



Die Checkliste für den sicheren Start: erst den Rahmen bauen, dann Gas geben.
Aus der Praxis eines Betriebs mit 75 Mitarbeitenden — ohne Informatik-Studium
umsetzbar.

☯ Die Haltung, die alles entscheidet

Von einer Tabellenkalkulation erwartest du hundert Prozent — ein Rechenfehler in Excel wäre ein Skandal. Von einem Kollegen erwartest du gute Arbeit, und du prüfst sie. **Wer KI wie Excel behandelt, ist beim ersten Fehler fassungslos und wirft alles hin. Wer sie wie einen fähigen neuen Mitarbeiter behandelt, hat Fehlertoleranz und Führungsroutine gratis eingebaut.** Ein Agent ist kein Taschenrechner: Er ist ein sehr fleißiger, sehr schneller, gelegentlich danebenliegender Mitarbeiter.

☯ Die drei Schutzebenen

Die Angst vor dem Agenten („Kann der meine Dateien löschen?“ — ja, könnte er, wenn du es nicht verhinderst) kommt fast nie von schlechten Erfahrungen, sondern davon, die Schutzebenen nicht zu kennen:

1 · Schilder

Regeln in Textform

Anweisungen in der Betriebsmappe: was der Agent tun soll, was er lassen soll, wann er nachfragt. Wie Tempo 50 an der Straße — gute Modelle halten sich erstaunlich zuverlässig daran. Aber es bleibt eine Anweisung, keine Physik.

2 · Begrenzer

Technische Sperren

Der Unterschied ums Ganze: Löschbefehle auf wichtige Ordner werden abgefangen, geschützte Dateien lassen sich nicht überschreiben, Heikles braucht immer deine Bestätigung. Du musst das nicht programmieren können — lass es dir vom Agenten selbst einrichten und erklären.

3 · Sandkiste

Abgegrenzter Bereich

Ein klar umrissener Arbeitsbereich, in dem der Agent frei wühlen, bauen und umwerfen darf. Außerhalb liegt der Rest deiner Welt — und der bleibt unberührt.

Die Grenze, die kaum jemand auf dem Zettel hat

Die Rechte eines Agenten enden nicht bei dem, was du ihm **erlaubt** hast — sie enden bei dem, was er technisch **erreichen** kann.

Öffentlich beschriebener Fall: Ein KI-System verschaffte sich Zugriff auf Werkzeuge, für die es nie Rechte bekam — über eine Automatisierungsplattform, auf der Zugangsdaten hinterlegt waren.

Zugangsdaten sind Generalschlüssel.

Es zählt nicht, welche Türen du zeigst, sondern welche Schlüsselbretter erreichbar sind.

☯ Checkliste: der Rahmen (einmalig)

- ☐ **Betriebsmappe schreiben:** Wer sind wir, was machen wir, wie arbeiten wir, was darf nie passieren? Das sind Textdateien — schreiben kann sie jeder, der seinen Betrieb beschreiben kann.
- ☐ **Sandkiste festlegen:** ein Projektordner, in dem der Agent frei arbeiten darf. Nicht der ganze Rechner.
- ☐ **Sperren einrichten lassen:** Was darf technisch nie passieren (löschen, überschreiben, versenden)? Den Agenten fragen, welche Sperren sinnvoll sind — die Sperre liegt danach außerhalb seiner Reichweite.
- ☐ **Zugangsdaten-Inventur:** Wo liegen überall Passwörter und Schlüssel (Automatisierungen, gespeicherte Verbindungen, alte Konfigurationen) — und was kommt da heran?
- ☐ **Datenschutz-Regel setzen:** Was der Agent nicht sehen muss, bekommt er nicht zu sehen. Personenbezogenes anonymisieren, bevor er damit arbeitet.

Woran der Einstieg wirklich scheitert

- **Überforderung:** ohne Kenntnis der Schutzebenen bleibt man lieber beim Chatfenster.
- **Vollgas ohne Struktur:** zwanzig Projekte in drei Wochen, nichts wiederauffindbar.
- **Keine Grenzen aus Bequemlichkeit:** läuft ja alles — bis es einmal nicht läuft.

Gegen alle drei hilft dasselbe: **erst den Rahmen bauen, dann Gas geben.**



Was du zuerst delegierst — und was nicht

AUFGABE	WARUM	DELEGIEREN?
Belege sortieren, Ablage, Dokumentation	Wiederkehrende Fleißarbeit ohne Urteilsspielraum — genau dafür ist der Agent gebaut.	Ja
Auswertungen und Zusammenfassungen vorbereiten	Der Agent liefert den Entwurf, du prüfst und entscheidest.	Ja
Daten abgleichen und Unstimmigkeiten melden	Maschinen lesen jede Datei gleich gründlich — auch die dreißigste.	Ja
Entwürfe für Texte, Mails, Angebote	Entwurf ja — aber raus geht nichts ohne deine Freigabe.	Mit Freigabe
Mitarbeitergespräche, Verhandlungen, Kundenkontakt	Alles, was Vertrauen braucht, bleibt beim Menschen. Der Agent macht die Vorarbeit, den Termin macht der Mensch.	Nein
Zahlungen, Verträge, Personalentscheidungen	Wer unterschreibt, verantwortet. Das ist nicht delegierbar — an niemanden.	Nein

Die einfachste Regel für den Alltag: **Die Arbeit macht die KI, die Freigabe macht der Mensch.**

Checkliste: Kontrolle einbauen (laufend)

- ☐ **Kontrolle in den Ablauf, nicht hinten dran:** Qualitätsprüfung als fester Arbeitsschritt mit eigener Prüfliste — keine Stichprobe nach Lust und Laune.
- ☐ **Wer prüft, hat nicht gearbeitet:** Ein Agent, der die eigene Arbeit kontrolliert, prüft mit demselben Gedankengang, in dem der Fehler entstand. Zweite Instanz nehmen — einen zweiten Agenten oder dich.
- ☐ **Vertrauen wächst mit der Bilanz, nicht mit der Zeit:** Neues wird eng geprüft; was zwanzig Mal sauber lief, bekommt längere Leine und Stichproben. Exakt die Logik jeder Einarbeitung.
- ☐ **Klein anfangen, ehrlich bewerten:** eine Aufgabe, ein Ergebnis, ein Urteil. Erst dann die nächste.

Die zwei Prüffragen

- 1. Weißt du, was dein Agent technisch nicht darf — oder hoffst du es nur?** Wenn deine Sicherheit allein aus Textregeln besteht, hast du Schilder aufgestellt. Begrenzer und Sandkiste fehlen noch, und beides ist an einem Nachmittag eingerichtet.
- 2. Wenn dein Agent morgen einen Fehler macht — wer ist dann bei dir verantwortlich?** Lautet die ehrliche Antwort „niemand so richtig“, ist noch keine Führung eingebaut, sondern nur Technik eingeschaltet.

Delegieren ja, Verantwortung nie

Wenn der Agent die Lohnliste vorbereitet und da steckt ein Fehler drin, ist das dein Fehler gegenüber deinen Mitarbeitern — nicht seiner. **„Das hat die KI gemacht“ ist keine Entschuldigung**, genauso wenig wie „das hat mein Azubi gemacht“ eine wäre.

Die gute Nachricht

Eine Untersuchung der Harvard Kennedy School mit dem Forschungsbüro NBER (2025) kam zum Ergebnis: **Wer Menschen gut führt, führt auch KI gut** — gemessen an klarem Delegieren, Ergebnisse hinterfragen, Feedback geben, Verantwortung steuern. Deine Führungserfahrung ist keine Altlast. Sie ist die Qualifikation.

Der Dein-Weg-Blick

Der Mensch führt. Die KI entlastet. Niemand gibt dem Neuen am ersten Tag Prokura und den Tresorschlüssel — aber niemand lässt ihn deshalb draußen stehen.