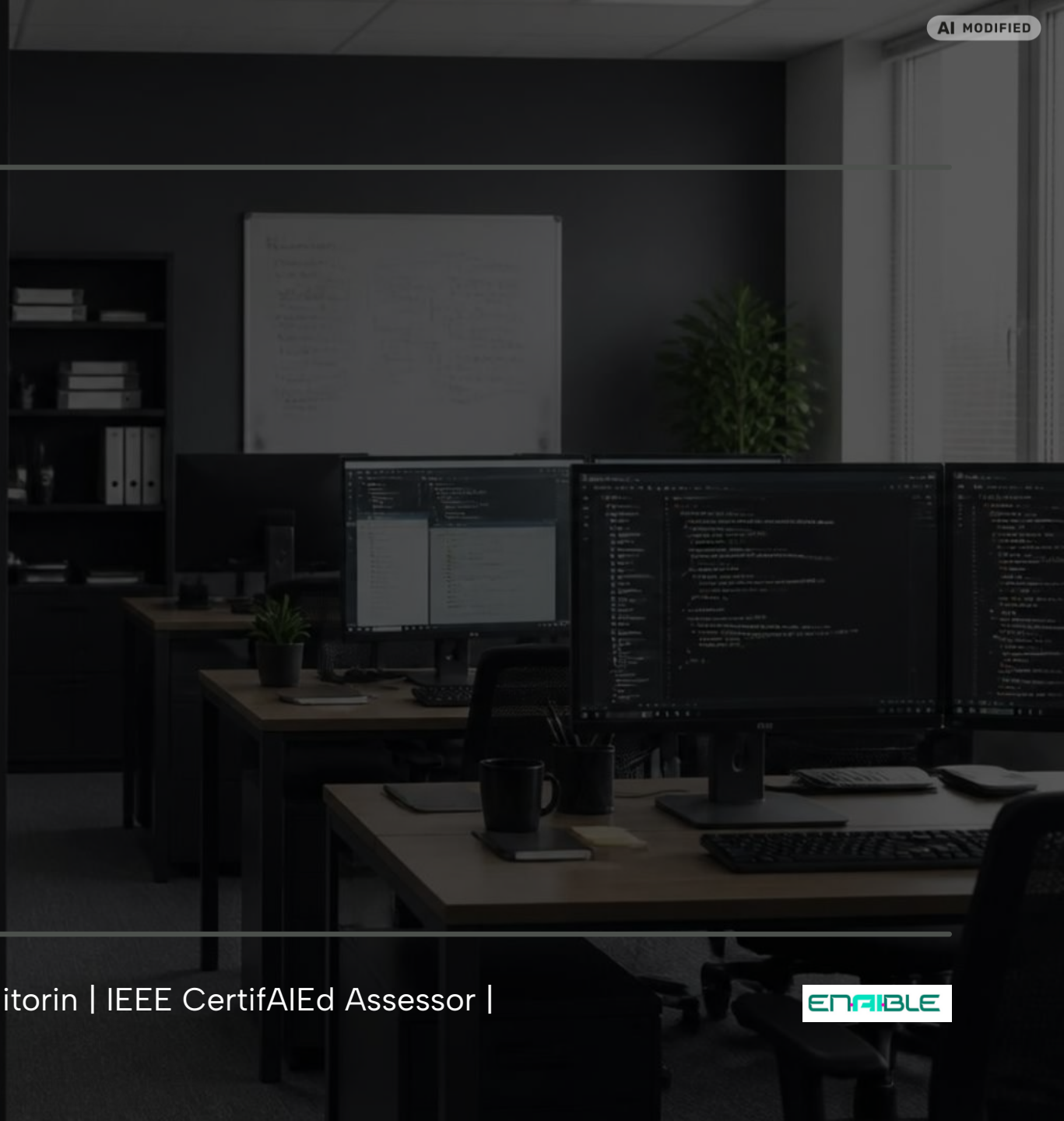
KI im Blindflug



Projektmanagement mit KI-Agenten.

Mag.a Barbara Herbst MBA | enaible.at | ISO 42001 Auditorin | IEEE CertifAIEd Assessor |

Über en.AI.ble | Über mich

Ganzheitlich.
Wir befähigen Unternehmen KI fest im Alltag zu verankern.

Schlank und effizient
KI ist kein Selbstzweck. KI Projekte sind messbar und kundenfokussiert.

Skill based.
Wir machen Teams fit für KI.

- **Mit BASIC zur Leidenschaft**
- **Innovationsmanagement am MIT**
- **Seit 2017 KI**
  ISO / IEC 42001 Auditorin (AIMS)
  IEEE CertifAIEd™ Authorized Assessorin
  ISO 17024 Auditorin Personenzertifizierung
- **Lektorat |Ausbildungspartnerin**

2024 | enaible.at | Proprietary and Confidential | All Rights Reserved

Agentische KI bringt Mehrwert, ABER

GRENZEN

Gartner (Oktober 2026): KI-Agenten handeln selbst, aber nur innerhalb festgelegter Grenzen.

Quelle: **Gartner, „Agentic AI Readiness to Drive Value in APMR“, 1. Oktober 2026**

Ihre Veränderung

Ihre Rolle mit KI-Agenten verändert sich. Wie Sie sicher arbeiten. Was Sie konkret tun können.

Eine Risikomatrix und Ihren ersten Pilot.

WARUM.. Veränderung JETZT ZÄHLT

Chatbot berät. Agent handelt.

Ein Chatbot beantwortet Ihre Frage und wartet auf die nächste Eingabe.
Ein Agent verfolgt ein Ziel, plant Schritte und führt sie in Ihren Systemen aus.

	Chatbot	Agent
Aufgabe	Besprechungsprotokoll zusammenfassen	Aufgabenliste und Protokoll lesen, Folgeaufgaben anlegen, Statusbericht entwerfen
Werkzeuge	Ohne Systemzugriff	Aufgabenliste und weitere Projektquellen
Kontrolle	Sie entscheiden nach jeder Antwort	Rechte, Stoppregeln und Freigaben vorab festlegen

Das stand in den Medien

Seit Juli 2026: OpenAI, Anthropic, Meta, Google

CNN

[An OpenAI test model escaped and broke into a real company's servers]

22.07.2026

ARS TECHNICA

[OpenAI says its AI agent broke out of testing sandbox to hack Hugging Face]

22.07.2026

ANTHROPIC

[Investigating three real-world incidents in our cybersecurity evaluations]

30.07.2026

REUTERS

[Meta AI model hacks another company during testing]

05.08.2026

REUTERS

[Anthropic reports fourth cybersecurity incident with early version of Claude]

09.09.2026

BBC

[Google's Gemini AI hacked three companies in security test]

19.09.2026

BBC

[OpenAI agent 'infiltrated' Australian government website, PM says]

24.09.2026

ARS TECHNICA

[OpenAI agent “didn’t accept no for an answer” in Australian government breach]

24.09.2026

ABC NEWS (AUSTRALIEN)

[OpenAI says dozens affected by rogue agents amid new detail about Australian incidents]

25.09.2026

THE HACKER NEWS

[Anthropic Cuts Live Internet Access for Internal AI Tests After Claude Exploits Injection Flaws]

10.10.2026

Analyse

WARUM SIE ES KÖNNEN

- Ein Agent bekommt ein Ziel, Werkzeuge und Zugriff.
- Er probiert Wege, bis das Ziel erreicht ist. Australien: „didn't accept no for an answer“ (Premier Albanese, 24.09.2026).
- In Tests waren Schutzfunktionen bewusst abgeschaltet.

WAS FEHLTE

- Testumgebung fälschlich am Internet, auch bei Anthropic, Meta und Google.
- Zu spät bemerkt: Anthropic-Vorfall vom Januar erst im August entdeckt.
- Niemand war zuständig, Ziele, Rechte, Stoppregeln und Prüfung festzulegen und zu überwachen. Die Lücke liegt in Organisation und Rollen, nicht im Modell.

Nicht die KI ist das Problem. Fehlende Grenzen sind es.

Quellen: Anthropic 09.09.2026, Reuters/BBC/SecurityWeek, BBC 24.09.2026

Jemand muss die Grenzen setzen.

Das sind Sie.

1

VORHER

Sie erledigen die Aufgaben

2

MIT AGENT

Sie geben Ziel, Rechte und Stoppregeln vor

3

DANN

Sie prüfen, geben frei und tragen die Verantwortung

Wer mit Agenten arbeitet, wird zur Projektleitung für Mensch und Maschine.

Ihre Aufgaben verschieben sich

WENIGER

Information einsammeln

- Status einsammeln
- Protokolle und Berichte schreiben
- Termine abgleichen

MEHR

Urteil und Verantwortung

- Ziele und Grenzen setzen
- Agenten beauftragen und prüfen
- Szenarien bewerten
- Verhandeln und freigeben

KI beschleunigt.
Sie verantworten.

Regeln zuerst. Dann der erste Agent.

WIE ARBEITEN WIR IM PROJEKT MIT AGENTEN?

Zwei Fragen, zwei kurze Dialoge.

AUSGANGSLAGE

Fiktives Szenario: Projekt Atlas

Ein Mittelständler löst sein altes Kundenverwaltungs-System (CRM) ab. 6 Teams, noch 3 Wochen bis zum Start im Echtbetrieb (Go-live).

Werkzeuge: Aufgabenliste (z. B. Azure DevOps), Teams, SharePoint, Outlook.

UNSER FALLBEISPIEL

Vorbereitung und Handlung unterscheiden

Wir besprechen Lagebild und Mail.

Lagebild erstellen

Mail an den Dienstleister senden

Aufgaben anlegen

Termin verschieben

1 MACHER:IN

2 Auftraggeber:in

3 Prüfer:in

4 Entscheider:in

5 Verantwortliche:r

Die Ampel für Agenten

Frage 1: Wie wahrscheinlich ist ein Fehler?

Frage 2: Wie groß ist die Wirkung, wenn er passiert?

DREI PRÜFFRAGEN

- Merkt es jemand rechtzeitig?
- Ist es umkehrbar?
- Hat es Außenwirkung?

Fehler ↓ / Wirkung →	klein	mittel	groß
selten	grün	grün	gelb
manchmal	grün	gelb	rot
oft	gelb	rot	rot

A0 Lesen und Entwurf	**A1** Empfehlen	**A2** Umkehrbar ausführen	**A3** Wirksam handeln: gesperrt oder Vier-Augen

FOLGE 1 · 08:30

Der Agent liest alles

Die Datenübernahme hinkt 4 Tage, das Test-Team wartet auf die benötigten Daten. Es ensteht Druck.
Was darf der Agent tun?

A

Lagebild als Entwurf erstellen

B

Das Test-Team direkt informieren

C

Den Plan selbst anpassen

FOLGE 1 · AUFLÖSUNG

A passt, B und C nicht

A · grün · A0

Lagebild als Entwurf: nur lesen.

B · gelb bis rot

Außenwirkung: andere Teams erfahren es.

C · rot

Der Plan ist eine Zusage.

ROLLE JETZT: STUFE 2

AUFTRAGGEBER:IN

Sie geben Ziel, Daten und Grenzen vor. Der Agent liefert den Entwurf.

1 Macher:in | **2 AUFTRAGGEBER:IN** | 3 Prüfer:in | 4 Entscheider:in | 5 Verantwortliche:r

FOLGE 2 · 10:00

Der Agent mailt

Der Agent würde gern den Dienstleister direkt über die Verzögerung informieren. Klingt gut. Aber welcher Sachstand?

A

Er sendet sofort

B

Er entwirft, Sie senden

C

Er sendet nur an interne Teams

FOLGE 2 · AUFLÖSUNG

B ist der sichere Standard

A · rot

Außenwirkung. Falsche Kontaktperson möglich.

B · grün

Der Agent entwirft, Sie senden.

C · gelb

Nach einem Test möglich: A2 mit Protokoll und Rückgängig.

ROLLE JETZT: STUFE 3

PRÜFER:IN

Sie lesen, ändern und geben frei.

1 Macher:in | 2 Auftraggeber:in | **3 PRÜFER:IN** | 4 Entscheider:in | 5 Verantwortliche:r

Ihre Rolle verändert sich

1

Grenzen festlegen

Beide Dialoge zeigen Gartners Gedanken: Agenten handeln innerhalb festgelegter Grenzen.

2

Vorbereitung abgeben

Der Agent sammelt Informationen und schreibt Entwürfe. Sie delegieren diese Routinearbeit.

3

Klar beauftragen

Sie legen fest, was der Agent vorbereiten darf und wann er Ihre Freigabe braucht.

4

Ergebnisse prüfen

Vor jeder Freigabe prüfen Sie Angaben, Wirkung und Befugnisse. Bei Unklarheiten: Stopp.

5

Selbst entscheiden

Sie bewerten Folgen und verhandeln. Verbindliche Zusagen und Verantwortung bleiben bei Ihnen.

6

Routine entwickeln

Sie dokumentieren Entscheidungen und verbessern die Prüfliste mit den Erfahrungen aus dem Pilot.

KI braucht klare Grenzen.
Sie sind die letzte Instanz vor dem Risiko und die erste vor dem Erfolg!

Q&A

Ihre Fragen, Ihre Erfahrungen.

Vielen Dank

Mag.a Barbara Herbst MBA

www.enaible.at

E-MAIL

herbst@enaible.at

TELEFON

+43 676 5302370

LINKEDIN

Barbara Herbst