

CASE-09 · SANITÀ & NUOVE TECNOLOGIE

Investire 25 milioni in AI diagnostica su tutta la rete di ospedali

Adottare un sistema più accurato (nei test del fornitore) per migliorare efficienza e qualità delle cure.

IL GIUDIZIO

tutta la rete → pilota validato

CONFIDENZA CALIBRATA

47%

SETTORE

Sanità



VERDETTO ESECUTIVO

Sui tuoi pazienti l'accuratezza **cala** — e la fiducia cieca crea errori nuovi.

IL GIUDIZIO

tutta la rete → pilota validato

CONFIDENZA

47%

IN UNA RIGA

Evitati **€25 mln** di tecnologia inutilizzata e un rischio clinico-legale da eccesso di fiducia nella macchina.

Il direttore sanitario della rete ospedaliera, con un mandato di innovazione digitale e un budget di trasformazione approvato dal consiglio di amministrazione per il triennio in corso.

Una rete di otto ospedali pubblici di medie dimensioni che serve una popolazione con caratteristiche demografiche eterogenee: alta percentuale di anziani, prevalenza di patologie croniche, forte variabilità etnica tra i presidi. Il sistema di diagnostica per immagini attuale è frammentato su tre piattaforme diverse. Il fornitore del nuovo sistema AI ha validato la propria soluzione su uno studio multicentrico internazionale di 120.000 esami.

§ BACKGROUND

I sistemi di AI diagnostica hanno raggiunto e in alcuni casi superato l'accuratezza media del radiologo umano nei benchmark pubblicati. Tuttavia la letteratura di implementazione reale racconta una storia più complessa: studi indipendenti hanno documentato riduzioni di performance dal 10 al 25% quando un modello addestrato su dati internazionali viene applicato a popolazioni locali con profili demografici differenti. Il fenomeno dell'automation bias — la tendenza dei clinici ad accettare il giudizio della macchina senza verifica critica — è ben documentato e si accentua quando il sistema mostra un'alta confidenza numerica. In assenza di protocolli espliciti di revisione umana e di chiari criteri di escalation, il rischio non è che il sistema non funzioni: è che funzioni abbastanza bene da sembrare affidabile, e poi fallisca in modo sistematico su un sottogruppo specifico di pazienti.

§ SCENARIO CONTROINTUITIVO

Sui pazienti veri l'accuratezza spesso **cala**: la tua popolazione è diversa da quella dei test (età, altre malattie, origini). E c'è un rischio opposto — i medici che si **fidano troppo** della macchina e smettono di controllare, introducendo errori nuovi. Il valore non è il software in sé, ma come lo integri nel lavoro di ogni giorno e come lo governi. Senza, sono 25 milioni di tecnologia che resta sullo scaffale, e per giunta rischiosa.

ASSUNZIONI IMPLICITE

- ◆ L'accuratezza dei test si ritroverà sui pazienti reali.
- ◆ I medici lo adotteranno e si fideranno dei risultati.
- ◆ Il ritorno arriva da meno errori e tempi più brevi.

PROVE DI FALSIFICAZIONE

- ◆ Provarlo prima sui TUOI pazienti, non sui test del fornitore.
- ◆ In un pilota: quante volte i medici lo correggono? Quanto si fidano troppo?
- ◆ Funziona allo stesso modo su tutti i gruppi di pazienti?

DOMANDE CHE ALZANO L'ONERE

- ◆ Qual è l'accuratezza sui tuoi pazienti reali, non su quelli pubblicati?
- ◆ Chi risponde, legalmente, quando la macchina sbaglia?

CONFIDENZA CALIBRATA & PROVENIENZA

che l'accuratezza dei test regga su tutta la rete

Provenienza: studio di validazione del fornitore · dati del pilota locale · letteratura sull'eccesso di fiducia nella macchina · base red-team.

Accuratezza dichiarata vs pilota locale

DATI ILLUSTRATIVI

Studio fornitore (pop. internazionale)

93%

Pilota locale (pop. rete)

81%

Accuratezza diagnostica del sistema AI nello studio di validazione del fornitore vs nel pilota condotto su un singolo presidio della rete con la popolazione locale (dati illustrativi).

Tasso di accettazione acritica in pilota

DATI ILLUSTRATIVI

64%

DEI REFERTI AI ACCETTATI
SENZA REVISIONE CRITICA NEL
PILOTA

Quota di referti AI accettati dai medici senza modifica né revisione critica nel pilota locale — segnale di automation bias già nella fase sperimentale (dati illustrativi).

ESITO

Estensione subordinata a una **validazione sul campo locale** e a regole chiare di governo e responsabilità medica.

VALORE

Evitati **€25 mln** di tecnologia inutilizzata e un rischio clinico-legale da eccesso di fiducia nella macchina.

NOTA METODOLOGICA — LEGGERE PRIMA

Casi compositi, nel metodo della Harvard Business Review: ricostruzioni basate su situazioni reali e ricorrenti in ciascun settore, fuse e anonimizzate per tutelare la riservatezza. Le dinamiche decisionali sono autentiche; nomi, cifre e dettagli sono modificati e non riferibili a un singolo cliente o caso. Le indicazioni di «provenienza» descrivono la tipologia di evidenza che il motore cita con tracciabilità in produzione. I valori Δ -CSI illustrano l'intensità della pressione esercitata dal contraddittorio sulle assunzioni.