# Kontrolle von kundenorientierten AI-Agenten HITL per Default

**Human-in-the-loop (HITL) *Default**

**Human-on-the-loop (HOTL)**

**Human-out-the-loop (HOOTL)**

## Setup «Human-in-the-loop»

Notwendige Kenntnisse «Human»

- Fach (Plausibilität)
- Technologie (z.B. LLM-Spezifische Fehler)
- Regulierung (z.B. Datenschutz, InfoSec, KI)
- Digital Ethik

Anforderungen an den AI-Agenten

- Nachvollziehbarkeit (z.B. durch Protokollierung aller durch AI-Agenten ausgeführten Schritte )

Organisatorische Massnahmen

- Accountability klar regeln
- Was wann zu tun ist (Action Plan)

## Kriterien AI-Agenten für weniger Kontrolle

Je höher das Risiko, desto kleiner der Spielraum für den AI-Agenten. Umgekehrt gilt: Steigt die Qualität, darf er mit mehr Autonomie arbeiten.

Risiko

- Schaden / Fairness
- Datenqualität
- Komplexität* / Freiheitsgrade

Qualität

- Zuverlässigkeit / Performance (Human als Baseline)
- Subjektives Vertrauen
- Use-Case alleine nicht entscheidend

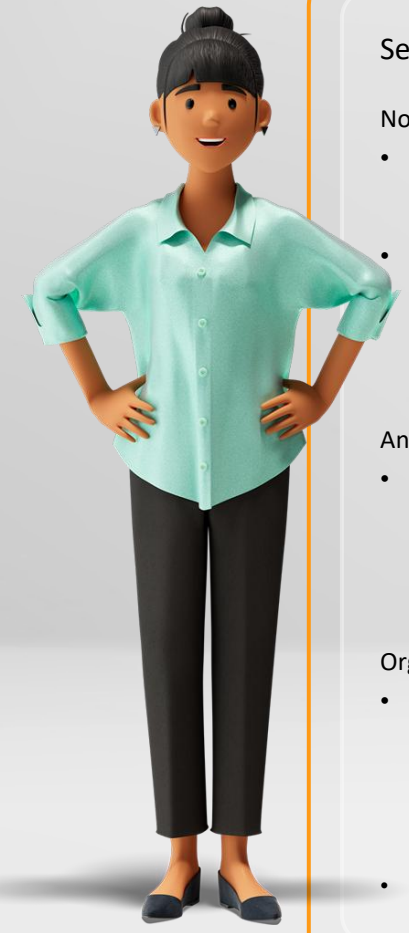
## Anforderungen aus Kundensicht

Must-have

- Möglichkeit für Einsprache Kundin
- Transparenz AI-Agent Einsatz

* Vlg.: https://arxiv.org/pdf/2502.02649 4 Agentic levels auf Seite 3

# Der AI Agent arbeitet, aber ein Mensch ist per Default immer dabei



## Setup «Human-in-the-loop»

### Notwendige Kenntnisse «Human»

- Fach
  Vicky bringt schon zwei Jahre Erfahrung als Service-Mitarbeiterin im Kundendienst mit.
- Technologie
  Sie hält sich durch regelmässige Trainings zu LLMs und AI Agents auf dem Laufenden. Besonders wichtig ist ihr, dass Datenschutz, EU AI Act etc. und digitale Ethik immer im Vordergrund stehen.

### Anforderungen an den AI-Agenten

- Nachvollziehbarkeit
  Der Agent muss seine Schritte so dokumentieren, dass sie im CRM jederzeit nachverfolgt werden können – Transparenz ist hier das A und O.

### Organisatorische Massnahmen

- Accountability
  Die Verantwortung bleibt klar verteilt: Entscheidungen trifft am Ende immer der Mensch. Bei HOOTL liegt die Verantwortung zusätzlich bei der Serviceleiterin – klare Regeln sorgen für Orientierung.
- Der Action Plan regelt, welches Team sich bei einem Vorfall darum kümmert.

## Kriterien AI-Agenten für weniger Kontrolle

### Risiko

- Schaden / Fairness
  Es wird geprüft, wie hoch das Risiko für die verschiedenen Arten von Kundenanfragen bei Fehlverhalten ist. Bei niedrigem Risiko ist weniger Kontrolle nötig.

### Qualität

- Zuverlässigkeit
  Entscheidend ist, wie oft der Service zur vollen Zufriedenheit der Kund:innen erledigt wird. Bei regelmässiger, grosser Zufriedenheit, braucht es weniger Kontrolle.
- Subjektives Vertrauen
  Damit AI Agents mehr Verantwortung übernehmen können, muss bewusst Vertrauen aufgebaut werden – durch Tests, klare Prozesse und transparente Ergebnisse.
- Beispiel Use Case
  Ein KI-Agent erkennt im Kundenservice eingehende E-Mail-Anfragen, bearbeitet diese automatisch und beantwortet sie selbstständig.

Beispiel Kundenservice

## Anforderungen aus Kundensicht

### Must-have

- Beim Eingang ihres Anliegens per E-Mail wird Lena informiert, dass ein automatisiertes System zur Bearbeitung eingesetzt wird. Sie hat jederzeit die Möglichkeit, dies abzulehnen – dann übernimmt ein Mensch den Fall.

## Beispiel Use Case

### Kundenservice

- Ein KI-Agent erkennt eingehende E-Mail-Anfragen, bearbeitet diese automatisch und beantwortet sie selbstständig.