Verantwoorde AI-afweging

Een beslissingspad: eerst of AI in het proces thuishoort, dan hoe je ze verantwoord inzet

AI is een middel, geen doel. Deze afweging volgt de beslissing in volgorde: stap één vraagt of AI wel het juiste gereedschap is, stap twee lijst de voorwaarden op om ze verantwoord in te zetten. Stoppen na stap één is een geldig resultaat, en vaak het beste.

Stap 1: is AI het juiste gereedschap

- [ ] Is de invoer echt rommelig: vrije tekst, beelden, inschattingen die zich niet tot regels laten herleiden?
- [ ] Heb je gecontroleerd dat een regel, een opzoektabel of een goed formulier het niet oplost?
- [ ] Is een fout antwoord goedkoop, of vangt een mens het op voor het schade aanricht?
- [ ] Moet de taak gebeuren op een schaal of tempo dat geen mens volhoudt?

Elk vakje dat je niet kan aanvinken is een stopteken. De eenvoudigere oplossing is sneller, goedkoper, testbaar en uitlegbaar, en niemand hoeft ze in het oog te houden. Bouw die eerst; lost ze het probleem al op, dan heb je je antwoord.

Stap 2: voorwaarden voor verantwoord gebruik

AI raakte door stap één. Dit zijn de voorwaarden voor ze echte mensen of echte data raakt:

- [ ] Je weet exact welke data het model ziet, en je stuurt het minimum dat de taak nodig heeft
- [ ] Voor persoonsgegevens of gevoelige data is er een rechtsgrond, en wat de modelaanbieder bewaart wordt behandeld als een verwerkingsvraag met een echt antwoord
- [ ] Een mens kijkt de uitvoer na voor ze de betrokkenen bereikt, overal waar een fout pijn zou doen
- [ ] Twijfelt het systeem, dan stopt het, in plaats van iets aannemelijks maar fouts te produceren
- [ ] Fouten hebben een eigenaar: wie ze opvangt, hoe ze rechtgezet worden, wie de betrokkene verwittigt
- [ ] De betrokkenen weten dat er AI meespeelt, in klare taal
- [ ] Kwaliteit wordt gemeten: een nulmeting vooraf, een evaluatieset met lastige gevallen nadien, en een hertest telkens het model verandert
- [ ] De lopende kost is begroot: monitoring, evaluaties, uitgavenplafonds

Elk vakje dat open blijft is werk om voor de lancering te plannen, geen nota voor later.

Rode vlaggen: waar AI niet mag beslissen

Vink aan wat op jouw situatie van toepassing is. Eén vinkje betekent: AI mag de beslissing voorbereiden, maar niet nemen:

- [ ] De beslissing heeft juridische, medische of financiële gevolgen voor een persoon
- [ ] Niemand in je organisatie zou kunnen uitleggen waarom het systeem iemand weigerde
- [ ] Een fout blijft onzichtbaar tot ze al schade heeft aangericht
- [ ] Mensen denken dat een mens dit werk doet, en zouden zich bedrogen voelen als het anders bleek
- [ ] Je zou aarzelen om dit gebruik van AI op je eigen homepage uit te leggen

Zo beslissen wij bij Punfyre, ook wanneer de beslissing nee is. Strandt jouw case op stap één, dan is dat een resultaat: de saaie oplossing wint op elke as.