

LLM-Copilots in ERP-Systemen: Eine Analyse von Nutzungspotenzialen, Risikofaktoren und quantifizierbaren Messmethoden

Harald Konnerth

Die Integration von Large Language Models (LLMs) in Enterprise-Resource-Planning-Systeme (ERP) leitet einen Paradigmenwechsel in der Mensch-Maschine-Interaktion ein. Diese Entwicklung transzendent inkrementelle Funktionserweiterungen und etabliert eine neue Form der Zusammenarbeit, die das Potenzial hat, die Effizienz und strategische Agilität von Organisationen nachhaltig zu steigern. An die Stelle starrer, menügeführter Benutzeroberflächen tritt eine dynamische, konversationelle und kontextsensitive Interaktionsebene. Die als „Copilots“ (z. B. Microsoft Copilot, SAP Joule) bezeichneten Systeme werden strategisch als Werkzeuge zur Erweiterung menschlicher Fähigkeiten positioniert, nicht zu deren Ersatz. Dieser Ansatz, der den Menschen in der Kontrollinstanz beläßt, fördert die Nutzerakzeptanz und mindert organisatorische Widerstände. Die Konsequenz ist eine Verschiebung im Change Management: weg von der Sorge um Arbeitsplatzverluste hin zur Weiterqualifizierung für eine effektive Mensch-Maschine-Kollaboration [1, 2]. Letztlich fungieren LLM-Copilots als Katalysatoren für die organisatorische Weiterentwicklung, indem sie die Kernziele von ERP-Systemen – Prozessoptimierung, Effizienzsteigerung und datengestützte Entscheidungsfindung – auf ein neues Niveau heben.

KI-Copilots gestalten die tägliche Arbeit mit Unternehmenssoftware von Grund auf neu. Sie sind weit mehr als reine Automatisierungshelfer – sie werden zu intelligenten Partnern, die bei komplexen Analysen und strategischen Planungen unterstützen. Ihr Nutzen zeigt sich in drei Kernbereichen:

- **Mehr Effizienz im Arbeitsalltag:**

Ein entscheidender Vorteil ist die Automatisierung von Routineaufgaben. Tätigkeiten wie die manuelle Dateneingabe, das Erstellen von Standardberichten oder die Rechnungsprüfung, die viel Zeit binden,

können von der KI übernommen werden. Das entlastet Fachkräfte spürbar und schafft Freiräume für anspruchsvollere, strategische Aufgaben, die echten Mehrwert schaffen.

- **Datenzugriff für alle:**

Die KI-Assistenten wirken wie ein Übersetzer zwischen Mensch und Maschine und machen komplexe ERP-Daten für jeden im Unternehmen zugänglich. Mitarbeitende müssen keine Systemexperten mehr sein oder SQL beherrschen. Eine einfache Frage in Alltagssprache genügt, zum Beispiel: „Zeige mir die Umsatzentwicklung für Produktgruppe A in der Region Süd“. Der Assistent versteht die Frage, sucht die passenden Daten heraus und präsentiert die Antwort klar und verständlich. Auf diese Weise können Entscheidungen im gesamten Unternehmen auf einer soliden Datengrundlage getroffen werden.

- **Fundiertere Entscheidungen treffen**

Generative KI erweitert die analytischen Fähigkeiten von ERP-Systemen entscheidend. Statt nur in die Vergangenheit zu blicken, ermöglichen

Lesen Sie:

- wie LLM-basierte KI-Copilots die Interaktion mit ERP-Systemen fundamental verändern
- welche Risiken die Einführung von KI-Copilots birgt und wie der Erfolg einer LLM-Copilot-Implementierung gemessen werden kann

die neuen Systeme auch Prognosen, die Simulation von „Was-wäre-wenn“-Szenarien oder das Aufspüren unauffälliger Muster in großen Datensätzen.

- Das markiert einen fundamentalen Wandel: Während frühere KI-Anwendungen wie RPA vor allem darauf abzielten, repetitive Aufgaben zu automatisieren, bieten die neuen Assistenten eine echte kognitive Unterstützung bei komplexen Denk- und Planungsprozessen.

Anwendungsfälle nach Funktionsbereich

Die transformative Wirkung von LLM-Copilots lässt sich anhand konkreter Anwendungsfälle in zentralen Unternehmensbereichen illustrieren:

- Finanzwesen:** Beschleunigung des Procure-to-Pay-Prozesses durch automatisierte Rechnungsverarbeitung. Unterstützung des Controllings durch die Erstellung narrativer Managementberichte mit kontextualisierten Kennzahlen.
- Supply Chain Management (SCM):** Verbesserung der Absatzprognosen durch Analyse externer Marktsignale. Simulation von Szenarien zur vorbeugenden Instandhaltung und Generierung intelligenter Lieferantenempfehlungen.
- Personalwesen (HR):** Automatisierung von Routineanfragen von Mitarbeitenden (z. B. zu Urlaubsansprüchen) und Unterstützung von Führungskräften bei der Erstellung von Stellenbeschreibungen oder Leistungsbeurteilungen.
- Softwareentwicklung und Customizing:** Beschleunigung der ERP-Anpassung durch Generierung von Code-Snippets und Unit-Tests (z. B. in ABAP), was den Entwicklungsaufwand reduziert und die Code-Qualität verbessert.

Architektonische Paradigmen und Implementierungen führender ERP-Anbieter

Die führenden ERP-Anbieter SAP, Microsoft und Oracle gehen bei der Einführung ihrer KI-Copilots jeweils eigene Wege. Ihre Strategien unterscheiden sich vor allem darin, wie tief die Assistenten in die Software integriert sind, auf welche Weise sie auf Unternehmensdaten zugreifen und welche Möglichkeiten zur individuellen Anpassung bestehen. Diese Unterschiede sind für Unternehmen relevant, da sie die Flexibilität und die Abhängigkeit vom jeweiligen Ökosystem beeinflussen.

SAP Joule: Das agentische Paradigma

SAP positioniert Joule als eine „agentische Plattform“. Spezialisierte KI-Agenten können komplexe, mehrstufige Geschäftsprozesse (z. B. Umplanung von Lieferketten) autonom über verschiedene Systeme hinweg orchestrieren. Die Datenanbindung (Grounding) erfolgt mittels Retrieval-Augmented Generation (RAG) auf Basis eines Katalogs von Systemfähigkeiten und Unternehmenswissen, wobei die Berechtigungen der Nutzer strikt eingehalten werden. Die Erweiterbarkeit wird durch die SAP Business Technology Platform (BTP) und das Joule Studio sichergestellt [3].

Microsoft Copilot: Tiefgreifende Ökosystem-Integration

Microsofts Strategie ist die nahtlose Integration des Copiloten in das gesamte Microsoft-365-Ökosystem (Dynamics 365, Word, Teams etc.). Als zentrale Komponente für die Datenanbindung dient der Microsoft Graph, der den gesamten Unternehmenskontext aus E-Mails, Dokumenten, Chats und ERP-Daten aggregiert. Dies ermöglicht hochgradig kontextualisierte Antworten unter Beachtung der bestehenden

Kriterium	SAP Joule	Microsoft Copilot for Dynamics 365	Oracle AI in Fusion Cloud
Architektonisches Kernprinzip	Agentische Prozessorchestrierung	Ökosystem-integrierter Assistent	Eingebettete funktionale Intelligenz
Mechanismus zur Datenanbindung (Grounding)	RAG über Knowledge/ Scenario Catalog	Microsoft Graph & Semantic Index	Native Daten der Fusion-Anwendungen
Framework zur Erweiterbarkeit	Joule Studio (auf BTP)	Copilot Studio (auf Power Platform)	OCI AI/ML Services
Primäre Integrationspunkte	SAP Cloud Portfolio (S/4HANA, SuccessFactors)	Microsoft 365 & Dynamics 365	Oracle Fusion Cloud Applications
Strategischer Differenziator	Autonome, mehrstufige Prozessausführung	Einheitliche User Experience über strukturierte und unstrukturierte Arbeit hinweg	„Touchless“-Automatisierung spezifischer Finanz- und SCM-Aufgaben

Tab. 1: Vergleichende Übersicht der LLM-Copilot-Architekturen..

Nutzerberechtigungen. Die Anpassung erfolgt über das Copilot Studio auf der Power Platform.

Oracle AI: Eingebettete, funktions-spezifische Intelligenz

Oracle fokussiert auf die Einbettung spezialisierter KI-Funktionen direkt in die Oracle Fusion Cloud Applications, um „Touchless Operations“ zu realisieren. Statt eines generalistischen Assistenten setzt Oracle auf spezifische Agenten (z. B. zur Dokumentenverarbeitung). Das Grounding erfolgt nativ auf den Daten innerhalb des Oracle-Ökosystems. Die Oracle Cloud Infrastructure (OCI) bietet zudem Plattformen für Fine-Tuning und die Entwicklung eigener KI-Lösungen [4] (Tabelle 1).

Identifikation und Mitigation inhärenter Risiken

Die Einführung von KI-Copilots in geschäftskritische ERP-Systeme bringt Risiken mit sich, für die klassische Sicherheitskonzepte nicht gemacht sind. Die unberechenbare, sprachbasierte Arbeitsweise der KI verlangt eine grundlegend neue Auseinandersetzung mit dem Thema Sicherheit, um einen verantwortungsvollen Einsatz zu gewährleisten.

Datensicherheit und Datenschutz in RAG-Architekturen

Retrieval-Augmented Generation (RAG) ist eine Schlüsseltechnologie, um KI-Assistenten mit verlässlichen Unternehmensdaten zu verbinden, schafft aber gleichzeitig neue Sicherheitsrisiken. So kann die Speicherung sensibler Informationen in speziellen Vektor-Datenbanken neue Lücken in der Abwehr eröffnen. Ein weiteres Problem: Oft fehlen in diesen Systemen die feingliedrigen Zugriffsberechtigungen, die aus ERP-Systemen bekannt sind. Dadurch könnten Mitarbeitende durch geschickt formulierte Fragen an Informationen gelangen, die sie eigentlich nicht sehen dürften.

Um diese Risiken zu minimieren, sind durchgehende Zugriffskontrollen im gesamten Datenprozess unerlässlich – von der Quelle bis zur finalen Antwort der KI. Wichtig sind zudem technische Vorkehrungen wie die automatische Unkenntlichmachung persönlicher Daten und die strikte Einhaltung gesetzlicher Vorschriften (z. B. Datenspeicherung innerhalb der EU).

KI-Halluzinationen und Informations-integrität

Ein inhärentes Risiko von LLMs ist die Generierung plausibler, aber sachlich falscher Informationen („Halluzinationen“). Im ERP-Kontext können fehlerhafte Finanzberichte oder strategische Empfehlungen gravierende finanzielle und rechtliche Folgen haben.

Der zentrale Lösungsansatz besteht in einem konsequenten Grounding des LLMs an hochwertige, aktuelle Unternehmensdaten mittels RAG. Zusätzlich helfen gezieltes Prompt Engineering und das Fine-Tuning von Modellen auf domänenspezifische Daten. Für alle geschäftskritischen Prozesse bleibt jedoch eine strikte menschliche Überwachung (Human-in-the-Loop) zur Validierung der Ergebnisse unerlässlich.

Prompt-Injection-Angriffe: Eine neue Bedrohungsklasse

Prompt-Injection-Angriffe, bei denen bösartige Anweisungen in Nutzereingaben eingebettet werden, stellen eine neuartige und schwer zu kontrollierende Bedrohung dar (von OWASP als Top-Risiko eingestuft). Ziel ist es, die ursprünglichen Systemanweisungen zu überschreiben und den Copiloten zu unerwünschten Aktionen zu verleiten, beispielsweise zur Preisgabe sensibler Daten oder der Ausführung nicht autorisierter Transaktionen [5].

Zur Risikominderung ist ein mehrschichtiger Verteidigungsansatz notwendig. Zu den Kernstrategien gehören die Validierung von Eingaben, die strikte Trennung von System- und Nutzerdaten sowie der Einsatz von KI-gestützten Filtern (Guardrails). Da eine absolute Sicherheit nicht erreichbar ist, stellt dies eine andauernde Bedrohung dar.

Dies bedeutet einen Paradigmenwechsel für die Sicherheit von ERP-Systemen: Die Angriffsvektoren verschieben sich von codebasierten Lücken hin zur sprachlichen Manipulation der KI. Sicherheitsteams müssen daher Kompetenzen in neuen Feldern wie Adversarial Prompting oder der KI-Verhaltensanalyse aufbauen.

Ein Framework zur Messung von Produktivität und Qualität

Die Bewertung des Erfolgs einer LLM-Copilot-Implementierung erfordert ein strukturiertes und mehrdimensionales Vorgehen. Sich ausschließlich auf einfache Effizienzkennzahlen wie die reine Zeitersparnis zu verlassen, ist unzureichend und kann zu irreführenden Schlussfolgerungen über den tatsächlichen Wertbeitrag führen. Ein ganzheitliches KPI-Framework, das neben der Prozesseffizienz auch die Ergebnisqualität, die Nutzerakzeptanz und den übergeordneten Geschäftswert berücksichtigt, ist für eine fundierte Evaluation unerlässlich. Die Messung muss dabei stets kontextbezogen erfolgen und Faktoren wie die spezifische Aufgabe und das Qualifikationsniveau der Anwender einbeziehen.

Das nachfolgende Gerüst bietet eine strukturierte Grundlage zur Erfolgsmessung von LLM-Copilots im ERP-Umfeld. Es ist als Balanced Scorecard konzipiert,

die es Organisationen ermöglicht, eine datengestützte Governance für ihre KI-Initiativen zu etablieren und

den Return on Investment (ROI) umfassend zu bewerten [6].

1. Prozesseffizienz		
KPI-Name	Definition und Messmethode	Relevanz im ERP-Kontext
Task Completion Time Reduction	Definition: Misst die prozentuale Reduktion der Bearbeitungszeit für einen Prozess (z. B. Auftragserstellung) im Vergleich zur manuellen Methode. Messmethode: A/B-Tests mit Kontrollgruppen.	Quantifiziert direkt die operative Geschwindigkeitssteigerung.
Manual Effort Reduction	Definition: Ermittelt die Verringerung der Anzahl von Klicks und Dateneingaben pro Aufgabe. Messmethode: Analyse mittels Process Mining und Aufzeichnung von Benutzersitzungen [7].	Quantifiziert die Reduktion monotoner Arbeit und setzt Kapazitäten für anspruchsvollere Tätigkeiten frei.

Tab. 2: KPI-Gerüst zur Erfolgsmessung von LLM-Copilots im ERP-Umfeld: Prozess-Effizienz.

2. Ergebnisqualität		
KPI-Name	Definition und Messmethode	Relevanz im ERP-Kontext
Error Rate Reduction	Definition: Reduktion menschlicher Fehler (z. B. bei Dateneingabe) durch die Nutzung des Copiloten. Messmethode: Audits der Prozessergebnisse und Vergleich der Fehlerquoten mit einer Baseline.	Misst die Verbesserung von Genauigkeit und Compliance, was besonders in regulierten Bereichen wie dem Finanzwesen wichtig ist.
Factual Consistency Score (Faithfulness)	Definition: Grad der Übereinstimmung der KI-Antwort mit den Quelldaten (Ground Truth). Messmethode: Kombination aus automatisierten Vergleichen und menschlicher Bewertung.	Dient der Messung der Systemzuverlässigkeit und hilft, das Risiko von KI-Halluzinationen zu kontrollieren.
Relevance Score	Definition: Das Ausmaß, in dem die KI-Ausgabe die Absicht und den Kontext der Nutzeranfrage trifft. Messmethode: Menschliche Bewertung (z. B. mittels Likert-Skalen) und semantische Analyse.	Stellt sicher, dass die Unterstützung nicht nur korrekt, sondern auch kontextbezogen und hilfreich ist.

Tab. 3: KPI-Gerüst zur Erfolgsmessung von LLM-Copilots im ERP-Umfeld: Ergebnisqualität.

3. Nutzerakzeptanz und -kompetenz		
KPI-Name	Definition und Messmethode	Relevanz im ERP-Kontext
Adoption Rate	Definition: Anteil der Mitarbeitenden, die den Copiloten regelmäßig (z. B. wöchentlich) aktiv nutzen. Messmethode: Analyse von System-Nutzungsprotokollen.	Eine fundamentale Kennzahl, die zeigt, ob das Werkzeug als wertvoll wahrgenommen wird. Eine niedrige Rate deutet auf Usability-Probleme hin.
Task Success Rate	Definition: Anteil der Aufgaben, die Nutzer erfolgreich mit dem Copiloten abschließen, ohne auf manuelle Methoden zurückzugreifen. Messmethode: Analyse von User-Interaction-Logs und Nutzerbefragungen.	Misst die Effektivität und Benutzerfreundlichkeit der Mensch-KI-Kollaboration.
Cognitive Load Reduction	Definition: Maß für die Verringerung der mentalen Anstrengung bei der Bewältigung komplexer Aufgaben. Messmethode: Standardisierte Fragebögen (z. B. NASA-TLX) und qualitative Interviews [8].	Erfasst den Wert der KI als kognitiver Partner, der Komplexität reduziert und die Nutzerzufriedenheit steigert.

Tab. 4: KPI-Gerüst zur Erfolgsmessung von LLM-Copilots im ERP-Umfeld: Nutzerakzeptanz und -kompetenz.

4. Geschäftswert		
KPI-Name	Definition und Messmethode	Relevanz im ERP-Kontext
Decision-Making Velocity	Definition: Misst die Verkürzung der Zeitspanne von einer geschäftlichen Fragestellung bis zu einer fundierten, datenbasierten Entscheidung mithilfe des Copiloten.	Eine fundamentale Kennzahl, die zeigt, ob das Werkzeug als wertvoll wahrgenommen wird. Eine niedrige Rate deutet auf Usability-Probleme hin.
Return on Investment (ROI)	Definition: Stellt den finanziellen Nutzen (z. B. Kosteneinsparungen, Mehrwert) den Gesamtkosten der Implementierung (Lizenzen, Aufwand, Schulungen) gegenüber. Messmethode: Finanzielle Modellierung, bei der operative KPIs in monetäre Werte übersetzt und den Kosten gegenübergestellt werden.	Die entscheidende Kennzahl zur finanziellen Beurteilung und Rechtfertigung der Investition auf Führungsebene.

Tab. 5: KPI-Gerüst zur Erfolgsmessung von LLM-Copilots im ERP-Umfeld: Geschäftswert.

Zusammenfassung und Ausblick

Die Art der Interaktion mit ERP-Systemen steht vor einem Umbruch. Durch KI-Assistenten entwickelt sich die reine Dateneingabe zu einer echten Kollaboration zwischen Mensch und Technologie. Der Mehrwert dieser Entwicklung geht weit über reine Effizienz hinaus; die Systeme werden zu kognitiven Partnern, die bei komplexen Entscheidungen unterstützen und damit den strategischen Beitrag der ERP-Software neu definieren.

Die Anwenderunternehmen stellen sich der Herausforderung, die Chancen der neuen Technologie zu nutzen, ohne dabei die entstehenden Risiken zu ignorieren. Besonders wichtig ist, Risiken wie Datenlecks, Falschinformationen durch die KI oder neue Angriffsarten fest im Griff zu haben. Ob die Einführung am Ende erfolgreich ist, zeigt sich nur, wenn der Nutzen klar messbar ist – von besseren Abläufen bis zum handfesten Geschäftsvorteil.

Zukünftig werden diese intelligenten Systeme noch autonomer agieren und ganze Prozessketten steuern. Diese Entwicklung wird die Rolle von Fachexperten

nachhaltig prägen: Ihre Aufgabe wird weniger die Ausführung, sondern vielmehr die strategische Steuerung, Überwachung und Qualitätssicherung der KI-gestützten Prozesse sein. Die Differenzierung zwischen Werkzeug und Partner verschwindet. Diese Entwicklung zwingt Unternehmen, sich schon heute mit den ethischen und strategischen Fragen von morgen auseinanderzusetzen.

Literatur

- [1] SAP. (2024). Joule Copilot from SAP. Abgerufen von <https://www.sap.com/products/artificial-intelligence/ai-assistant.html>.
- [2] Microsoft. (2024). Copilot and generative AI in Dynamics 365. Microsoft Learn. Abgerufen von <https://learn.microsoft.com/en-us/dynamics365/copilot>.
- [3] Lewis, P., Perez, E., Piktus, A., Karpukhin, V., Goyal, N., Küttler, H., ... Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.
- [4] Oracle. (2024). Oracle AI for Fusion Applications – Get Started. Oracle Help Center. Abgerufen von <https://docs.oracle.com/en/cloud/saas/fusion-ai/>
- [5] OWASP Foundation. (2023). OWASP Top 10 for Large Language Model Applications. Abgerufen von <https://owasp.org/www-project-top-10-for-large-language-model-applications/>
- [6] Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard – Measures That Drive Performance. Harvard Business Review, 70(1), 71–79.
- [7] van der Aalst, W. M. P. (2016). Process Mining: Data Science in Action. Springer.
- [8] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139–183). North-Holland.

Stichwörter:

ERP-Systeme, LLM-Copilots, Generative KI, Prompt-Injection. Messmethoden für Produktivität und Qualität



Als Professor für Prozesse und Systeme an der FH Oberösterreich in Steyr liegen die Schwerpunkte von Harald Konnerth auf betrieblichen Informationssystemen und Datenmanagement. Diese Expertise setzt er auch als Geschäftsführer praktisch um, indem er Unternehmen bei der Implementierung von ERP-Systemen und bei der digitalen Prozessautomatisierung unterstützt.

DI (FH) Harald Konnerth MBA
Upper Austria University of Applied Sciences
FH OÖ Studienbetriebs GmbH
Wehrgrabengasse 1-3, 4400 Steyr, Österreich
Tel.: +43 664 2268190
E-Mail: harald.konnerth@fh-steyr.at