



Ergänzung

Kapitel 8: KI-Systeme gestalten

Ein KI-System vom Nutzen her entwerfen und menschliche Aufsicht einbauen, die wirklich greift.

- 1 Vom Nutzen her entwerfen
- 2 Menschliche Aufsicht, die greift

Lektion 1 von 2

Vom Nutzen her entwerfen

Ziel: Ich kann ein KI-System von den Menschen her entwerfen, die es nutzen, seinen Informationsfluss aufschreiben und ein Vorhaben in Teile zerlegen.

Erst die Menschen, dann das Werkzeug

Ein KI-System zu entwerfen fängt nicht beim Prompt an, sondern bei zwei Fragen: Wem soll es nützen? Und was passiert diesen Menschen, wenn es sich irrt?

Eine Vorlage für Rezeptideen darf sich öfter irren als eine, die Fristen aus Briefen vom Amt herausschreibt. Der mögliche Schaden bestimmt, wie viel Prüfung ein Entwurf braucht.

Merksatz

Wie viel Sorgfalt ein System braucht, richtet sich nach dem Schaden, den ein Fehler bei Menschen anrichtet.

Der Informationsfluss

Schreib auf, bevor du baust, wie die Daten laufen. Drei Fragen reichen dafür.

Dabei gilt der Grundsatz der Datenminimierung aus der DSGVO (Artikel 5): Hinein kommt nur, was die Aufgabe wirklich braucht.

- Hinein: welche Angaben das System bekommt und von wem
- Heraus: was es zurückgibt und wohin das Ergebnis geht
- Sichtbar: wer Eingaben und Ergebnisse lesen kann, auch der Anbieter des Dienstes

Merksatz

Datensparsam ist ein Entwurf, wenn jede Eingabe einen Grund hat, den du nennen kannst.

Barrieren und Fairness

Ein System nützt nur denen, die es bedienen können und für die es gleich gut funktioniert. Frag deshalb auch, wer damit nicht zurechtkommt. Drei Fragen helfen.

- Sprache: Klappt es in einfacher Sprache und für Leute, die Deutsch lernen?
- Zugang: Geht es mit Vorleseprogramm, auf dem Handy und ohne eigenes Konto?
- Gruppen: Liefert es für alle, die es nutzen, gleich gute Ergebnisse?

Merksatz

Ein Entwurf ist erst fertig, wenn du weißt, für wen er schlechter funktioniert.

Ein Vorhaben zerlegen

Ein großes Vorhaben wie „KI für den Verein“ lässt sich nicht entwerfen. Zerleg es in drei oder vier Teile, die je eine Aufgabe, eigene Nutzende und einen eigenen möglichen Schaden haben.

Dann entwirfst du Teil für Teil, und bei jedem fragst du neu, ob KI dort überhaupt hilft.

Merksatz

Ein Vorhaben ist klein genug zerlegt, wenn du bei jedem Teil sagen kannst, wem er nützt und was ein Fehler kostet.

Lektion 2 von 2

Menschliche Aufsicht, die greift

Ziel: Ich kann menschliche Aufsicht so in ein KI-System einbauen, dass ein Mensch wirklich eingreifen kann und nicht nur abnickt.

Zwei Arten von Aufsicht

Ein Mensch vor der Entscheidung prüft ein Ergebnis, bevor es wirkt: Die Mail geht erst raus, wenn jemand sie freigibt. Ein Mensch, der nur beobachtet, sieht erst hinterher, was das System getan hat.

Je größer der mögliche Schaden, desto eher gehört der Mensch vor die Entscheidung.

Merksatz

Aufsicht greift nur, wenn der Mensch vor dem Schaden an der Reihe ist, nicht erst danach.

Wenn aus Prüfen Abnicken wird

Wer jeden Tag Vorschläge einer Maschine bestätigt, bestätigt irgendwann auch die falschen. Das heißt Automatisierungsverzerrung, englisch automation bias: Man nickt ab, was die Maschine sagt, besonders wenn sie oft richtig liegt und viel zu tun ist.

Die KI-Verordnung der EU verlangt für Hochrisiko-Systeme, dass die Menschen, die sie beaufsichtigen, sich dieser Neigung bewusst bleiben (Artikel 14).

Merksatz

Ein Häkchen unter hundert Vorschlägen ist noch keine Aufsicht.

Eingriffspunkte planen

Aufsicht braucht Stellen, an denen ein Mensch wirklich etwas ändern kann. Vier Werkzeuge gehören dazu, und du planst sie beim Entwurf ein, nicht erst nach dem ersten Fehler.

Die KI-Verordnung nennt für Hochrisiko-Systeme ausdrücklich eine „Stopptaste“ oder ein ähnliches Verfahren.

- Freigabe: Vor einem Schritt, der nach außen wirkt, entscheidet ein Mensch.
- Stopp: Jemand kann das System anhalten, und danach passiert nichts mehr von selbst.
- Stichprobe: Zufällig gezogene Ergebnisse werden regelmäßig gelesen, auch wenn alles gut läuft.
- Protokoll: Festgehalten wird, was das System tat und wer was freigegeben hat.

Merksatz

Ein Stopp, den niemand findet oder auslösen darf, ist kein Eingriffspunkt.

Ein Name und ein Weg zur Fachperson

Aufsicht braucht eine benannte Person, die weiß, dass sie es ist, und Zeit dafür hat. Zwei Dinge machen diese Rolle echt: ein Name statt „das Team“ und ein Weg nach oben.

Bei einem Fall, der über die eigene Kenntnis hinausgeht, gibt die Person ihn an eine Fachperson weiter, etwa an die Ausbilderin, die Datenschutzbeauftragte oder eine Beratungsstelle, und hält fest, warum.

Merksatz

Aufsicht hat einen Namen, Zeit im Kalender und eine Fachperson, an die sie weitergeben kann.

Das nimmst du mit

- Ich kann ein KI-System von den Menschen her entwerfen, die es nutzen, seinen Informationsfluss aufschreiben und ein Vorhaben in Teile zerlegen.
- Ich kann menschliche Aufsicht so in ein KI-System einbauen, dass ein Mensch wirklich eingreifen kann und nicht nur abnickt.

Eine Kette im Studierendenwerk sortiert Anträge vor und meldet einen Fall, der in keine Kategorie passt. Die zuständige Person ist unsicher. Was ist richtig?

- A** Sie gibt den Fall an eine Fachperson weiter und notiert, dass und warum sie es tat
- B** Sie legt den Fall zurück in die Kette, damit die KI ihn beim nächsten Lauf einsortiert
- C** Sie übernimmt die Einschätzung der KI, weil die schon viel mehr Fälle gesehen hat

Auflösung

Richtig ist A: Sie gibt den Fall an eine Fachperson weiter und notiert, dass und warum sie es tat

Ein Fall außerhalb aller Kategorien ist genau der, bei dem die KI am wenigsten Halt hat. Zurücklegen oder Übernehmen schiebt die Entscheidung an die Maschine zurück, die Weitergabe holt Fachwissen dazu.