

Workshop-Daten

Agent / Use Case

Datum

Moderation

1. Zweck des Agenten

Welches konkrete Problem soll der Agent loesen? Was ist ausdruecklich nicht sein Auftrag?

2. Werkzeuge und Systemzugriffe

Welche Tools, APIs und Zielsysteme sind zwingend notwendig? Welche bleiben gesperrt?

3. Datenklassifizierung

Welche Daten werden gelesen, verarbeitet oder veraendert? Wie sensibel sind sie?

4. Erlaubte Autonomiestufe

Stufe 1 Information, 2 Assistenz, 3 begrenzte Ausfuehrung oder 4 Multi-Agenten-Betrieb?

Kontrollen und Verantwortung

Formulieren Sie pruefbare Regeln, benannte Owner und einen konkreten Stop-Pfad.

5. Fehlergrenze und Stop-Regel

Welche Abweichung stoppt den Agenten automatisch? Wie funktioniert der Kill-Switch?

6. Eskalations- und Freigabepfad

Wer prueft oder gibt frei? In welcher Frist und ueber welchen dokumentierten Kanal?

7. Kennzahlen und Monitoring

Welche Kennzahl zeigt sichere Wirkung statt blosser Aktivitaet? Wer kontrolliert sie wie oft?

8. Verantwortlichkeit und Owner

Wer traegt fachliche, technische sowie Datenschutz- und Sicherheitsverantwortung?

Abschlussfrage: Wer bemerkt eine Fehlentscheidung zuerst - und wie schnell kann der Agent gestoppt werden?