

Human-in-the-Loop

Menschen prüfen, korrigieren und steuern, bevor etwas wirksam wird

Nicht jeder KI-Prozess braucht einen Menschen in der Schleife. Aber sobald eine Entscheidung spürbare Folgen für Personen, Geld oder die Rechtslage hat, ist die Vollautomatisierung ein Risiko. Die Kunst liegt nicht darin, alles zu kontrollieren, sondern die richtigen Stellen zu kontrollieren.

DREI AUFGABEN, DIE DER MENSCH ÜBERNIMMT

PRÜFEN

Ist das Ergebnis plausibel, sachlich richtig, frei von systematischer Schieflage?

KORRIGIEREN

Fehlenden Zusammenhang ergänzen und die Entscheidung an den Einzelfall anpassen.

STEUERN

Regeln setzen, Grenzen ziehen und im Zweifel gegen die Empfehlung entscheiden.

WARUM AUFSICHT ÜBERHAUPT NÖTIG IST

- **Ausnahmen und Grauzonen.** Modelle sind stark im Regelfall und schwach genau dort, wo ein Fall vom Muster abweicht.
- **Verzerrung.** Aus historischen Daten gelernte Schieflagen wirken weiter, etwa bei Kredit und Bewerbervorauswahl.
- **Verantwortung.** Ein Modell kann rechtlich und ethisch nicht haften. Ein Mensch schon.

WANN EINE FREIGABE ZWINGEND IST

- Wenn es Einzelpersonen spürbar trifft: Kündigung, Kreditabsage, Therapie
- Wenn eine rechtliche oder vertragliche Bindung entsteht
- Wenn eine ethische Abwägung nötig wird
- Wenn ein Reputationsschaden wahrscheinlich ist

Sonst genügen Stichproben. Interne Zusammenfassungen, Ideensammlungen und unverbindliche Entwürfe brauchen keine Freigabeschleife.

WAS AM KONZEPT SELBST SCHEITERN KANN

- **Kosten.** Wenn am Ende jeder Fall von Hand validiert wird, verpufft der Effizienzgewinn.
- **Flaschenhals.** Je engmaschiger die Freigabe, desto schwerer skaliert der Prozess.
- **Blindes Durchwinken.** Unter Zeitdruck werden lange KI-Texte nur überflogen. Der Mensch in der Schleife ist dann nur noch Dekoration.
- **Datenschutz.** Wer manuell sichtet, sieht oft mehr als für die Aufgabe nötig. Es braucht ein Rollenkonzept.

MERKSÄTZE

Der AI Act verlangt für Hochrisiko-Systeme, dass Menschen sie überblicken und jederzeit eingreifen können. Human-in-the-Loop ist die praktische Umsetzung davon.

Ein Mensch in der Schleife ist keine Qualitätsgarantie. Ohne Zeit und Checkliste winkt er durch.

Die Maschine liefert Wahrscheinlichkeiten. Haftung übernimmt sie keine.

SO SIEHT DAS KONKRET AUS

VERTRAGSENTWURF

Die KI erstellt den Entwurf aus Vorlage und Stichworten. Er geht mit einer festen Checkliste an die zuständige Person: Sind alle Parteien richtig genannt, sind die Fristen plausibel, entsprechen die Klauseln der aktuellen Richtlinie? Erst nach dem Okay geht das Dokument hinaus.

DREI KONTROLLSTUFEN

VOLLAUTOMATISCH

Geringe Folgen, hohe Menge.

STICHPROBE

Mittlere Folgen. Regelmäßig, aber nicht jeder Fall.

FACHFREIGABE

Hohe Folgen. Kein Ergebnis verlässt das Haus ohne Okay.

IN FÜNF SCHRITTEN AUFSETZEN

- 1 Je Anwendungsfall nüchtern bewerten, was ein Fehler anrichten würde
- 2 Daraus die Kontrollstufe ableiten
- 3 Festlegen, wer haftet und wer die letzte Entscheidung hat
- 4 Prüfkriterien als feste Checkliste standardisieren
- 5 Laufend auswerten: immer derselbe Fehler heißt, der Fehler liegt im System