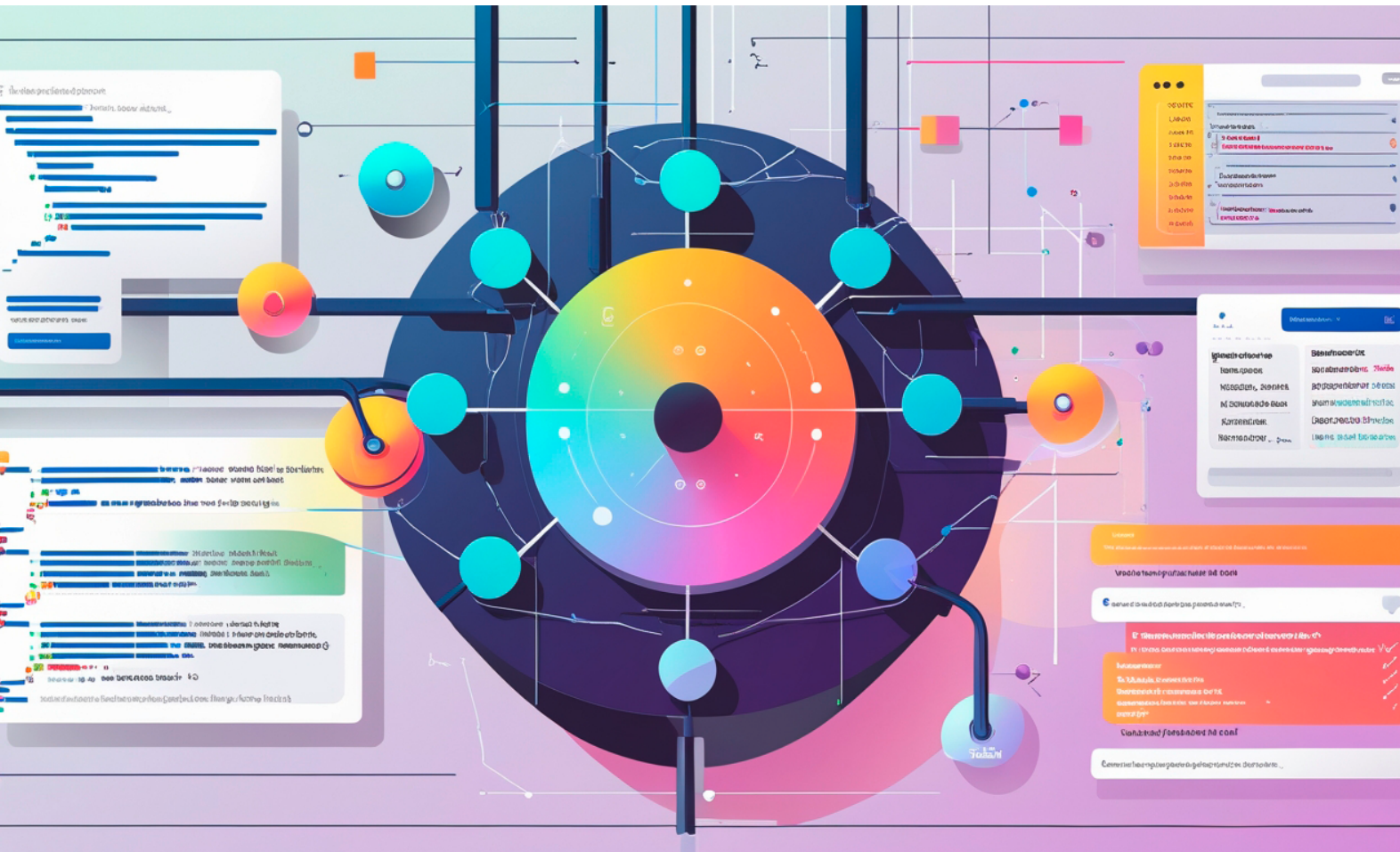
Agentische Systeme markieren einen Wendepunkt in der Anwendung von KI-Systemen: Sie sind nicht länger passive Werkzeuge, die nur auf Anfragen reagieren, sondern aktive Akteure, die planen, handeln und sich koordinieren können. Mit **Open-Source-Frameworks** wie beispielsweise **LangChain**,

AutoGen, **CrewAI**, **Smolagents**, **LlamaIndex** und **PydanticAI** steht Entwickler:innen heute ein reichhaltiges Ökosystem zur Verfügung, um eigene Agenten zu entwerfen.

Gleichzeitig zeigen industrielle Systeme wie Devin oder **ChatDev**, dass agentische Ansätze mehr sind als nur Forschungsexperimente – sie erobern bereits reale Anwendungsfelder. Und die Forschung arbeitet daran, diese neue Generation von KI zu systematisieren, zu standardisieren und sicher nutzbar zu machen.

Die Vision ist klar: In naher Zukunft werden Agenten nicht nur **digitale Tools** bedienen, sondern als eigenständige digitale Kolleg:innen auftreten, die gemeinsam mit Menschen komplexe Probleme lösen.

Eine Möglichkeit wie diese Vision von der gemeinsamen Arbeit umgesetzt werden kann sind **Retrieval-Augmented Generation (RAG)** Agenten. Diese Agenten können nicht nur tun was man ihnen sagt, sondern verstehen den Überblick und können dementsprechend eigenständig handeln. Wie genau das funktioniert, erklären wir Ihnen im nächsten Kapitel.



Kompetenzwissen

Agentic RAG: Die nächste Evolutionsstufe für intelligente KI-Systeme

David Golchinfar

Die Welt der künstlichen Intelligenz entwickelt sich rasant weiter. Große **Sprachmodelle (Large Language Models, LLMs)** wie GPT und Gemini haben bereits die Art und Weise, wie wir mit Informationen interagieren, revolutioniert. Doch ihre Abhängigkeit von statischen Trainingsdaten begrenzt ihre Fähigkeit, auf aktuelle, spezifische oder unternehmensinterne Informationen zuzugreifen. Hier kam die „**Retrieval-Augmented Generation (RAG)**“ ins Spiel – eine Technik, die **LLMs** mit externen **Wissensdatenbanken** verbindet. Doch nun steht die nächste, noch leistungsfähigere Stufe dieser Entwicklung bereit: **Agentic RAG**.

Dieser Beitrag taucht tief in die Welt von **Agentic RAG** ein. Wir erklären, was es ist, wie es funktioniert und warum es das Potenzial hat, die Grenzen dessen, was KI leisten kann, neu zu definieren.

Kurz wiederholt: Was war RAG noch mal?

Um **Agentic RAG** zu verstehen, werfen wir einen kurzen Blick auf das traditionelle **RAG**-Modell. Stellen Sie sich ein **LLM** als einen extrem sachkundigen, aber in der Vergangenheit lebenden Gelehrten vor. Sein Wissen ist riesig, aber auf den Zeitpunkt

seines „Trainingsabschlusses“ beschränkt.

RAG gibt diesem Gelehrten eine Bibliothek zur Hand. Wenn eine Anfrage hereinkommt, sucht das **RAG**-System zunächst in dieser Bibliothek (einer externen **Wissensdatenbank**, z. B. einer **Vektordatenbank** mit Unternehmensdokumenten) nach relevanten Informationen. Diese gefundenen „Schnipsel“ werden dann zusammen mit der ursprünglichen Anfrage an das **LLM** weitergegeben. Das Modell nutzt diesen zusätzlichen Kontext, um eine genauere, aktuellere und fundiertere Antwort zu generieren.

Dieser Ansatz löst bereits viele Probleme, insbesondere bei der Reduzierung von Falschaussagen („**Halluzinationen**“) und der Nutzung **proprietärer Daten**. Aber der Prozess ist noch immer recht linear und starr.

Der Sprung nach vorn: Was ist Agentic RAG?

Agentic RAG geht einen entscheidenden Schritt weiter. Statt eines starren, sequenziellen Prozesses wird ein **autonomer KI-Agent** in die Architektur integriert. Dieser Agent agiert wie ein intelligenter Projektmanager für die Beantwortung von Anfragen. Er kann eigenständig denken, planen, Entscheidungen treffen und verschiedene Werkzeuge nutzen.

Man kann es sich so vorstellen: Während traditionelles **RAG** einem Bibliothekar entspricht, der auf Anweisung ein bestimmtes Buch aus dem Regal holt, ist **Agentic RAG** ein erfahrener Forscher. Dieser Forscher versteht die Anfrage, überlegt, welche Informationsquellen (nicht nur eine!) am besten geeignet sind, kann bei Bedarf auch im Internet suchen, einen Taschenrechner benutzen oder sogar eine andere KI um Hilfe bitten.

Die vier Säulen von Agentic RAG

Die Funktionsweise und die Vorteile von **Agentic RAG** basieren auf vier zentralen Prinzipien:

1. Autonome Entscheidungsfindung: Der Agent analysiert eine Anfrage und entscheidet selbstständig, ob und welche zusätzlichen Informationen benötigt werden. Er kann eine Anfrage auch in mehrere Teilschritte zerlegen.

2. Dynamische Informationsbeschaffung: Im Gegensatz zum traditionellen **RAG**, das oft nur eine einzige Wissensquelle anzapft, kann der Agent auf ein ganzes Arsenal von Werkzeugen zugreifen. Dazu gehören **Vektordatenbanken**, Web-Su-

Anwendungsfälle: Wo Agentic RAG glänzt

Die Einsatzmöglichkeiten für Agentic RAG sind vielfältig und erstrecken sich über zahlreiche Branchen:

Kundenservice:

Ein **Agentic-RAG**-Chatbot kann nicht nur Standardfragen aus einer **Wissensdatenbank** beantworten, sondern bei einer komplexen Anfrage auch den Bestellstatus über eine API prüfen, Produktdetails aus einer anderen Datenbank abrufen und dem Kunden eine umfassende, personalisierte Antwort geben.

Markt- und Datenanalyse:

Ein System kann beauftragt werden, einen Bericht über die neuesten Markttrends zu erstellen. Der Agent würde aktuelle Finanznachrichten aus dem Web abrufen, Verkaufszahlen aus internen Datenbanken analysieren und die Ergebnisse in einer kohärenten Zusammenfassung präsentieren.

Wissensmanagement in Unternehmen:

Mitarbeiter können komplexe Fragen stellen wie: „Welche Leistungsbestandteile benötige ich für ein Angebot an einen Kunden mit den Anforderungen X, und wie hat sich dessen bisherige **Customer Journey** entwickelt?“ Der Agent würde Informationen aus dem internen Wiki, dem Ticketsystem und dem **CRM** aggregieren, um eine fundierte Empfehlung zu geben.

Gesundheitswesen:

Ein KI-Assistent könnte Ärzte unterstützen, indem er die neuesten Forschungsergebnisse zu einem bestimmten Krankheitsbild recherchiert, diese mit den anonymisierten Patientendaten abgleicht und mögliche Diagnose- oder Behandlungsoptionen vorschlägt.

chen, **APIs** zu Unternehmenssoftware oder sogar Taschenrechner. Der Agent wählt dynamisch das passende Werkzeug für den jeweiligen Teilschritt.

3. Erweiterte Generierung durch Iteration: Der Prozess ist nicht nach einer einzigen Suche beendet. Der Agent kann die ersten Ergebnisse bewerten. Sind sie ausreichend? Beantworten sie die Frage vollständig? Wenn nicht, kann er seine Strategie anpassen, die Suchanfrage umformulieren und erneut auf die Suche gehen. Dieser iterative Prozess führt zu wesentlich besseren und umfassenderen Ergebnissen.

4. Kontinuierliches Lernen und Feedback: Agentic RAG-Systeme können aus Interaktionen lernen. Durch die Analyse von Ergebnissen und (implizitem oder explizitem) Feedback können sie ihre zukünftigen Strategien zur Informationsbeschaffung verfeinern und so mit der Zeit immer effizienter und intelligenter werden.

Die Vorteile in der Praxis

Die Einführung eines autonomen Agenten in den **RAG**-Prozess bringt erhebliche Vorteile mit sich:

- **Höhere Genauigkeit und Relevanz:** Durch die Fähigkeit, mehrere Quellen zu nutzen und Suchanfragen zu verfeinern, sind die generierten Antworten deutlich präziser.
- **Komplexere Aufgabenbewältigung:** **Agentic RAG** kann mehrstufige Arbeitsabläufe handhaben, die weit über einfache Frage-Antwort-Szenarien hinausgehen.
- **Gesteigerte Effizienz:** Durch die Automatisierung von Recherche- und Analyseprozessen werden menschliche Mitarbeiter entlastet und können sich auf strategischere Aufgaben konzentrieren.
- **Flexibilität und Anpassungsfähigkeit:** Das System ist nicht auf eine einzige Datenquelle beschränkt und kann flexibel auf unterschiedlichste Anfragen reagieren.

Fazit: Die Zukunft ist „agentisch“

Agentic RAG ist mehr als nur ein weiteres Schlagwort in der KI-Welt. Es repräsentiert einen fundamentalen Wandel von reaktiven Informationssystemen hin zu **proaktiven**, autonomen Problemlösern. Indem wir KI-Systemen die Fähigkeit geben, zu planen, zu handeln und aus ihren Handlungen zu lernen, erschließen wir ein enormes Potenzial für die Automatisierung und Verbesserung

komplexer Arbeitsabläufe.

Während die Implementierung technisches Know-how erfordert, ist klar, dass Unternehmen, die auf intelligente und datengestützte Lösungen setzen, an **Agentic RAG** nicht vorbeikommen werden. Die Fähigkeit, autonom auf das gesamte Wissen eines Unternehmens und der Welt zuzugreifen und es intelligent zu nutzen, wird der entscheidende Wettbewerbsvorteil der nahen Zukunft sein.

Während diese neuen Errungenschaften in der KI-Technologie den Betriebsalltag erleichtern können und viele Potenziale mit sich bringen, gibt es durch die Anwendung von KI aber auch Hürden in Bezug auf die Kompetenzen von Mitarbeitenden im Betrieb. Während es logisch ist, dass die Menschen, die KI einsetzen auch die Kompetenzen zur Bedienung haben müssen, wurde dies jetzt auch auf den ganzen Betrieb ausgeweitet und im **EU AI Act** europaweit festgelegt. Was genau die Vorgaben sind und welche Verpflichtungen Betriebe haben die KI nutzen möchten, erfahren Sie im nächsten Kapitel.





Basiswissen

EU AI ACT: Potenziale und Hürden

Franziska Herrmann

Was der EU AI Act für KMU bedeutet: KI-Kompetenz als neue Pflicht

Der **EU AI Act** ist beschlossene Sache und bringt für Unternehmen, die **Künstliche Intelligenz (KI)** entwickeln, einsetzen oder betreiben, klare Regeln mit sich. Eine dieser Vorschriften betrifft etwas, das auf den ersten Blick abstrakt klingt, aber für kleine und mittlere Unternehmen (KMU) ganz konkrete Auswirkungen hat: die Pflicht zur Förderung von **KI-Kompetenzen (AI Literacy)** im eigenen Haus. Nicht nur die Entwickler von KI-Systemen sind angesprochen, sondern auch die KI-nutzenden Unternehmen.

Was bedeutet das für KMU konkret?

Auch wenn Ihr Unternehmen keine eigene KI entwickelt, sondern etwa ein digitales Tool wie ChatGPT, eine automatisierte Bewerbermanagementsoftware oder eine KI-basierte Bilderkennung einsetzt, gel-

Was genau fordert der AI Act?

Im Gesetzestext heißt es sinngemäß:

Anbieter und Betreiber von KI-Systemen müssen sicherstellen, dass ihr Personal sowie andere Personen, die mit dem Betrieb oder der Nutzung der KI-Systeme betraut sind, über ein ausreichendes Maß an KI-Kompetenz verfügen – abhängig von deren Wissen, Erfahrung und dem Nutzungskontext.

Konkret heißt das: Wer KI in seinem Unternehmen einsetzt, muss dafür sorgen, dass die betroffenen Mitarbeitenden verstehen, was sie tun, und die nötigen Kompetenzen haben, so dass sie verantwortungsvoll mit KI-Systemen umgehen können.

ten Sie als „**Deployer**“ (**Betreiber**) im Sinne des Gesetzes. Damit haben Sie Pflichten.

Die Schulung ist kein „Nice-to-have“ mehr, sondern eine Pflicht. Alle Mitarbeitenden, die mit einem KI-System arbeiten oder dessen Ergebnisse verwenden, müssen hinsichtlich folgender Fragestellungen angemessen geschult sein:

- Wie funktioniert das System grundsätzlich?
- Wo liegen seine Grenzen?
- Welche Risiken (z. B. **Verzerrungen**, Fehlentscheidungen) können auftreten?
- Was muss beachtet werden im Umgang mit (personenbezogenen) Daten?

Die Schulung muss zielgruppenspezifisch sein. Sie muss sich nach dem konkreten Einsatzkontext und dem Vorwissen der Mitarbeitenden richten. Das könnte beispielsweise bedeuten:

- IT-Team: vertiefte technische Schulung.
- HR-Team: Fokus auf diskriminierungsfreie Nutzung im Recruiting.
- Führungskräfte: Überblick über Risiken und **Compliance**.

Welche Schritte sollten KMU jetzt unternehmen?

1. Bestandsaufnahme machen: Welche KI-Systeme werden eingesetzt? Wer nutzt sie?
2. **Risikoeinstufung** prüfen: Gehört das eingesetzte System laut **EU AI Act** zu einer **Hochrisiko-Kategorie**?
3. Schulungsbedarf identifizieren: Wer muss was wissen?
4. Maßnahmen planen: Eigene Schulungen aufbauen oder externe Angebote nutzen.
5. Verantwortlichkeiten festlegen: Wer ist für Schulung, Dokumentation und **Nachhaltigkeit** zuständig?

Fazit: KI-Kompetenz ist Teil der digitalen Sorgfaltspflicht

Der **EU AI Act** macht deutlich: Wer KI nutzt, muss sie verstehen – und das gilt nicht nur für große

Tech-Konzerne. Auch KMU sind in der Pflicht, ihr Team fit zu machen im Umgang mit KI.

Der **EU AI Act** fordert von KMU seine Mitarbeiter:innen auf den neusten Wissensstand zu bringen, um gewissenhaft mit KI umgehen zu können und ihren Betrieb nach vorne zu bringen. Was aber, wenn der Betrieb durch **Cyberattacken** lahmgelegt wird, weil es keine KI gab, um vor diesen Angriffen zu schützen? Was, wenn der generelle Umgang mit **Cybersicherheit** nicht auf dem neusten Stand ist, und welche Implikationen haben die Entwicklungen in der KI eigentlich für die **Cybersicherheit**?

Weitere Informationen zum Thema EU AI Act können Sie in einer **Broschüre des Digital-Zentrum Kaiserslautern** finden.

Nachdem wir in den ersten Kapiteln erklärt haben, was KI eigentlich alles kann, möchten wir in den nächsten Kapiteln erklären, was Menschen alles können und wie KMU sich und ihre Mitarbeiter:innen vor **Cyberattacken** schützen können, ohne KI.



Basiswissen

Social Engineering: Die unsichtbare Waffe der Cyberkriminellen – und warum KI alles verändert

Michelle Walther

Cyberkriminalität ist längst nicht mehr nur eine technische Bedrohung. Während Firewalls, Virens Scanner und Sicherheitsupdates unsere Systeme schützen sollen, fokussieren sich viele Angriffe auf den schwächsten Punkt: den Menschen. Hier setzt **Social Engineering** an – eine Methode, bei der Kriminelle gezielt psychologische Tricks nutzen, um Vertrauen aufzubauen, Ängste zu schüren oder Neugier auszunutzen.

Was ist Social Engineering?

Unter **Social Engineering** versteht man die Manipulation von Menschen, um an sensible Informationen, Geld oder Zugänge zu gelangen. Dabei nutzen

Angrifer menschliche Bedürfnisse wie Hilfsbereitschaft, Vertrauen oder den Wunsch nach Sicherheit und Anerkennung.

Angriffe verlaufen dabei in mehreren Phasen: von der sorgfältigen Vorbereitung über den eigentlichen Kontakt bis hin zur Nachbereitung, bei der Spuren verwischt und Daten weiterverwertet werden.

Warum Social Engineering so gefährlich ist

Das Tückische: **Social Engineering** tarnt sich als Alltagssituation. Eine E-Mail wirkt wie eine gewöhnliche Nachricht, ein Anruf wie eine legitime Nach-

Typische Taktiken sind:

- **Autorität:** „Ich bin von der IT-Abteilung, geben Sie mir bitte Ihr Passwort.“
- **Knappheit & Zeitdruck:** „Nur heute: sichern Sie sich Ihr Gratis-Upgrade!“
- **Sympathie & Ähnlichkeit:** Ein vermeintlicher „Kollege“ meldet sich mit vertraut klingenden Details.
- **Neugier & Spieltrieb:** Verlockende Links oder Anhänge, die angeblich spannende Inhalte versprechen.

frage. Genau diese Alltäglichkeit macht es schwer, Manipulationen rechtzeitig zu erkennen.

Die Ziele können vielfältig sein: von digitalem Vandalismus über Rufschädigung bis hin zu Erpressung oder groß angelegtem **Identitätsdiebstahl**. In vielen Fällen ist Social Engineering das Einfallstor für größere Angriffe wie **Ransomware** oder Betrugs-maschinen im E-Commerce.

Künstliche Intelligenz (KI) macht Social Engineering noch raffinierter

Mit dem Aufkommen generativer KI-Systeme verschärft sich die Lage dramatisch. Während früher schlecht übersetzte Phishing-Mails oder auffällige Fake-Anrufe noch Warnsignale waren, können heute:

- KI-generierte Texte nahezu perfekt formuliert und personalisiert sein.
- **Deepfake**-Stimmen und -Videos echte Personen täuschend echt imitieren.
- **Chatbots** in Echtzeit überzeugende Gespräche führen, angepasst an die Reaktionen des Opfers.

Das bedeutet: Die bisherigen Warnsignale verschwinden zunehmend. Was früher durch Tippfehler oder unnatürliche Formulierungen auffiel, wirkt heute professionell, individuell zugeschnitten und glaubwürdig.

Prävention: Mensch und Maschine gemeinsam

Absolute Sicherheit gibt es nicht, aber **digitale Resilienz** ist möglich. Die wichtigsten Maßnahmen sind:

1. Bewusstsein schaffen: Schulungen und Aufklärung, damit Mitarbeitende typische Manipulationstechniken erkennen.

2. Technische Prävention: KI-basierte Betrugs-erkennung, die verdächtige Muster automatisch analysiert.

3. Gesunde Skepsis: Keine überstürzten Entscheidungen bei Zeitdruck oder ungewöhnlichen Anfragen treffen.

4. Mehrstufige Sicherheitsmechanismen: Multi-Faktor-Authentifizierung, um Zugriffe trotz erbeuteter Daten zu erschweren.

Fazit: Der Mensch als letzte Verteidigung

Social Engineering zeigt: **Cyberkriminalität** ist nicht nur ein technisches, sondern vor allem ein menschliches Problem. Mit KI wird die Erkennung noch schwieriger, da Angriffe glaubwürdiger und individueller wirken. Umso wichtiger ist es, dass wir lernen, kritische Fragen zu stellen – sowohl im privaten Alltag als auch in Unternehmen. Denn die beste Verteidigung ist nicht nur Technologie, sondern ein wachsameres Bewusstsein.

Eine der wohl häufigsten **Social Engineering Cyber-rattaken** sind Phishing E-Mails, die meist darauf zielen relevante Informationen (z.B. Zugangsdaten) von ihren Opfern zu erhalten. Wie Phishing genau definiert wird und was Mitarbeitende tun können, um diese zu erkennen, erläutern wir im folgenden Kapitel.



Basiswissen

Phishing erkennen und vermeiden: So schützen Sie sich vor digitalen Ködern

Michelle Walther

Phishing ist eine der häufigsten und gefährlichsten Formen von Cyberangriffen. Dabei geben sich Kriminelle als vertrauenswürdige Institutionen oder Personen aus, z.B. per E-Mail, SMS oder über gefälschte Websites, um an Ihre vertraulichen Daten zu gelangen. Das Ziel: Zugang zu Konten, Passwörtern oder finanziellen Informationen.

Warum ist Phishing so erfolgreich?

Weil es auf eines setzt: Unachtsamkeit. Die Angreifer bauen künstlichen Druck auf, nutzen psychologische Tricks und tarnen ihre Nachrichten geschickt – etwa als Nachricht vom Chef, von der Bank oder einem Online-Shop.

Ein Beispiel: So könnte eine Phishing-Mail aussehen

Die Sprache wirkt seriös, der Absender wirkt bekannt, aber die E-Mail ist gefälscht.

Absender: support@bank-sicherheit24.com

Betreff: Wichtiger Sicherheitshinweis

Text:

Sehr geehrter Kunde,

aufgrund eines Sicherheitsvorfalls benötigen wir eine Bestätigung Ihrer Zugangsdaten. Bitte klicken Sie auf folgenden Link, um Ihr Konto zu schützen:

[Jetzt bestätigen](#)

Vielen Dank,

Ihre Sparkasse

Was tun bei Verdacht?

- **Nicht klicken!** – Verdächtige Links oder Anhänge niemals öffnen.
- **Keine Daten eingeben!** – Geben Sie niemals vertrauliche Informationen preis.
- **Adresse prüfen!** – Stimmt die Absenderadresse wirklich mit der Institution überein?
- **IT kontaktieren!** – Im Zweifel sofort die IT-Abteilung oder eine Vertrauensperson informieren.
- **E-Mail melden oder löschen** – Phishing-Mails sollten nicht unbeantwortet im Postfach liegen bleiben.

Für Unternehmen: Wie Sie Mitarbeitende schützen

- **Aufklärung & Schulung** – Regelmäßige Awareness-Trainings helfen, Phishing zu erkennen. Hier bietet Mittelstand-Digital vielfältige Unterstützung. Insbesondere das Mittelstand-Digital Zentrum Mensch und die [Transferstelle Cybersicherheit](#) sind hier passende Anlaufstellen die bei der Orientierung weiterhelfen können.
- **Phishing-Simulationen** – Durch kontrollierte Tests kann man Reaktionsmuster verbessern.
- **Klare Meldewege schaffen** – Wer sich unsicher ist, muss wissen, an wen er sich wenden kann.

Fazit: Wachsamkeit ist der beste Schutz

Phishing ist kein technisches Problem – es ist ein menschliches. Wer weiß, worauf er achten muss, kann sich gut schützen. Wenn Sie sich bei einer E-Mail unsicher sind: Lieber einmal mehr nachfragen als einmal zu viel klicken.

Eine weitere [Social Engineering](#) Attacke sind gefälschte Kundenbewertungen. Diese werden genutzt, um die eigenen Produkte und Services zu promoten oder konkurrierende Produkte und Services zu demoten. Was gefälschte Kundenbewertungen genau sind, wie Sie sie erkennen und wie Sie mit ihnen umgehen sollten, erfahren Sie im nächsten Kapitel.

Die sechs häufigsten Erkennungsmerkmale von Phishing

1. Dringender Handlungsbedarf

„Ihr Konto wird gesperrt, wenn Sie nicht innerhalb von 24 Stunden reagieren!“

2. Drohungen oder Konsequenzen

„Wir haben verdächtige Aktivitäten festgestellt – bestätigen Sie Ihre Identität sofort!“

3. Aufforderung zur Eingabe vertraulicher Daten

z. B. PIN, TAN, Passwort oder Kreditkartennummer

4. Unerwartete Links oder Formulare

Sie erhalten einen Link zu einer Seite, obwohl Sie keine Aktion erwartet haben.

5. Verdächtige

URLs Der sichtbare Link sieht korrekt aus, führt aber zu einer ganz anderen Adresse. (Tipp: Mauszeiger über den Link bewegen – nicht klicken!)

6. Falsche Absenderadresse und unprofessionelles Layout

Die E-Mail scheint von Ihrer Bank oder Ihrem Vorgesetzten zu kommen – aber die E-Mail-Adresse passt nicht. Häufig sind auch die Anrede unpersönlich („Sehr geehrter Kunde“) oder das Layout wirkt chaotisch.



Basiswissen

Gefälschte Kundenbewertungen: Warum der richtige Umgang mit ihnen so wichtig ist

Michelle Walther

Warum Kundenbewertungen so mächtig sind

Ob wir ein Hotel buchen, einen neuen Staubsauger kaufen oder die nächste Powerbank aussuchen – Kundenbewertungen beeinflussen unsere Kaufentscheidungen heute mehr denn je. Studien zeigen: Bei gleichbleibendem Preis und Funktionsumfang hängen bis zu 90 % der Kaufentscheidungen von Kundenbewertungen ab. Positive Bewertungen schaffen Vertrauen und steigern die Verkäufe. Negative senken den Kaufwunsch erheblich.

Doch nicht alle Bewertungen sind echt.

Was sind Fake Reviews?

Fake Reviews sind gefälschte Kundenbewertun-

gen, die mit der Absicht verfasst werden, potenzielle Käufer:innen zu täuschen und ihre Entscheidungen zu manipulieren. Sie stammen oft von Personen ohne Produkterfahrung oder werden sogar gezielt im Auftrag von Unternehmen erstellt.

Man unterscheidet drei Haupttypen gefälschter Bewertungen:

- **Unwahre Aussagen:** frei erfundene Lob- oder Kritikpunkte
- **Markenbezogene Bewertungen:** Kommentare zur Firma, nicht zum Produkt
- **Werbung für andere Produkte:** Bewertungen, die auf Konkurrenzangebote verlinken oder ablenken

Warum werden Fake Reviews geschrieben?

Hinter **Fake Reviews** stecken oft profitorientierte Motive:

- Unternehmen wollen durch gekaufte Rezensionen besser gelistet oder Konkurrenzprodukte schlecht dargestellt sehen.
- Rezensenten erhalten im Gegenzug Geld, Gutscheine oder Gratisprodukte.
- Manche Nutzer posten aus Langeweile oder um „**Top-Bewerter**“-Abzeichen zu erhalten – auch das kann zu irreführenden Inhalten führen.

Wie erkennt man gefälschte Bewertungen?

Fake Reviews zu entlarven ist nicht leicht. Dennoch gibt es typische Merkmale, die Misstrauen wecken sollten:

- Extremes Urteil – ausschließlich 1- oder 5-Sterne-Bewertungen
- Unnötige Details – Wiederholung offensichtlicher Produktmerkmale
- **Emotionale Sprache** – ohne sachlichen Inhalt oder differenzierte Sicht
- Wenig bis keine Bilder
- Einseitigkeit – fehlender Ausgleich von Vor- und Nachteilen

Wie kann man sich schützen?

Für Konsument:innen heißt das: Bewertungen immer im Kontext betrachten. Kombinationen aus Nutzerfeedback (z. B. auf Amazon, Google), professionellen Tests (z. B. Stiftung Warentest) und unabhängigen Blogs bieten ein realistischeres Bild.

Für Unternehmen gilt:

- Regelmäßig Reviews überprüfen
- Verdächtige Einträge melden oder löschen lassen
- Konstruktiv und empathisch auf Kritik reagieren
- Rechtsweg nutzen, wenn Bewertungen nach-

Tipp:

Bewertungen, die wie Werbetexte klingen oder keine echten Nutzungserfahrungen beschreiben, sollten kritisch betrachtet werden.

weislich falsch sind (z. B. § 187 StGB oder § 3 UWG)

So reagieren Sie auf verschiedene Kundenbewertungen richtig

- Positive Bewertung mit Inhalt: Bedanken und auf die genannten Punkte eingehen.
- Positive Bewertung ohne Inhalt: Bedanken und nach konkretem Feedback fragen.
- Verbesserungsvorschläge enthalten? Bedanken und mitteilen, dass man das Feedback berücksichtigen wird.
- Negative Bewertung (vermutlich fake): Mitgefühl zeigen, nach Details fragen, gegebenenfalls melden oder löschen lassen.
- Kritik am Produkt oder Kaufprozess: Mitgefühl ausdrücken, sachlich reagieren, Kontaktmöglichkeiten anbieten.

Fazit: Transparenz schlägt Täuschung

Fake Reviews gefährden das Vertrauen in digitale Plattformen und damit langfristig auch den Umsatz. **Authentizität**, Transparenz und ein professioneller Umgang mit Kundenstimmen sind das wirksamste Mittel gegen **Bewertungsbetrug**.

Tipp für Unternehmen: Integrieren Sie Kundenbewertungen aktiv in Ihre Marketingstrategie und motivieren Sie zufriedene Kund:innen zur ehrlichen Bewertung – z. B. per QR-Code auf Rechnungen oder direkt im Ladengeschäft.

Mittlerweile werden gefälschte Kundenbewertungen aber nicht mehr nur im kriminellen Untergrund geschrieben, sondern Kriminelle nehmen aktiv Kontakt mit meist kleineren Unternehmen auf und versuchen diesen mit negativen Kundenbewertungen zu drohen, um eine Art Lösegeld oder Schutzgeld zu erhalten. Wie das genau funktioniert, wie man es erkennt, was man tun kann und wie man sich schützt, erfahren Sie im anschließenden Kapitel.



Basiswissen

Kundenbewertungen als Erpressung: Wenn Fake-Bewertungen zum Geschäftsrisiko werden

Michelle Walther

Immer mehr Unternehmen werden Opfer von **Bewertungs-Erpressungen**, bei denen Kriminelle Geld oder kostenlose Leistungen fordern, um negative **Fake-Bewertungen** zu verhindern oder zu löschen. Betroffene sollten keinesfalls zahlen, sondern Beweise sichern, die Bewertungen melden und gegebenenfalls rechtliche Schritte einleiten.

Online-Bewertungen sind ein mächtiges Werkzeug. Sie beeinflussen Kaufentscheidungen, das Vertrauen in eine Marke und letztlich den wirtschaftlichen Erfolg eines Unternehmens. Doch dieses System wird zunehmend von Kriminellen missbraucht: Immer häufiger sehen sich Unternehmen mit sogenannten **Bewertungs-Erpressungen** (**Review Extortion Scams**) konfrontiert.

Was ist ein Kundenbewertungs-Erpressungsscam?

Bei dieser Masche fordern Kriminelle Geld oder kostenlose Produkte im Austausch dafür, dass sie ent-

weder keine negativen Bewertungen veröffentlichen oder bereits veröffentlichte Falscheinträge wieder löschen. Oft erhalten Unternehmen E-Mails mit Drohungen wie:

„Wenn Sie uns nicht innerhalb von 48 Stunden antworten, werden wir dutzende 1-Stern-Bewertungen auf Google und anderen Plattformen veröffentlichen.“

Diese Drohungen kommen häufig von anonymen Konten oder unter Verwendung gestohlener Identitäten. Manchmal setzen die Täter ihre Drohung sofort um und fluten das Profil mit schlechten Bewertungen, um Druck auszuüben.

Wie erkennt man einen Erpressungsver-such?

- Ungewöhnliche E-Mail-Anfragen, in denen direkt Geld oder Leistungen im Gegenzug für Bewertungen gefordert werden.

Was sollte man tun, wenn man betroffen ist?

1. Nicht zahlen!

Auch wenn der Druck groß ist: Niemals auf die Forderung eingehen. Eine Zahlung schützt nicht vor weiteren Angriffen – im Gegenteil, sie signalisiert Schwäche.

2. Beweise sichern

Speichern Sie Screenshots der E-Mails, Bewertungen, Profilnamen und Zeitpunkte. Diese Informationen helfen später bei der Meldung an Plattformen oder Behörden.

3. Den Vorfall melden

- **Google:** Melden Sie unfaire Bewertungen direkt über Ihr Google-Unternehmensprofil. Nutzen Sie dabei die Option „Verstoß gegen Richtlinien“.
- **Trustpilot, Yelp, Amazon etc.:** Auch diese Plattformen bieten Meldefunktionen.
- **Polizei:** Erstellen Sie Anzeige wegen versuchter Erpressung, idealerweise mit allen gesicherten Beweisen.

4. Öffentlich transparent reagieren

Antworten Sie sachlich und professionell auf fragwürdige Bewertungen – etwa mit:

„Wir können Sie als Kunde leider nicht zuordnen. Bitte kontaktieren Sie uns direkt, damit wir das klären können.“

So zeigen Sie anderen Kunden, dass Sie seriös und offen mit Kritik umgehen.

5. Rechtlichen Beistand einholen

Ein auf IT-Recht oder Medienrecht spezialisierter Anwalt kann Sie unterstützen – insbesondere bei wiederholten Angriffen oder schwerwiegenden Reputationsschäden.

- Ein plötzlicher Schwall negativer Bewertungen, meist ohne konkrete Beschreibung oder mit auffälligen Formulierungen.
- Bewertungen von Profilen ohne Historie, mit identischem Wortlaut oder aus ähnlichen Regionen.
- Drohungen via E-Mail, Chat oder Kontaktformulare, die Bewertungen als Druckmittel einsetzen.

Wie kann man sich vorbeugend schützen?

- **Monitoring** einrichten: Tools wie unter anderem **Google Alerts** oder **Reputationsmanagement**-Software helfen, plötzliche Bewertungsänderungen früh zu erkennen.
- Echte Kunden zur Bewertung motivieren: Ein solides Fundament an authentischen Bewertungen macht Sie widerstandsfähiger gegen einzelne **Fake-Bewertungen**.
- Interne Prozesse klären: Schulen Sie Mitarbeiter im Umgang mit verdächtigen Mails und Kontaktanfragen.

Fazit: Digitale Erpressung nicht unterschätzen

Bewertungs-Erpressung ist kein Einzelfall mehr – sie ist ein ernstzunehmendes Risiko im digitalen Geschäftsalltag. Unternehmen sollten vorbereitet sein, professionell reagieren und rechtliche Möglichkeiten ausschöpfen. Wichtig ist: Lassen Sie sich nicht einschüchtern, sondern setzen Sie auf Transparenz, Integrität und rechtlich sauberes Vorgehen.

Diese Transparenz, Integrität und rechtlich sauberes Vorgehen ist aber natürlich nicht nur bei Kundenbewertungen einzuhalten, sondern zieht sich durch alle Themen, die wir in diesem Kurzlehrbuch besprochen haben. Im nächsten Kapitel möchten wir Ihnen nochmals die wichtigsten Erkenntnisse zusammenfassen und Ihnen erläutern, wie das Mittelstand-Digital Zentrum Fokus Mensch Sie und Ihre Unternehmen bei Fragen zu KI und **Digitaler Souveränität** unterstützen kann.

Schlusswort

Mit diesem Kurzlehrbuch möchten wir kleinen und mittleren Unternehmen (KMU) eine fundierte und praxisnahe Grundlage bieten, um die Herausforderungen und Chancen rund um **Cybersicherheit** und **generative Künstliche Intelligenz (GenKI)** zu erkennen und durch Einsatz gezielter Maßnahmen souverän zu meistern. Die digitale Transformation schreitet unaufhaltsam voran und mit ihr verändern sich Arbeitsprozesse, Geschäftsmodelle und Anforderungen an Mitarbeitende. Unser Ziel war es, die wichtigsten Wissensbausteine kompakt und verständlich aufzubereiten, damit Unternehmen nicht nur reagieren, sondern aktiv gestalten können.

Dabei steht nicht nur die Technologie im Mittelpunkt, sondern vor allem der Mensch. Denn **digitale Souveränität** bedeutet, Technologien bewusst und selbstbestimmt einzusetzen – mit einem klaren Verständnis für ihre Möglichkeiten, Grenzen und Risiken. Mitarbeitende sind dabei keine Schwachstelle, sondern ein entscheidender Schutzfaktor. Wenn sie verstehen, wie KI-Systeme funktionieren, wie sie mit ihnen zusammenarbeiten können und wie sie sich gegen Cyberbedrohungen schützen können, entsteht eine resiliente Unternehmenskultur, die Innovation und Sicherheit vereint.

Um den Transfer von Wissen in die Praxis zu erleichtern, möchten wir zum Abschluss die fünf zentralen Erkenntnisse dieses Kurzlehrbuches hervorheben und zeigen, wie das Mittelstand-Digital Zentrum Fokus Mensch Sie konkret unterstützen kann:

1. **Die Arbeitswelt verändert sich – wir müssen uns mit verändern.** Neue Technologien und digitale Bedrohungen erfordern neue Kompetenzen. Wir bieten Workshops, Vorträge und Austauschformate, um Sie und Ihre Mitarbeitenden auf dem neuesten Stand zu halten.
2. **Generative KI kann schon bei kleinen Aufgaben große Wirkung entfalten.** Ob Textgenerierung, Ideenfindung oder Prozessoptimierung – wir zeigen anhand konkreter Use Cases, wie **GenKI** sinnvoll eingesetzt werden kann und begleiten Sie in der Kompetenzentwicklung.
3. **Die Entscheidung für KI ist individuell – ebenso die Wahl der passenden Anwendung.** Es gibt nicht die eine „richtige“ KI. Wir helfen Ihnen dabei, die für Ihre Ziele und Rahmenbedingungen geeigneten Lösungen zu identifizieren und strategisch zu integrieren.
4. **KI-Nutzung erfordert rechtliche Orientierung – insbesondere im Hinblick auf den EU AI Act.** Der rechtliche Rahmen ist komplex, aber entscheidend. Wir unterstützen Sie dabei, den **EU AI Act** zu verstehen und zeigen anhand von Beispielen, wie er in Unternehmensprozesse eingebunden werden kann.
5. **Menschen müssen Teil der Lösung sein – sowohl bei KI als auch bei Cybersicherheit.** Awareness, Schulung und Beteiligung sind entscheidend. Wir bieten Formate, die Ihre Teams sensibilisieren, befähigen und stärken – für eine sichere und verantwortungsvolle digitale Zukunft.

Wir hoffen, dass dieses Kurzlehrbuch Ihnen wertvolle Impulse liefert, Orientierung bietet und konkrete Handlungsmöglichkeiten aufzeigt. Die digitale Transformation ist kein Selbstläufer, sie braucht Wissen, Mut und Zusammenarbeit. Als Mittelstand-Digital Zentrum Fokus Mensch stehen wir Ihnen dabei zur Seite: mit Expertise, Praxisnähe und einem klaren Fokus auf den Menschen.

Lassen Sie uns gemeinsam daran arbeiten, den Mittelstand zukunftsfähig, sicher und innovativ aufzustellen.

Herzliche Grüße Ihr Team vom Mittelstand-Digital Zentrum Fokus Mensch

Kooperationspartner: Denkschmiede

Glossar

ACE-Modell – Architekturmodell, das Agenten in drei Schichten gliedert: Aspirational (Ziele), Cognitive (Denken), Executive (Handeln). Dient als konzeptionelle Struktur für agentische Systeme.

Adapter – Softwarekomponente, die eine Verbindung zwischen einem **Sprachmodell** und einer Anwendung herstellt und definiert, welche Funktionen oder Daten das Modell nutzen darf.

Agentic RAG – Erweiterte Form der Retrieval-Augmented Generation (**RAG**), bei der ein **autonomer KI-Agent** eingesetzt wird. Dieser kann selbstständig planen, Entscheidungen treffen und verschiedene Informationsquellen oder Tools nutzen, um komplexe Aufgaben zu lösen.

Alerts – Automatische Benachrichtigungen bei bestimmten Ereignissen oder Fehlern im System.

Allow List – Liste erlaubter Funktionen, die das Modell aufrufen darf. Verhindert, dass ungewollte oder gefährliche Aktionen ausgeführt werden.

Anbieter (Provider) – Unternehmen oder Organisationen, die ein KI-System entwickeln oder auf den Markt bringen. Sie sind verantwortlich für dessen Sicherheit und Konformität mit dem **EU AI Act**.

API (Application Programming Interface) – Programmierschnittstelle, über die (Software-) Programme miteinander kommunizieren, Daten austauschen oder Funktionen aufrufen können, z. B. zur Nutzung externer KI-Modelle; **MCP** verwendet API-ähnliche Nachrichtenformate. **Agentic-RAG**-Systeme nutzen APIs, um auf externe Tools oder Datenbanken zuzugreifen.

Audit Trail – Lückenlose Aufzeichnung von Abläufen und Systemaktionen, um **Compliance**-Anforderungen zu erfüllen.

Auditierbarkeit – Nachvollziehbarkeit und Protokollierung von Aktionen eines Systems, wichtig für **Compliance** und Sicherheit.

Authentizität – Echtheit und Glaubwürdigkeit von Bewertungen als Grundlage für Vertrauen und langfristigen Geschäftserfolg.

AutoGen – Framework von Microsoft zum Aufbau von **Multi-Agent-Systemen**. Ermöglicht die Zusammenarbeit mehrerer spezialisierter Agenten und unterstützt hybride Mensch-KI-Teams.

Automatisch übersetzter Benchmark – Ein aus dem Englischen maschinell übertragener Vergleichstest; kann durch **Übersetzungsartefakte** seine Aussagekraft verlieren.

Autonomer KI-Agent – Selbstständig agierendes KI-System, das Aufgaben planen, Entscheidungen treffen und mehrere Werkzeuge nutzen kann, ohne direkte menschliche Steuerung.

Benchmark – Vergleichstest zur Bewertung der Leistungsfähigkeit von **Sprachmodellen** anhand standardisierter Aufgaben. z. B. hinsichtlich Zielerreichung oder Ergebnisqualität.

Betreiber (Deployer) – Unternehmen, die KI-Systeme einsetzen, aber nicht selbst entwickeln. Auch sie unterliegen Pflichten im Rahmen des **EU AI Act**, z. B. Schulungspflichten und Risikoprüfungen.

Bewertungs-Erpressung – Form der digitalen Erpressung, bei der Kriminelle Geld oder kostenlose Leistungen fordern, um negative **Fake-Bewertungen** zu verhindern oder zu löschen. **Review Extortion Scam** im englischen.

Bewertungsbetrug – Systematische Manipulation von Bewertungsplattformen durch gefälschte Rezensionen oder koordiniertes Verhalten.

CAD (Computer-Aided Design) – Software zur Konstruktion technischer Zeichnungen und Modelle; kann über **MCP** sprachgesteuert durchsucht werden.

Chain-of-Thought – Textuell oder logisch dargestellte Folge von Zwischenschritten; hilft Nutzenden, die Argumentation der KI besser nachzuvollziehen.

Chatbot – Software, die mittels KI sowohl schriftlich als auch in gesprochener Sprache natürliche Gespräche mit Menschen führen kann. Wird im **Social Engineering** genutzt, um in Echtzeit überzeugende, personalisierte Kommunikation zu simulieren.

ChatDev – Experimentelles **Multi-Agent-System**, das eine virtuelle Softwarefirma simuliert, in der verschiedene Agenten (z. B. CEO, Entwickler, Tester) zusammenarbeiten.

Cloud-Plattform – Externe Serverinfrastruktur, auf der Anwendungen oder KI-Modelle betrieben werden, ohne lokale Hardware zu benötigen.

Compliance – Einhaltung rechtlicher, regulatorischer und ethischer Vorgaben beim Einsatz von KI-Systemen. Dazu zählen **Datenschutz**, Transparenz, Fairness und Nachvollziehbarkeit.

Compliance-Verstoß – Nichteinhaltung rechtlicher/organisatorischer Vorgaben (z. B. durch ungeeignete Modelle oder fehlerhafte Evaluierungen).

CRM (Customer-Relationship-Management) – System zur Verwaltung von Kundenbeziehungen, Kontakten und Interaktionen. Eine wichtige Datenquelle für Agentic-RAG-Systeme.

CrewAI – Framework für kollaborative **Multi-Agent-Systeme** mit klar definierten Rollen (z. B. Researcher, Planner, Executor). Besonders geeignet für Geschäftsprozesse und Projektmanagement.

Customer Journey – Gesamter Verlauf der Interaktionen eines Kunden mit einem Unternehmen. **Agentic RAG** kann Daten aus **CRM**-, Support- und Marketing-Systemen kombinieren, um diesen Verlauf zu analysieren.

Cyberattacke – Gezielter Angriff auf IT-Systeme, um Schaden zu verursachen oder Daten zu entwenden. Beispiele sind Phishing, **Ransomware** oder Datenlecks.

Cyberkriminalität – Sammelbegriff für alle kriminellen Handlungen, die mit Computern, Netzwerken oder digitalen Technologien durchgeführt werden.

Cybersecurity/Cybersicherheit – Schutz von IT-Systemen, Netzwerken und Daten vor Angriffen, Manipulation oder Diebstahl. Eng verknüpft mit dem sicheren Betrieb von KI-Systemen.

Datenethiker:in – Expert:in, die sich mit ethischen und gesellschaftlichen Fragen im Umgang mit KI und Daten beschäftigt.

Datenschutz – Schutz personenbezogener und sensibler Daten vor unbefugtem Zugriff oder Nutzung, besonders relevant bei KI-Anwendungen.

Datensouveränität – Kontrolle über die eigenen Daten, insbesondere darüber, wo sie gespeichert und verarbeitet werden.

Deepfake – Durch KI generiertes oder manipuliertes Audio-, Bild- oder Videomaterial, das reale Personen täuschend echt imitieren kann. Wird häufig für Identitätsbetrug oder gezielte Desinformation eingesetzt.

Demokratisierung von Wissen – Verbreitung von Fachwissen durch KI, die komplexe Fähigkeiten breiteren Nutzergruppen zugänglich macht.

Deterministisches Parsing – Eindeutige Auswertung strukturierter Daten, sodass gleiche Eingaben immer gleiche Ergebnisse liefern.

Digitale Erpressung – Allgemeiner Begriff für die Nutzung digitaler Mittel, um Unternehmen oder Personen zu bedrohen und zur Herausgabe von Geld oder Leistungen zu zwingen.

Digitale Köder – Täuschungsmanöver im Internet, die darauf abzielen, Nutzer zu bestimmten Handlungen (z. B. Klicks, Dateneingaben) zu verleiten.

Digitale Resilienz – Fähigkeit eines Unternehmens, sich gegen digitale Risiken wie Cyberangriffe oder Systemausfälle zu schützen und nach Zwischenfällen schnell wieder funktionsfähig zu werden.

Digitale Sorgfaltspflicht – Verantwortung eines Unternehmens, bei Einsatz digitaler Technologien – insbesondere KI – sicherzustellen, dass Risiken erkannt, Mitarbeitende geschult und Systeme verantwortungsvoll betrieben werden.

Digitale Souveränität – Fähigkeit eines Unternehmens, einer Einrichtung, Person oder Staates, KI-Systeme und Daten unabhängig, sicher und nachvollziehbar zu betreiben, ohne von externen **Anbietern** abhängig zu sein.

Digitale Tools – Softwarelösungen, die Arbeitsprozesse unterstützen oder ersetzen.

DSGVO (Datenschutz-Grundverordnung) – Europäische Verordnung zum Schutz personenbezogener Daten; zentrale rechtliche Grundlage für KI- und Datenanwendungen in Europa. Regelt den Umgang mit personenbezogenen Daten und stellt hohe Anforderungen an Transparenz und Sicherheit.

Emergentes Verhalten – Komplexes, unerwartetes Verhalten, das aus der Interaktion vieler Agenten entsteht, ohne explizit programmiert zu sein.

Emotionale Sprache – Übermäßig gefühlsbetonte oder werbliche Formulierungen statt sachlicher Nutzungserfahrungen.

ERP (Enterprise Resource Planning) – Unternehmenssoftware zur Steuerung von Geschäftsprozessen wie Einkauf, Produktion oder Finanzen.

EU AI Act – Europäische Verordnung zur Regulierung von Künstlicher Intelligenz. Sie legt Anforderungen an Entwicklung, Einsatz und Betrieb von KI-Systemen fest, um Sicherheit, Transparenz und Grundrechte zu gewährleisten.

Evaluationsmetriken – Messgrößen, mit denen die Qualität, Genauigkeit und Verständlichkeit von KI-Erklärungen bewertet werden können.

Explainable AI (XAI) – Forschungs- und Anwendungsfeld der Künstlichen Intelligenz, das Methoden entwickelt, um Entscheidungen, Vorhersagen oder generierte Inhalte von KI-Systemen transparent, nachvollziehbar und überprüfbar zu machen. Ziel ist es, Vertrauen zu schaffen und Fehlentscheidungen zu vermeiden.

Fake-Bewertung – Falsche oder manipulierte Online-Bewertung, die nicht von echten Kundinnen oder Kunden stammt.

Framework – Technisches Gerüst oder Baukasten, der die Entwicklung von KI-Agenten vereinfacht, indem er wiederverwendbare Komponenten und Schnittstellen bereitstellt.

Function-Calling – Mechanismus, mit dem ein **Sprachmodell** gezielt externe Funktionen oder Tools aufrufen kann. Das Modell gibt strukturierte Daten (meist im **JSON**-Format) aus, die von einer Anwendung interpretiert und ausgeführt werden können.

Gefälschte Kundenbewertungen (Fake Reviews) – Gefälschte oder manipulierte Rezensionen, die erstellt werden, um Kaufentscheidungen zu beeinflussen oder den Ruf von Produkten/Firmen zu manipulieren. Z.B. durch eigene Produkte oder Dienstleistungen besser oder Konkurrenzprodukte schlechter darzustellen.

Generative KI (GenKI) – Künstliche Intelligenz, die neue Inhalte wie Texte, Bilder, Videos oder Stimmen erzeugen kann. Grundlage vieler moderner Social-Engineering-Angriffe, da sie täuschend echte Inhalte erstellt.

Governance – Regelwerk zur Steuerung und Kontrolle des Einsatzes digitaler Technologien, insbesondere im Hinblick auf Sicherheit und Verantwortung.

Google Alerts – Kostenloses Tool von Google, das Nutzer automatisch über neue Online-Erwähnungen zu bestimmten Schlagwörtern informiert.

Guardrails – Sicherheitsmechanismen, die Eingaben und Ausgaben überwachen. Schützen vor Missbrauch, prüfen Parameter und sorgen für Einhaltung von Geschäftsregeln.

Halluzination – Falschaussage eines **Sprachmodells**, die plausibel klingt, aber nicht auf realen Daten basiert. **RAG** reduziert dieses Risiko, indem es externe, überprüfbare Informationen einbindet.

Hochrisiko-KI-System – Kategorie des **EU AI Act** für besonders sicherheits- oder grundrechtsrelevante Anwendungen, etwa im Gesundheitswesen, Personalmanagement oder bei kritischer Infrastruktur. Solche Systeme unterliegen strengen Anforderungen an Transparenz, Dokumentation und Kontrolle.

Hugging Face – Plattform und Community für Open-Source-KI-Modelle, Werkzeuge und Datensätze.

Human-in-the-Loop – Prinzip, bei dem Menschen weiterhin in den Entscheidungsprozess eingebunden bleiben, um Fehler zu vermeiden und Kontrolle zu behalten.

Hybrid-Szenario – Kombination aus lokal betriebenen (**on-premises**) und Cloud-basierten Modellen oder Tools, um Flexibilität und Kontrolle zu vereinen.

Hybrid-Strategie – Kombination von **Open-Source**- und proprietären KI-Modellen, um Kosten, **Datenschutz** und Leistungsfähigkeit auszubalancieren.

Hybrides Setup – Kombination aus autonomen Agenten und menschlicher Kontrolle, um Flexibilität und Sicherheit in **Multi-Agent-Systemen** zu gewährleisten.

Identitätsdiebstahl – Unbefugte Nutzung personenbezogener Daten, um sich als jemand anderes auszugeben, meist mit betrügerischer Absicht.

Iteration – Wiederholter Prozess, bei dem ein Agent Zwischenergebnisse prüft, Suchstrategien anpasst und erneut nach besseren Informationen sucht. Erhöht Genauigkeit und Vollständigkeit der Ergebnisse.

JSON – Abkürzung für JavaScript Object Notation. Ein strukturiertes, maschinenlesbares Datenformat, das zur eindeutigen Ausgabe von Funktionsaufrufen dient.

JSON-Schema – Definition, wie die JSON-Daten aufgebaut sein müssen, welche Felder erlaubt sind und welche Datentypen verwendet werden.

KI-basierte Agenten-Systeme – Systeme, die große **Sprachmodelle** mit Werkzeugnutzung, Gedächtnis und Entscheidungslogik verbinden. Sie ermöglichen es KI, selbstständig zu planen, zu handeln und Ziele zu verfolgen.

KI-basierte Betrugserkennung – Systeme, die mithilfe von Künstlicher Intelligenz verdächtige Verhaltensmuster analysieren und potenzielle Angriffe oder Täuschungen automatisch erkennen.

KI-Kompetenz (AI Literacy) – Fähigkeit, die Funktionsweise, Möglichkeiten und Risiken von KI-Systemen zu verstehen. Beinhaltet technisches Grundwissen, kritisches Denken und verantwortungsvollen Umgang mit KI-Ergebnissen.

KI-Strateg:in – Fachperson, die den gezielten und verantwortungsvollen Einsatz von KI in Organisationen plant.

KI-Verordnung (EU AI Act) – EU-Regelwerk zur Regulierung des Einsatzes Künstlicher Intelligenz mit Fokus auf Transparenz, Sicherheit und ethische Standards.

Künstliche Intelligenz (KI) – Technologien, die menschenähnliche Denk- und Entscheidungsprozesse simulieren, z. B. durch maschinelles Lernen oder Sprachverarbeitung.

LangChain – **Open-Source-Framework** zur Verbindung von **Sprachmodellen** mit Tools, Datenquellen und Gedächtnis. Unterstützt das **ReAct-Pattern** zur schrittweisen Handlungslogik.

Large Language Model (LLM) – Großes **Sprachmodell**, das auf umfangreichen Textdaten trainiert wurde, natürliche Sprache verstehen und erzeugen kann und komplexe Aufgaben wie Textverständnis oder Dialog führen kann. Beispiele: GPT, Gemini, Claude.

LLaMA / Mistral – **Open-Source-Sprachmodelle**, die lokal oder auf eigenen Servern betrieben werden können und als Alternativen zu kommerziellen Lösungen gelten.

LlamaIndex – **Framework** zur Anbindung von **Sprachmodellen** an Wissensquellen wie Datenbanken, Dokumente oder **APIs**. Ermöglicht wissensgestützte Agenten.

Logging – Protokollierung aller Modellaufrufe, Entscheidungen und Systemereignisse, um Nachvollziehbarkeit und Transparenz zu gewährleisten.

Low-Code / No-Code – Plattformen, die Softwareentwicklung ohne oder mit minimaler Programmierung ermöglichen.

Meldewege – Festgelegte interne Prozesse, über die Mitarbeitende verdächtige Nachrichten oder Sicherheitsvorfälle melden können.

Model Context Protocol (MCP) – Standardisierte Schnittstelle, die eine gemeinsame Sprache zwischen **Sprachmodellen** und Unternehmenssystemen schafft. Ermöglicht strukturierte, kontrollierte und sichere Interaktion mit internen Datenquellen und Tools.

Monitoring – Systematische Beobachtung von Online-Aktivitäten, um Veränderungen, Bedrohungen oder negative Bewertungen frühzeitig zu erkennen.

Multi-Agent-System – System, in dem mehrere KI-Agenten gemeinsam agieren, kommunizieren und Aufgaben arbeitsteilig lösen.

Multi-Faktor-Authentifizierung (MFA) – Sicherheitsverfahren, das mehrere unabhängige Nachweise erfordert (z. B. Passwort + Einmalcode).

Nachhaltigkeit (im Kontext von KI) – Sicherstellung, dass Schulungs- und Kontrollmaßnahmen langfristig wirksam bleiben und regelmäßig überprüft oder aktualisiert werden.

On-Premise – Betrieb von Software oder KI-Modellen auf eigener Hardware im Unternehmen. Bietet volle Kontrolle über Daten, erfordert aber mehr Wartung.

Open-Source – Software, deren Quellcode öffentlich zugänglich ist und von Dritten genutzt, geprüft und angepasst werden darf.

Open-Source LLM – Sprachmodell, dessen Quellcode oder Modellparameter öffentlich zugänglich sind und das frei angepasst oder lokal betrieben werden kann.

OpenAI-kompatible API – Programmierschnittstelle, die dem OpenAI-Standard folgt und es ermöglicht, Modelle oder Tools auszutauschen, ohne Anwendungen neu entwickeln zu müssen.

Open-Source-Framework – Frei verfügbares Software-Framework, dessen Quellcode offen zugänglich ist und von der Community erweitert werden kann.

Open-Source-Modelle – Frei verfügbare KI-Modelle, deren Architektur und Trainingsdaten offen zugänglich sind. Sie fördern Transparenz, Anpassbarkeit und Unabhängigkeit.

Ökosystem-Tools – Unterstützende Software, Plattformen und Communities, die Entwicklung, Training und Einsatz von KI-Modellen erleichtern.

Orchestrator – Systemkomponente, die zwischen **Sprachmodell** und Unternehmenssystemen vermittelt. Führt Funktionsaufrufe aus, prüft Ergebnisse und gibt sie an das Modell zurück.

Proaktives System – KI-System, das nicht nur auf Anfragen reagiert, sondern aktiv plant, recherchiert und Vorschläge zur Problemlösung erarbeitet.

Proprietäre Daten – Unternehmensinterne oder vertrauliche Informationen, die nicht öffentlich zugänglich sind.

Proprietäre KI-Systeme – Geschlossene, kommerzielle KI-Modelle, die von Unternehmen kontrolliert werden.

Proprietäres LLM – Geschlossenes, kommerzielles **Sprachmodell**, das über Cloud oder API genutzt wird; Code und Datenverarbeitung bleiben beim Anbieter.

Prompt Engineer – Fachperson, die präzise Eingaben (Prompts) für KI-Modelle entwickelt, um optimale Ergebnisse zu erzielen.

Prompt Injection – Manipulation eines **Sprachmodells** durch gezielte Eingaben, um es zu ungewollten oder gefährlichen Aktionen zu verleiten. **MCP** hilft, dieses Risiko zu minimieren.

Prompting-Strategien – Vorgehensweisen zur gezielten Gestaltung von Eingaben (Prompts), um stabile, nachvollziehbare und qualitativ hochwertige Ausgaben zu erzeugen.

PydanticAI – Framework mit Fokus auf strenge Typprüfungen und Validierung von Agenten-Ausgaben. Ideal für produktionsreife Anwendungen mit hohen Qualitätsanforderungen.

RAG (Retrieval-Augmented Generation) – Verfahren, das ein Sprachmodell (**LLM**) mit externen Wissensquellen verbindet.

Reasoning – Verfahren zur expliziten Darstellung oder Prüfung von Schlussfolgerungsschritten eines KI-Modells, um Entscheidungen besser bewerten zu können. Es dient der Transparenz und performativer Verbesserung.

ReAct-Pattern – Vorgehensweise, bei der ein Agent abwechselnd denkt („thought“), handelt („action“) und reflektiert („observation“), um nachvollziehbare Entscheidungsprozesse zu erzeugen.

Regulierte Branchen – Bereiche wie Gesundheit, Recht oder Verwaltung, in denen Fehlbewertungen zusätzlich **Compliance**-Risiken erzeugen.

Reputationsmanagement – Maßnahmen zur Überwachung und Pflege des Online-Rufs eines Unternehmens.

Review Extortion Scam – Englischer Begriff für **Bewertungs-Erpressung**; Betrugsversuche, bei denen Bewertungen als Druckmittel eingesetzt werden.

Ransomware – Schadsoftware, die Daten verschlüsselt und deren Freigabe erst nach Zahlung eines Lösegelds verspricht.

Risikoeinstufung – Bewertung eines KI-Systems gemäß **EU AI Act** nach potenziellen Auswirkungen auf Sicherheit, Grundrechte und Gesellschaft.

Salesforce Einstein GPT Agents – Industrielles Beispiel für orchestrierte Multi-Agent-Workflows im Kundenservice. Kombiniert **Sprachmodelle** und **CRM**-Funktionalität.

Smolagents – Minimalistisches **Framework** von **Hugging Face** zur schnellen Entwicklung kleiner Agenten-Prototypen.

Social Engineering – Methode der Manipulation von Menschen, um an vertrauliche Informationen, Zugänge oder Geld zu gelangen. Angreifer nutzen psychologische Tricks.

Sovereign Cloud – Europäische Cloud-Infrastruktur, die **Datenschutzstandards** wie die **DSGVO** einhält und Datenhoheit sicherstellt.

Sprachmodell (LLM – Large Language Model) – KI-System, das auf großen Textmengen trainiert ist und menschliche Sprache verstehen und erzeugen kann. Grundlage vieler generativer KI-Anwendungen.

Sprachspezifische Evaluierung – Bewertung eines Modells in der Zielsprache (hier Deutsch) mit passenden Aufgaben und Daten statt reiner Übernahme englischer Tests.

SwarmAgentic – Forschungsframework (2025), das Schwarmintelligenz zur automatischen Koordination und Emergenz von Agentenverhalten nutzt.

Technische Prävention – Einsatz technischer Schutzmaßnahmen wie KI-gestützte Anomalieerkennung, Firewalls oder E-Mail-Filter.

Top-Bewerter-Anreiz – Nutzer, die aus Statusgründen oder Gamification-Belohnungen Bewertungsaktivität zeigen.

Transformer – Architekturtyp moderner **Sprachmodelle**, der Informationen parallel und verteilt verarbeitet.

Typologieagentik – Methodischer Ansatz zur Klassifikation agentischer Systeme anhand von Kriterien wie Autonomie, Kognition und Umweltanpassung.

Übersetzungsartefakt – Unbeabsichtigte Veränderung von Inhalt, Bedeutung oder Struktur durch automatische Übersetzung, die Aufgaben verfälscht.

URL – Abkürzung für Uniform Resource Locator; die Internetadresse, über die eine Website aufgerufen wird.

Unsicherheitsindikatoren – Metriken oder Signale, die anzeigen, wie sicher oder zuverlässig eine generierte Erklärung oder ein Ergebnis ist.

Verifikationsmechanismen – Verfahren zur unabhängigen Überprüfung, ob die von der KI gelieferten Erklärungen korrekt oder konsistent sind.

Vektordatenbank – Datenbank, die Informationen in Form von Vektoren speichert, um semantisch ähnliche Inhalte schnell zu finden.

Verzerrung (Bias) – Systematische Fehlentscheidung oder Ungleichbehandlung durch ein KI-Modell, verursacht durch unausgewogene Trainingsdaten oder algorithmische Voreingenommenheit.

Wissensdatenbank – Externe Datensammlung, aus der ein **RAG**- oder **Agentic-RAG**-System relevante Informationen abrufen.

Wissensmanagement – Systematische Erfassung, Speicherung und Nutzung von Wissen innerhalb eines Unternehmens.

Zero-Trust-Prinzip – IT-Sicherheitsansatz, bei dem kein System oder Nutzer automatisch vertraut wird; jeder Zugriff muss überprüft und autorisiert werden.

Autorenverzeichnis



Dr. Daryoush Daniel Vaziri

Daryoush Vaziri ist promovierter Wirtschaftsinformatiker und Experte für das Thema generative KI im Mittelstand. Er besitzt jahrelange Erfahrung bei der Konzeption und Gestaltung von KI-basierten Systemen und tragfähigen Geschäftsmodellen, insbesondere für Innovationen der natürlichen Sprachverarbeitung und generativen KI.

Kontakt: d.vaziri@digitalzentrum-fokus-mensch.de

David Golchinfar

David Golchinfar ist leitender Entwickler im Bereich generative Sprachmodelle und absolvierte sein Informatikstudium an der Hochschule Bonn-Rhein-Sieg. Beim Mittelstand-Digital Zentrum Fokus Mensch arbeitet er an KI-Systemen, die digitale Souveränität demonstrieren, und ist Experte für generative Sprachmodelle. Seine Promotion an der Universität Siegen befasst sich mit zukünftigen Arbeitsformen, speziell mit natürlicher sprachiger Mensch-Roboter-Kollaboration in Dienstleistungsbranchen.

Kontakt: d.golchinfar@digitalzentrum-fokus-mensch.de



Darius Hennekeuser

Darius Hennekeuser ist ein Mitarbeiter des Mittelstand-Digital Zentrums Fokus Mensch im Bereich der Digitalen Souveränität, wo er organisatorische Aufgaben im KI-Trainer übernimmt. Zugleich hat er sich auf den Bereich der generativen Sprachmodelle spezialisiert. Diese Expertise bringt er in die Entwicklung von Sprachassistenzsystemen und **Chatbots** für die Projekte des Mittelstand-Digital Zentrum Fokus Mensch ein.

Er promoviert an der Universität Siegen und seine Forschung konzentriert sich auf den Einsatz und die Implementierung von generativen Sprachmodellen. Dieser Bereich ist einer der Schlüssel zum Verständnis und zur Entwicklung von zukunftsweisender KI-gesteuerter Sprachtechnologie. Es ist seine Mission, die Usability von Technologie durch innovative Einsatzmöglichkeiten künstlicher Intelligenz zu verbessern.

Kontakt: d.hennekeuser@digitalzentrum-fokus-mensch.de

Diana Yakimchuk

Diana Yakimchuk ist Expertin im Bereich Marketing und fungiert als Ansprechpartnerin für Öffentlichkeitsarbeit sowie KI-Trainer. Ihre Tätigkeitsschwerpunkte am Mittelstand-Digital Zentrum Fokus Mensch liegen insbesondere im Bereich der Digitalen Souveränität. Frau Yakimchuk hat ihren Master in der Fachrichtung Marketing an der Hochschule Bonn-Rhein-Sieg absolviert. Bereits während ihres Studiums war Frau Yakimchuk als wissenschaftliche Hilfskraft an der HBRS beschäftigt und unterstützte dabei die Projekte zum Thema Sprachassistenten.

Kontakt: d.yakimchuk@digitalzentrum-fokus-mensch.de



Franziska Herrmann

Franziska Herrmann ist Informatikerin und seit Mai 2019 Wissenschaftliche Mitarbeiterin an der NORDAKADEMIE Hochschule der Wirtschaft am Fachbereich Informatik. Seit 2021 ist sie außerdem Mitarbeiterin im Mittelstand-Digital-Netzwerk (Mittelstand Kompetenzzentrum Usability, seit Mai 2023 Mittelstand-Digitalzentrum Fokus Mensch). Ihr Background ist Nutzerzentrierung, Usability und User Experience. Sie ist Botin des agilen Mindsets und hält Workshops und Vorträge dazu. Ihr Anliegen ist es, zielgruppengerecht und niedrigschwellig IT-Kompetenzen zu vermitteln, vor allem auch im Bereich KI.

Kontakt: f.herrmann@digitalzentrum-fokus-mensch.de



Michelle Walther

Michelle Walther befasst sich mit dem Thema Öffentlichkeitsarbeit und Evaluation. Des Weiteren liegt ihr Fokus auf Social Engineering Attacks und betrügerische Kundenrezensionen online und hilft Betrieben im Umgang mit diesen resilienter zu werden. Als wissenschaftliche Mitarbeiterin beschäftigt Michelle Walther sich außerdem mit der Wahrnehmung von Betrugsmöglichkeiten im Bereich Online Shopping aus Nutzerperspektive. Ihren Master hat Frau Walther in der Fachrichtung Psychologie an der Maastricht University Twente in den Niederlanden absolviert.

Kontakt: m.walther@digitalzentrum-fokus-mensch.de



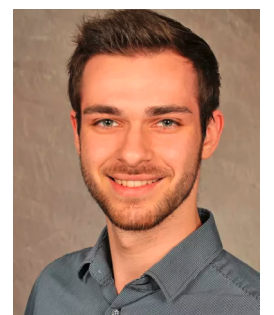
Gastbeitrag

Benedikt Haag (Denkschmiede)



Benedikt Haag ist wissenschaftlicher Mitarbeiter an der Hochschule Bonn-Rhein-Sieg. Er beschäftigt sich seit mehreren Jahren mit der erfolgreichen Anwendung von generativer Künstlicher Intelligenz. Im Drittmittelprojekt Denkschmiede unterstützt er kleine und mittelständische Unternehmen aus der Region Rhein-Sieg dabei, Potenziale von generativer KI zu erkennen und praxisnah umzusetzen.

Kontakt: benedikt.haag@h-brs.de



Impressum

Dieses Dokument wurde als Teil des Mittelstand-Digital Netzwerkes im Rahmen des vom BMWF geförderten Mittelstand-Digital Zentrum Fokus Mensch erstellt.

Das Mittelstand-Digital Netzwerk bietet mit den Mittelstand-Digital Zentren und der Initiative *IT-Sicherheit in der Wirtschaft* umfassende Unterstützung bei der Digitalisierung mit dem Schwerpunkt Künstliche Intelligenz. Das Bundesministerium für Wirtschaft und Energie ermöglicht die kostenfreie Nutzung der Angebote von Mittelstand-Digital. Weitere Informationen finden Sie unter: www.mittelstand-digital.de

Herausgeber

Mittelstand-Digital Zentrum Fokus Mensch in Zusammenarbeit mit der Denkschmiede¹

Stand

Februar 2026

Unter Redaktion von:

Michelle Walther, Benedikt Haag, Diana Yakimchuk

Gender-Hinweis

Die in diesem Kurzlehrbuch verwendeten Personenbezeichnungen beziehen sich immer gleichermaßen auf weibliche und männliche Personen. Es wurde versucht eine Doppelnennung und gegenderte Bezeichnungen umzusetzen. Für den Fall, dass Fälle übersehen wurden, möchten wir uns entschuldigen.

Bildnachweise

Das Titelbild und die Bilder von den Seiten 6, 9, 12, 14, 16, 18, 21, 24, 27, 28, 30, 32, 34, 36 wurden mit Leonardo.Ai erstellt. Prompts von Diana Yakimchuk. Die Bilder von den Seiten 46-47 gehören zu den entsprechenden Autor:innen.

Ergänzend lesenswertes

Mittelstand-Digital Kongress (2025). Selbstbestimmt handeln im digitalen Raum – warum Digitale Souveränität für KMU immer wichtiger wird - Interview mit Prof. Dr. Dennis-Kenji Kipker vom cyberintelligence.institute. [Klicken Sie hier für weitere Informationen.](#)

Mittelstand-Digital Kaiserslautern (2025). KI-Verordnung - Praxisbeispiele zur Orientierung für KMU. [Klicken Sie hier für weitere Informationen.](#)

©2026 Mittelstand-Digital Zentrum Fokus Mensch

¹ Finanziert vom Ministerium für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen & Kofinanziert von der Europäischen Union.



Mittelstand-Digital
Zentrum
Fokus Mensch



www.digitalzentrum-fokus-mensch.de

Gefördert durch:



Bundesministerium
für Wirtschaft
und Energie

Mittelstand-
Digital

aufgrund eines Beschlusses
des Deutschen Bundestages