

# MACCHINE CHE PENSANO, MEDICI CHE DECIDONO

*Storia dei Clinical Decision Support Systems:  
dall'utopia cibernetica  
all'intelligenza artificiale in ostetricia  
e ginecologia*



Carlo Piscicelli

2026



Nel 1959, un medico di nome Keeve Brodman affermò che la diagnosi poteva essere eseguita da una macchina. Aveva ragione nel principio. Sbagliava i tempi — di decenni. Questa è la storia di quei decenni.

Macchine che pensano, medici che decidono ripercorre sessant'anni di Clinical Decision Support Systems — i sistemi informatici progettati per aiutare i medici a prendere decisioni cliniche — con la struttura narrativa di un romanzo e il rigore di un saggio scientifico. Dai primi sistemi esperti degli anni Settanta, come MYCIN — capace di battere gli specialisti di malattie infettive eppure mai entrato in un ospedale — alle reti neurali che oggi leggono le mammografie meglio dei patologi e scoprono, per pura sorpresa statistica, che quelle stesse immagini rivelano il rischio di infarto cardiaco.

Il libro racconta i pionieri che hanno tentato di costruire la mente artificiale del medico: Edward Shortliffe, il giovane studente di Stanford che codificò la conoscenza di un esperto in seicento regole logiche; Jack Myers, il grande clinico di Pittsburgh che dettò trent'anni di esperienza diagnostica a un calcolatore; Homer Warner, il cardiologo che applicò il teorema di Bayes alla diagnosi congenita prima che esistesse il termine intelligenza artificiale. Racconta perché sistemi brillanti e validati scientificamente non hanno mai raggiunto le corsie — e cosa è cambiato nel 2022, quando un sistema addestrato a prevedere la prossima parola ha improvvisamente cominciato a ragionare.

In un'epoca in cui Mustafa Suleyman — figlio di un taxista siriano, co-fondatore di DeepMind, CEO di Microsoft AI — avverte che ci stiamo avvicinando a una soglia critica nella storia della nostra specie, e in cui la mammografia di una donna può rivelare, grazie all'intelligenza artificiale, non soltanto un tumore ma anche il rischio di un infarto, questo saggio si propone come guida storica e strumento critico per chi lavora in medicina e vuole capire il mondo che sta per cambiare.

Il punto di vista è quello di un ginecologo-ostetrico che ha vissuto dall'interno questa trasformazione — costruendo strumenti digitali per la propria pratica clinica, usando l'intelligenza artificiale come co-autore, e rimanendo convinto che la mente che guida debba restare umana.

## L'autore

---

Carlo Piscicelli è specialista in Ostetricia e Ginecologia con oltre trent'anni di esperienza clinica, chirurgica e accademica. Ha diretto l'Unità Operativa di Ostetricia e Ginecologia dell'Ospedale Cristo Re di Roma dal 2013 al 2023, e opera attualmente come Senior Consultant presso la Fondazione Policlinico Universitario Agostino Gemelli IRCCS di Roma. Esercita come libero professionista in diversi poliambulatori della capitale ed è docente universitario presso l'Università LUMSA.

Autore di oltre 90 pubblicazioni scientifiche in letteratura internazionale, ha partecipato come relatore a numerosi congressi nazionali e internazionali di ostetricia, ginecologia e medicina perinatale. È consulente medico-legale in ambito ostetrico-ginecologico.

Da sempre interessato all'applicazione dell'informatica alla clinica, ha sviluppato negli anni Novanta e Duemila GinHelp® e GynDesk® — realizzati insieme al professor Serra — tra i primi software nazionali di supporto decisionale per la ginecologia, distribuiti su scala nazionale. Negli ultimi anni ha costruito GINOS (Gynecology Integrated Network Operating System), una suite gratuita di applicazioni web di Clinical Decision Support a uso clinico, disponibile all'indirizzo [ginos-cdss.it](http://ginos-cdss.it).

L'interesse per i CDSS e per l'intelligenza artificiale applicata alla medicina — coltivato attraverso la costruzione diretta di strumenti clinici con il supporto di modelli linguistici di nuova

generazione — è il filo conduttore di questo saggio, scritto in parte in collaborazione con Claude (Anthropic), secondo il modello del clinico-autore che queste pagine propongono.



# Indice

---

<b>Premessa</b> .....	<b>10</b>
<b>Prologo</b> .....	<b>16</b>
<b>Capitolo I</b> .....	<b>18</b>
La profezia di William Schwartz.....	18
Lo sfondo culturale: perché i medici temevano le macchine.....	19
<b>Capitolo II</b> .....	<b>20</b>
Due modi diversi di imitare il medico.....	20
Ledley e Lusted: la diagnosi come teorema di Bayes.....	21
Homer Warner a Salt Lake City: il primo sistema statistico funzionante.....	23
Il tallone d’Achille dell’approccio probabilistico.....	24
<b>Capitolo III</b> .....	<b>25</b>
Nomenclatura come problema politico.....	25
Clement McDonald e i «clinical reminders»: la via pragmatica.....	26
La definizione che cambiò tutto.....	27
Il quadro di Wright e Sittig: quattro generazioni di architetture.....	28
<b>Capitolo IV</b> .....	<b>30</b>
Un giovane studente a Stanford.....	30
Come funzionava MYCIN.....	31
La valutazione del 1979: meglio degli specialisti.....	32
Il paradosso: celebre ma inutilizzato.....	32
<b>Capitolo V</b> .....	<b>34</b>
Jack Myers, il grande clinico di Pittsburgh.....	34
Il progetto enciclopedico.....	35
Il confronto con i discussant del Massachusetts General Hospital.....	36
Perché anche INTERNIST-I rimase un esperimento.....	36
<b>Capitolo VI</b> .....	<b>37</b>
<b>Capitolo VII</b> .....	<b>39</b>
Dal centro di calcolo al reparto.....	39
L’HITECH Act: la svolta normativa.....	40
<b>Capitolo VIII</b> .....	<b>41</b>
Quando le macchine smettono di seguire regole e cominciano ad imparare.....	41
I cinque modi di mescolare regole ed esperienza.....	42
<b>Capitolo IX</b> .....	<b>43</b>

Un cervello addestrato a prevedere la prossima parola.....	43
La logica come proprietà emergente.....	44
Il problema delle allucinazioni — e come è cambiato.....	45
Il cervello dietro l'AI non parla inglese come madrelingua.....	47
<b>Capitolo X.....</b>	<b>48</b>
Quando la guardia del sistema diventa il suo tallone d'Achille.....	48
Le strategie per uscire dal loop.....	49
<b>Capitolo XI.....</b>	<b>50</b>
L'ostetricia come terreno privilegiato.....	50
Il CTG — la promessa più antica.....	51
Il problema dell'interpretazione: quando gli esperti non sono d'accordo.....	53
Trent'anni di sistemi computerizzati: SisPorto e INFANT.....	52
L'intelligenza artificiale e il CTG: performance promettenti, risultati irrisolti.....	54
Il paradosso ACOG 2025: una raccomandazione contro i CDSS.....	56
Un caveat necessario: l'AI del 2025 non è quella del 2020.....	57
Il cambio di paradigma: dall'interpretazione categoriale alla lettura fisiopatologica.....	58
I primi tentativi di CDSS fisiopatologico: FIT-CAT e TWERIS Mini.....	60
Perché non esistono ancora altri CDSS fisiopatologici.....	63
Lo screening del cancro cervicale: quando la macchina batte il patologo.....	64
La predizione delle complicanze in gravidanza: il sogno dell'ostetricia di precisione.....	65
L'ecografia ostetrica nell'era dell'AI: verso gli „ultrasuoni 5D“.....	68
Una lezione che attraversa tutto il capitolo.....	69
<b>Capitolo XII.....</b>	<b>70</b>
Il problema dell'implementation gap.....	70
Il modello del clinico-autore.....	70
<b>Capitolo XIII.....</b>	<b>72</b>
Una macchina che inventa e una macchina che legge.....	72
Che cos'è il RAG — spiegato ai ragazzi delle medie.....	73
Perché il RAG non allucina.....	74
Dal libro generico al manuale personalizzato sulla tua paziente.....	75
Il RAG come architettura naturale per i CDSS.....	76
Una promessa concreta.....	77
<b>Capitolo XIV.....</b>	<b>78</b>
Il paradosso al cuore della medicina basata sull'evidenza.....	78
Tre problemi che nessun documento può risolvere.....	79
La linea guida come architettura, il CDSS come ingegnere.....	81
GINOS come laboratorio di traduzione delle linee guida.....	82
Le linee guida tacciono: il valore del consenso clinico implicito.....	83
<b>Capitolo XV.....</b>	<b>85</b>
Il framework TASS: tecnica, stakeholder e società.....	85

Il paradosso dell'esperto.....	86
<b>Capitolo XVI.....</b>	<b>87</b>
Il framework PRISM e il design centrato sull'utente.....	87
I cinque diritti del CDS.....	88
<b>Epilogo.....</b>	<b>89</b>
Il momento eccezionale e inquietante.....	89
La mammografia che guarda il cuore — una scoperta non cercata.....	91
Centauri, Cyborg — e la domanda che resta.....	92
<b>Nota finale: Fare il medico nel 2026.....</b>	<b>94</b>
<b>Il Progetto GINOS.....</b>	<b>100</b>
<b>Fonti e riferimenti bibliografici principali.....</b>	<b>107</b>
<b>Dichiarazione di trasparenza.....</b>	<b>97</b>

## Premessa

Ho incontrato i Clinical Decision Support Systems due volte nella mia vita professionale. La prima volta senza sapere che si chiamassero così.

Era la fine degli anni Novanta. Lavoravo all'Ospedale Cristo Re di Roma, in una sala parto che somigliava a tutte le sale parto del mondo: il ritmo irregolare dei monitor fetali, le luci al neon, il conto alla rovescia silenzioso che precede ogni nascita. Capitava — come capita in ogni reparto di ostetricia del mondo — che nonostante tutta l'attenzione, nonostante il voler applicare le regole dell'evidenza di quel momento storico, nascessero bambini che non stavano bene. Bambini con asfissia. Bambini con segni di encefalopatia ipossico-ischemica. Era deprimente.

Non nel senso clinico del termine, ma in quello umano: la sensazione che qualcosa avrebbe potuto essere fatto meglio, che il tracciato cardiocografico avesse detto qualcosa che forse non avevamo letto abbastanza bene, che le variabili cliniche avessero disegnato un quadro che avremmo potuto interpretare diversamente.

Ma c'era qualcosa di ancora più difficile da gestire sul piano intellettuale: il CTG è uno strumento imperfetto nel senso più tecnico del termine. Un tracciato normale non garantisce che il bambino stia bene — ci sono neonati che nascono compromessi con un CTG che fino a pochi minuti prima sembrava rassicurante. Un tracciato alterato non significa che il bambino sia in pericolo — la grande maggioranza dei feti che mostrano decelerazioni, tachicardia o perdita di variabilità sopportano il travaglio senza conseguenze. Ed è proprio questo il problema statistico che rende così difficile la gestione del monitoraggio intrapartum: una qualche

alterazione del tracciato si verifica in oltre la metà di tutti i travagli. Se ogni alterazione portasse a un cesareo, i tassi operatori sarebbero insostenibili. Se ogni alterazione venisse ignorata nell'attesa che le cose si chiarissero da sole, qualche bambino avrebbe meno tempo di quanto pensassimo.

A questa complessità clinica si affiancava una preoccupazione più silenziosa ma non meno reale: il rischio medico-legale. Come facevo a dimostrare, a posteriori, di aver seguito correttamente la lettura del tracciato CTG? Di aver applicato le regole giuste nel momento giusto? In caso di esito avverso — e gli esiti avversi capitano anche quando si fa tutto bene, anche quando non si poteva fare nulla di diverso — la domanda del perito è sempre la stessa: qual era il vostro ragionamento? Su quali dati si basava? Quali regole stavate seguendo? Erano appena uscite le linee guida del Royal College of Obstetricians and Gynaecologists sul monitoraggio intrapartum. Erano rigorose, sistematiche, basate sull'evidenza. Ma non erano uno strumento clinico: erano un documento che richiedeva di essere tradotto in decisioni, passo dopo passo, in tempo reale, da un ostetrico stanco alla terza ora di guardia.

Così, insieme al professor Serra — che in quel periodo dirigeva il nostro reparto — cominciammo a lavorare su qualcosa di diverso. Decidemmo di dare un peso a ciascuna variabile del tracciato CTG e a ciascuna variabile clinica rilevante: la frequenza cardiaca di base, la variabilità, le decelerazioni, la loro forma e il loro rapporto con le contrazioni, l'epoca gestazionale, le condizioni cliniche della madre. Ciascuna variabile aveva un punteggio. La loro combinazione produceva uno score. Lo score ci diceva come dovevamo ragionare: questo profilo di tracciato, in queste

condizioni cliniche, suggerisce questa categoria di rischio, richiede questa risposta.

Ma la funzione più preziosa dello score non era quella prospettica. Era quella retrospettiva: documentava, in forma strutturata e verificabile, cosa avevamo fatto, perché lo avevamo fatto, e quale ragionamento avevamo seguito. Questo va capito bene, perché potrebbe sembrare una forma di medicina difensiva — uno strumento costruito per proteggersi in caso di contenzioso. Non era così. Era esattamente il contrario.

Quando un medico sa di poter dimostrare di aver seguito una logica corretta, di aver applicato le regole disponibili al momento della decisione, di aver ragionato in modo documentato e verificabile — quella certezza lo libera. Lo libera dalla paura. E un medico libero dalla paura decide in modo diverso: non si chiede «cosa devo fare per non essere attaccato?» ma «cosa è giusto fare per questa paziente?». La documentazione del ragionamento non produce medicina difensiva. La previene.

È questa la catena causale che spiega quello che osservammo nel tempo: il numero di cesarei per sofferenza fetale scese. Non perché il sistema scoraggiasse il cesareo — anzi, in certi profili lo indicava chiaramente come scelta necessaria — ma perché togliere al medico l'ansia della giustificazione a posteriori significava restituirgli la libertà di ragionare in avanti. Di fare la cosa giusta, non la cosa più difendibile. Il medico che documenta il proprio ragionamento non ha bisogno di proteggersi con il bisturi.

Quello score nacque su carta. Era un foglio strutturato — variabili elencate, punteggi da assegnare, totale da calcolare — che si compilava a mano durante il travaglio. Solo in un secondo tempo lo informatizzammo, costruendo un programma che automatizzava

i calcoli e archiviava i dati in forma digitale. Ed è a quel punto, tecnicamente, che divenne un vero CDSS: la definizione formale che il campo adottò negli anni Ottanta richiedeva esplicitamente un supporto elettronico. Il foglio cartaceo era la logica clinica. Il software era il CDSS. Ma senza il foglio, il software non sarebbe mai esistito.

Ma non fu l'ultimo strumento che costruimmo. Cominciammo a costruirne altri, su problemi clinici diversi che avevano in comune la stessa struttura di fondo: troppe variabili da tenere a mente simultaneamente, troppe conseguenze asimmetriche tra le opzioni disponibili, e la necessità — clinica prima ancora che medico-legale — di documentare che la scelta fatta era stata ragionata.

Uno degli esempi che ricordo con più chiarezza era l'algoritmo per decidere se prescrivere eparina a basso peso molecolare dopo un intervento chirurgico ginecologico. Se dai l'eparina rischi l'emorragia; se non la dai rischi la trombosi. Entrambe le complicanze possono verificarsi anche quando l'emostasi è stata impeccabile, anche quando il rischio trombotico sembrava basso. Un'emorragia postoperatoria in una paziente che hai deciso di anticoagulare ti lascia con la domanda: eri giustificato? Una trombosi in una paziente che non hai anticoagulato ti lascia con la stessa domanda, rovesciata. L'algoritmo non eliminava il rischio — niente lo elimina — ma rendeva esplicita la logica della scelta: questi fattori di rischio trombotico, questi fattori di rischio emorragico, questa la bilancia, questa la raccomandazione. Dimostrare di aver fatto la cosa giusta non è soltanto una questione medico-legale. È rispetto per la paziente. È segno di serietà professionale. È il modo in cui la medicina si distingue dall'improvvisazione.

Li chiamavamo algoritmi clinici. Non avevamo un'altra parola. E in fondo era la parola giusta: percorsi logici strutturati, costruiti per aiutare a ragionare in modo sistematico davanti a un problema complesso. Solo molto più tardi avremmo scoperto che il mondo dell'informatica medica li chiamava in un altro modo.

Non lo chiamavamo così. Non conoscevamo MYCIN, non avevamo letto Shortliffe, non sapevamo che vent'anni prima un chirurgo di Leeds di nome Francis Timothy de Dombal aveva fatto qualcosa di concettualmente simile — pesare le variabili cliniche, costruire un modello probabilistico, confrontarlo con il ragionamento dei medici — e aveva scoperto che il sistema non funzionava perché fosse più intelligente dei clinici, ma perché li obbligava a essere più rigorosi. Li usava come, scrisse trent'anni dopo, «un focus educativo e uno stimolo alla buona pratica clinica». Lo score CTG che avevamo costruito funzionava esattamente allo stesso modo.

La seconda volta che ho incontrato i CDSS è stata nel 2022. Avevo sentito parlare di ChatGPT come tutti, con la curiosità un po' scettica con cui i medici della mia generazione accolgono le novità tecnologiche. Poi un pomeriggio ho aperto il browser, ho descritto a quel sistema il percorso clinico per la gestione dell'osteoporosi in una donna in menopausa, e ho visto apparire sullo schermo qualcosa che somigliava — non perfettamente, ma sorprendentemente — a ciò che avrei scritto io. In pochi minuti. Senza errori di impostazione logica. Con la struttura di un percorso decisionale che avrei impiegato giorni a costruire.

Da quel momento, non ho smesso. Ho costruito GINOS — una suite gratuita di applicazioni web per il supporto decisionale in ostetricia e ginecologia, accessibile su [ginos-cdss.it](http://ginos-cdss.it) — usando i modelli di intelligenza artificiale come collaboratori. Non come

oracoli. Come interlocutori intelligenti con cui sviluppare, in linguaggio naturale, la logica clinica che avevo in testa. La stessa logica che trent'anni prima avevo tradotto prima in un foglio di carta e poi in un programma informatico.

Questo testo nasce dalla stessa passione. Volevo capire — e poi raccontare — da dove venivano i sistemi che stavo usando. Volevo sapere chi erano le persone che avevano tentato, cinquant'anni prima di me, la stessa cosa che stavo tentando io. Volevo capire perché avevano ragione troppo presto. E volevo capire cosa è cambiato oggi, perché improvvisamente quello che sembrava impossibile sembra inevitabile.

Ho scritto questo saggio con l'aiuto dell'intelligenza artificiale — usando Claude come co-autore per trasformare la materia grezza della letteratura scientifica in narrazione. Le idee, il filo narrativo, i personaggi storici, le tesi argomentative sono miei. I contenuti sono stati sviluppati con il supporto dell'AI. Questo è esattamente il modello che Ethan Mollick chiama lavoro da centauro: io decido la strategia, verifico ogni passaggio, mi assumo la responsabilità del risultato. Non mi rende meno autore: mi rende un autore che usa strumenti nuovi con spirito critico vecchio.

Il professor De Dombal aveva capito cinquant'anni fa che lo strumento non sostituisce il ragionamento del medico. Lo attiva. Lo rende più rigoroso. Garry Kasparov, dopo aver perso contro Deep Blue nel 1997, capì che la risposta non era rinunciare alle macchine ma imparare a cavalcarle — e inventò il concetto di «centauro». Io lo chiamo fare il medico nel 2026.

*Carlo Piscicelli*  
*Roma, aprile 2026*

## Prologo: Una promessa lunga sessant'anni

C'è un momento preciso in cui l'essere umano smette di considerare la propria mente come l'unica sede del pensiero affidabile. Per la medicina, quel momento arriva nel 1959, nella sala di un convegno a New York, quando un medico di nome Keeve Brodman prende la parola e pronuncia una frase destinata a diventare profetica e, al tempo stesso, scandalosa: la formulazione di corrette interpretazioni diagnostiche dei sintomi può essere un processo in tutti gli aspetti logici e così completamente definito da poter essere eseguito da una macchina.

**Keeve Brodman (1904–1992)** — *Medico internista, professore al New York Hospital-Cornell Medical Center. Sviluppò il Cornell Medical Index (CMI), uno dei primi questionari standardizzati per la raccolta strutturata della storia clinica, tradotto in diciotto lingue e usato in tutto il mondo. Nel 1959 pubblicò su Science — con Erdmann e Wolff — uno dei primissimi studi sull'uso dei calcolatori per la classificazione diagnostica dei sintomi.*

Brodman non era un visionario eccentrico. Era un clinico serio, un professore di psicologia medica al Weill Cornell Medical College di New York, che negli anni Quaranta e Cinquanta aveva sviluppato il Cornell Medical Index, uno dei primi strumenti standardizzati per raccogliere la storia clinica dei pazienti in forma strutturata. Aveva capito, prima di quasi tutti i suoi colleghi, che la diagnosi medica non era soltanto intuizione, arte, tocco — quell'ineffabile qualità del grande clinico — ma anche, e soprattutto, elaborazione di informazioni. E le informazioni, a differenza delle emozioni, possono essere trattate sistematicamente.

L'entusiasmo di Brodman era figlio di un'epoca. Erano gli anni dello Sputnik, della corsa allo spazio, dell'automazione industriale, della nascita dei primi calcolatori elettronici. L'America

usciva dalla Seconda Guerra Mondiale con la certezza di poter risolvere qualsiasi problema con abbastanza ingegno e abbastanza denaro. La scienza era il nuovo Prometeo. La tecnologia era il suo fuoco.

In questo clima, l'idea che un calcolatore potesse aiutare — o addirittura sostituire — il ragionamento diagnostico del medico non sembrava fantascienza. Sembrava il passo successivo ovvio di una civiltà in marcia. E così, mentre le prime macchine IBM riempivano intere stanze climatizzate con il loro ronzio di ventole e il crepitio delle schede perforate, i medici più visionari cominciarono a chiedersi: e se le macchine potessero pensare come noi? E se potessero pensare meglio di noi?

La storia che state per leggere racconta quella domanda — e le sue risposte, parziali, contraddittorie, sorprendenti, a volte tragicamente incompiute. È la storia dei Clinical Decision Support Systems, i sistemi informatici progettati per aiutare i medici a prendere decisioni cliniche. Una storia che attraversa sei decenni, muta pelle mille volte, conosce trionfi dimenticati e fallimenti celebrati, e approda oggi, tra le corsie degli ospedali e le sale ecografiche, nell'era dell'intelligenza artificiale.

Una storia, in fondo, di esseri umani che cercano di capire come pensano — e di costruire macchine che lo facciano al posto loro.



## Capitolo I: Gli anni dell'ottimismo ingenuo (1959–1972)

### La profezia di William Schwartz

Nel 1970, un nefrologo di Boston di nome William B. Schwartz pubblicò sul *New England Journal of Medicine* un articolo destinato a fare epoca. Il titolo era quasi timido nella sua precisione burocratica: *Medicine and the Computer: The Promise and Problems of Change*. Ma le parole contenute in quelle pagine erano tutt'altro che timide.

Schwartz era un clinico di primo piano alla Tufts University School of Medicine, autore di ricerche fondamentali sull'equilibrio acido-base e sulla fisiologia renale. Non era un informatico, non era un matematico. Era un medico che aveva dedicato trent'anni ad ascoltare corpi umani e a costruire, lentamente, un'enciclopedia mentale di pattern diagnostici. Ed era convinto che quella enciclopedia potesse, un giorno, essere trascritta su un calcolatore.

*"Entro l'anno 2000, i computer avranno assunto un ruolo completamente nuovo in medicina, agendo come potente estensione dell'intelletto del medico."*

**William B. Schwartz (1922–2009)** — Nefrologo di fama internazionale alla Tufts University School of Medicine di Boston, autore di ricerche fondamentali sull'equilibrio acido-base e sulla fisiologia renale. Pubblicò nel 1970 sul *New England Journal of Medicine* un articolo profetico — *Medicine and the Computer: The Promise and Problems of Change* — che anticipò con straordinaria lucidità le sfide dell'informatica medica. Non era un tecnologo: era un clinico che aveva compreso, prima di quasi tutti i colleghi, che la medicina era nella sua sostanza un'attività di elaborazione dell'informazione, e che questa elaborazione poteva essere sistematizzata. Continuò a scrivere sull'AI in medicina per decenni, tornando periodicamente a valutare quanto del 1970 si fosse realizzato — con risultati misti, e con una franchezza intellettuale che lo rese un punto di riferimento del settore.

Era una previsione audace per il 1970. I calcolatori dell'epoca erano macchine enormi, inaccessibili, che richiedevano squadre di tecnici specializzati per funzionare. L'idea di un medico che consultasse un computer al letto del paziente come consulterebbe

un collega sembrava, al massimo, un sogno per il futuro lontano. E tuttavia Schwartz ci credeva. E aveva, in qualche misura, ragione.

Aveva ragione nel principio: la medicina è, nella sua sostanza più profonda, un'attività cognitiva. Il medico raccoglie informazioni — sintomi, segni, esami — e le elabora per produrre una decisione: questa diagnosi, questa terapia, questo follow-up. Se l'elaborazione delle informazioni può essere formalizzata, può essere automatizzata. Se può essere automatizzata, può essere migliorata, standardizzata, resa accessibile a tutti, non soltanto ai grandi clinici delle grandi città.

Aveva torto, invece, nei tempi. Il 2000 arrivò e i computer in medicina c'erano, ma il ruolo immaginato da Schwartz era ancora di là da venire. Non perché la tecnologia non fosse sufficiente. Ma perché tra la tecnologia e la medicina si frapponeva qualcosa di molto più resistente: la cultura professionale dei medici, i sistemi organizzativi degli ospedali, la complessità ingovernabile della malattia umana.

### **Lo sfondo culturale: perché i medici temevano le macchine**

Per capire la storia dei CDSS bisogna capire, prima di tutto, chi erano i medici degli anni Sessanta e Settanta e che cosa rappresentava per loro l'idea di una macchina diagnostica.

Il medico di quella generazione era al centro di una delle professioni più prestigiose della società occidentale. Il suo potere era fondato, in parte, sulla conoscenza — anni di studio, di tirocini estenuanti, di notti insonni — e, in parte, sul mistero di quella conoscenza. Il paziente non capiva perché il medico facesse certe domande, toccasse certi punti, ordinasse certi esami. Quella incomprendibilità era potere. Era la base della fiducia, dell'autorità, del compenso.

Un sistema informatico che codificasse la logica diagnostica non era soltanto una macchina utile. Era una minaccia esistenziale all'identità professionale. Se una macchina poteva fare ciò che il medico faceva, che cosa restava del medico? Se la diagnosi era un algoritmo, chi aveva bisogno di venti anni di formazione specialistica?

Questa resistenza non era irrazionale. Era umana, comprensibile, radicata in timori legittimi. E avrebbe segnato, in modi diversi, tutta la storia successiva dei sistemi di supporto alle decisioni cliniche.



## Capitolo II: I due paradigmi alle origini — regole contro probabilità

### Due modi diversi di imitare il medico

Prima che MYCIN esistesse, prima che Shortliffe iniziasse il suo dottorato a Stanford, un'altra tradizione di ricerca stava silenziosamente costruendo le proprie fondamenta in un laboratorio di Washington. Era una tradizione radicalmente diversa nell'approccio, nei protagonisti, nella filosofia di fondo. Non cercava di codificare le regole dei medici esperti. Cercava di imitare la statistica della malattia.

Per capire la differenza tra i due paradigmi bisogna tornare al cuore del problema diagnostico. Quando un medico vede un paziente con febbre, dolore toracico e dispnea, come arriva alla diagnosi? Ci sono, in sostanza, due risposte possibili. La prima: applica regole. Sa che febbre più dolore pleuritico più dispnea uguale probabilmente polmonite, e naviga l'albero decisionale fino

alla conclusione. La seconda: ricorda. Sa che di mille pazienti che ha visto con quella triade di sintomi, settecento avevano la polmonite, centocinquanta una pericardite, cinquanta una tromboembolica. E scommette sulla diagnosi più frequente.

Il primo approccio è logico-deduttivo. Il secondo è statistico-probabilistico. Il primo è esplicito, trasparente, giustificabile. Il secondo è implicito, basato sull'esperienza accumulata, difficile da verbalizzare. Entrambi i metodi esistono nel ragionamento clinico reale — spesso in modo intrecciato, inseparabile. Ma quando i pionieri degli anni Cinquanta e Sessanta cercarono di replicarli in una macchina, li separarono nettamente. E questa separazione generò due storie parallele che avrebbero dovuto, decenni dopo, ricongiungersi.

### **Ledley e Lusted: la diagnosi come teorema di Bayes**

Il 9 ottobre 1959, la rivista *Science* pubblicò un articolo che pochi lettori avrebbero saputo apprezzare nella sua piena portata. Il titolo era astratto, quasi ostico per un medico clinico: *Reasoning Foundations of Medical Diagnosis*. Gli autori erano Robert Stuart Ledley e Lee Byron Lusted.

Ledley non era un medico. Era un biomedical engineer, un ingegnere biomedico formato alla Columbia University, che lavorava per la National Academy of Sciences a Washington. Aveva una mente matematica applicata alla biologia — una combinazione rara per l'epoca — e un interesse quasi ossessivo per la logica formale. Lusted era invece un radiologo dell'University of Rochester, uno specialista che passava le giornate a interpretare immagini cliniche, a tradurre sfumature di grigio in diagnosi. Sembravano due mondi lontani. Ma Lusted aveva capito, nella sua pratica quotidiana, che la lettura di una lastra radiografica era un

processo probabilistico — mai una certezza assoluta, sempre una stima di verosimiglianza.

Insieme, Ledley e Lusted proposero qualcosa di radicale: la diagnosi medica poteva essere formalizzata attraverso il teorema di Bayes. La logica era elegante. Se sappiamo con quanta frequenza una certa malattia si presenta nella popolazione generale — la probabilità a priori — e se sappiamo quanto spesso quella malattia produce certi sintomi — la probabilità condizionata — allora possiamo calcolare matematicamente la probabilità che un paziente con quei sintomi abbia quella malattia. Niente intuizioni. Niente esperienza soggettiva. Solo matematica.

#### Il teorema di Bayes spiegato a un ragazzo di terza media

Immagina che a scuola qualcuno stia starnutando. Quanto è probabile che abbia il raffreddore?

Prima ancora di sentirlo starnutire, sai già qualcosa di utile: in quella stagione, il 10% dei ragazzi della scuola ha il raffreddore. Questa è la probabilità di partenza — quella che Bayes chiama probabilità a priori: ciò che sai prima di osservare il sintomo.

Poi arriva il sintomo: lo starnuto. Sai anche che chi ha il raffreddore starnutisce spesso — diciamo nel 90% dei casi. Ma sai anche che ogni tanto si starnutisce per altri motivi — polvere, allergia — e che il 20% dei ragazzi senza raffreddore starnutisce comunque almeno una volta al giorno.

Il teorema di Bayes combina queste tre informazioni — quanto è comune la malattia, quanto spesso produce il sintomo, quanto spesso il sintomo compare anche senza la malattia — e calcola la probabilità aggiornata. In questo caso, circa il 33%: uno starnuto triplica la probabilità di raffreddore rispetto a quella di partenza.

In medicina: la probabilità a priori è quanto è frequente una malattia nella popolazione; la probabilità condizionata è quanto spesso quella malattia produce un certo sintomo; la probabilità a posteriori è la risposta che vogliamo — quanto è probabile che questo paziente, con questo sintomo, abbia quella malattia. Bayes ci dà una formula per calcolarla in modo rigoroso, senza affidarsi solo all'istinto.

*"La formulazione matematica della diagnosi rivela la struttura logica del problema clinico e apre la strada a strumenti computazionali capaci di assistere il medico nel ragionamento sistematico."*

L'articolo di Ledley e Lusted fu una pietra miliare intellettuale. Dimostrò che il ragionamento diagnostico aveva una struttura formale che poteva essere trattata matematicamente. Aprì un dibattito che avrebbe occupato le riviste di informatica medica per due decenni. E seminò il seme di quello che sarebbe

diventato, molto più tardi, il machine learning in medicina. Non a caso, Lusted fondò poco dopo il Medical Information Research Laboratory, e nel 1967 contribuì a istituire la Society for Medical Decision Making — la prima istituzione dedicata alla formalizzazione del ragionamento clinico.

### **Homer Warner a Salt Lake City: il primo sistema statistico funzionante**

Se Ledley e Lusted avevano costruito la teoria, un cardiologo di Salt Lake City di nome Homer Richardson Warner ne fece la pratica.

Warner era una figura singolare nel panorama medico americano degli anni Cinquanta. Cardiologo all'University of Utah, aveva due passioni apparentemente incompatibili: la fisiologia cardiaca e la matematica computazionale. In un'epoca in cui i calcolatori erano macchine industriali inaccessibili ai clinici, Warner aveva ottenuto l'accesso ai computer dell'università e aveva cominciato a usarli per analizzare i dati dei suoi pazienti. Aveva capito — prima di quasi chiunque altro nel mondo clinico — che un grande dataset di pazienti cardiopatici conteneva informazioni che nessun singolo clinico poteva elaborare mentalmente: pattern statistici nascosti, correlazioni non intuitive, segnali deboli dispersi nel rumore della variabilità individuale.

Nel 1961, Warner e i suoi collaboratori pubblicarono su *The American Journal of Cardiology* uno studio che rappresenta una pietra miliare dimenticata nella storia dell'intelligenza artificiale medica: un sistema bayesiano capace di diagnosticare le cardiopatie congenite nei neonati. Il sistema prendeva in ingresso un insieme di segni clinici — il tipo di cianosi, i caratteri dell'auscultazione cardiaca, i reperti radiografici — e calcolava, per

ciascuna delle principali diagnosi cardiologiche congenite, la probabilità secondo il teorema di Bayes. Poi segnalava la diagnosi più probabile.

I risultati erano impressionanti per l'epoca: un'accuratezza diagnostica molto alta sui casi di addestramento, con risultati promettenti anche su pazienti nuovi. Warner aveva dimostrato che un approccio statistico poteva competere con il ragionamento clinico tradizionale. Andò oltre: negli anni seguenti, costruì all'University of Utah uno dei primi sistemi informativi ospedalieri integrati del mondo — il HELP (Health Evaluation through Logical Processing) system — che avrebbe poi incorporato elementi di supporto decisionale e sarebbe rimasto in uso, continuamente aggiornato, per decenni. Warner è forse il personaggio più longevo di questa storia: continuò a pubblicare sull'informatica medica fino agli anni Novanta, e visse abbastanza a lungo da vedere parte dei suoi sogni realizzati.

### **Il tallone d'Achille dell'approccio probabilistico**

Il teorema di Bayes funziona in modo ottimale quando le variabili di ingresso sono statisticamente indipendenti tra loro — cioè quando la presenza di un sintomo non influenza la probabilità di un altro. Ma in medicina questa condizione è quasi sempre violata. La febbre e la tachicardia, la dispnea e l'ortopnea, il dolore addominale e la nausea — i sintomi clinici sono profondamente correlati tra loro. Un sistema bayesiano che li tratta come indipendenti commetterà errori sistematici in tutti i casi in cui queste correlazioni sono clinicamente rilevanti.

Questo problema fu riconosciuto dai ricercatori molto presto. E generò un dibattito vivace sulla natura stessa della diagnosi medica: era davvero un processo formalizzabile in termini

probabilistici? O c'era qualcosa nel ragionamento clinico — il riconoscimento di pattern complessi, la ponderazione del contesto, l'intuizione costruita su anni di esperienza — che sfuggiva alla matematica?

Il secondo problema era pratico: il sistema bayesiano aveva bisogno di dati. Tanti dati. Per calcolare le probabilità a priori e le probabilità condizionate con sufficiente precisione, occorrevo database di migliaia di pazienti con diagnosi certe e dati clinici completi e strutturati. Nel 1961, questi database non esistevano. Le cartelle cliniche erano di carta. I dati erano frammentati, incompleti, incompatibili tra ospedali diversi. Warner aveva costruito il suo sistema su un dataset locale, specifico, controllato. Ma generalizzarlo richiedeva un'infrastruttura di raccolta dati che il mondo medico degli anni Sessanta non era in grado di offrire.

Alla fine degli anni Sessanta, i due paradigmi — regole e probabilità — avevano entrambi mostrato le proprie promesse e i propri limiti insuperabili. Non si erano ancora incontrati. Ma entrambi stavano scavando gallerie separate verso lo stesso punto: l'idea che una macchina potesse essere utile al clinico nel momento critico della decisione.



## Capitolo III: Un nome per un'idea — la nascita del termine CDSS

### Nomenclatura come problema politico

Le grandi idee scientifiche trovano il loro nome spesso tardi, e quasi mai in modo ordinato. Prima arriva la cosa, poi arriva la parola. E la parola, quando arriva, porta con sé un carico di

implicazioni, di confini disciplinari, di rivendicazioni di paternità che non è mai neutro.

Così è stato per i Clinical Decision Support Systems — i CDSS. Il concetto aveva preceduto il nome di almeno un decennio. Quando Ledley e Lusted scrivevano nel 1959, non chiamavano il loro sistema un CDSS. Quando Warner costruiva il suo programma di diagnosi cardiologica nel 1961, non usava quella sigla. Quando Shortliffe sviluppava MYCIN nel 1972, il termine non esisteva ancora. I ricercatori usavano etichette diverse — «computer-assisted diagnosis», «medical expert systems», «clinical information systems» — ciascuna con sfumature diverse e, spesso, con implicazioni politiche altrettanto diverse.

La parola «diagnosi», in particolare, era quasi tabù. Dire che un computer poteva «fare diagnosi» sembrava un'usurpazione intollerabile della prerogativa medica, una dichiarazione di guerra all'identità professionale. Era meglio dire che il computer «assisteva» la diagnosi, che «supportava» la decisione — un linguaggio volutamente più morbido, più compatibile con l'ego dei clinici. Questa scelta retorica non era superficiale. Era una strategia di sopravvivenza per un campo che aveva bisogno della collaborazione dei medici per esistere.

### **Clement McDonald e i «clinical reminders»: la via pragmatica**

Uno dei precursori del concetto moderno di CDSS era un internista e informatico dell'Università dell'Indiana di nome Clement Joseph McDonald. Negli anni Settanta, McDonald aveva costruito a Indianapolis, presso il Regenstrief Institute for Health Care — un istituto di ricerca sull'assistenza sanitaria fondato dall'industriale Samuel Regenstrief con la precisa ambizione di rendere la medicina più efficiente attraverso la tecnologia — uno

dei primi sistemi informatici ospedalieri integrati del mondo: il Regenstrief Medical Record System.

Il Regenstrief non era soltanto un sistema di archiviazione. McDonald aveva cominciato a sperimentare quello che chiamava *clinical reminders*: avvisi automatici generati dal sistema per ricordare al medico azioni cliniche appropriate nel momento in cui visitava il paziente. Se il sistema sapeva che un paziente diabetico non aveva fatto l'emoglobina glicata negli ultimi sei mesi, lo segnalava. Se sapeva che un paziente iperteso non aveva misurato la creatinina nell'ultimo anno, lo ricordava. Non diagnosi — reminder. Non decisioni — suggerimenti.

McDonald pubblicò nel 1976 su *The Annals of Internal Medicine* uno studio randomizzato controllato — uno dei primissimi di quel tipo in informatica medica — che dimostrò in modo rigoroso che questi promemoria automatici miglioravano la pratica clinica: i medici che li ricevevano prescrivevano più spesso i farmaci indicati dalle linee guida, ordinavano più spesso gli esami di follow-up raccomandati. Era una dimostrazione empirica — non teorica — che il supporto informatico poteva cambiare il comportamento clinico in modo misurabile.

Il lavoro di McDonald era meno appariscente di MYCIN, meno ambizioso di INTERNIST-I. Ma era, in un senso importante, più vicino a quello che i CDSS sarebbero diventati: non sistemi che sostituiscono il ragionamento del medico, ma sistemi che lo integrano, che colmano i vuoti della memoria e dell'attenzione, che portano la linea guida al momento giusto nel posto giusto.

## La definizione che cambiò tutto

Il termine «clinical decision support system» come categoria nominale consolidata cominciò ad affermarsi negli anni Ottanta,

grazie al lavoro di diversi gruppi — a Vanderbilt, a Indiana, a Stanford — che cercavano un'etichetta comune per sistemi concettualmente affini ma tecnicamente diversissimi. La sfida era costruire una definizione abbastanza precisa da escludere ciò che non era un CDSS — le semplici banche dati, i sistemi amministrativi, le cartelle elettroniche di pura archiviazione — e abbastanza ampia da includere i diversi approcci che la ricerca stava sviluppando.

La definizione che gradualmente si impose identificava i CDSS come sistemi elettronici progettati per aiutare direttamente nel processo decisionale clinico, nei quali le caratteristiche dei singoli pazienti vengono utilizzate per generare valutazioni o raccomandazioni specifiche per il paziente. Tre elementi erano essenziali: l'elettronicità, che distingueva i CDSS dai protocolli cartacei; la specificità per il paziente, che li distingueva dalle linee guida generiche; la finalità decisionale, che li distingueva dai sistemi puramente informativi.

Questa definizione, apparentemente tecnica, aveva implicazioni politiche e scientifiche profonde. Includeva sistemi molto diversi tra loro: calcolatori di rischio, alert farmacologici, promemoria preventivi, sistemi di supporto alla diagnosi differenziale. La categoria era ampia — forse troppo — ma aveva il merito di mettere ordine in un campo dove la confusione terminologica era essa stessa un ostacolo alla ricerca e al dialogo tra comunità scientifiche diverse.

## **Il quadro di Wright e Sittig: quattro generazioni di architetture**

Nel 2008, Adam Wright e Dean Sittig pubblicarono sul *Journal of Biomedical Informatics* una revisione storica (Wright &

Sittig, 2008) che organizzava l'evoluzione del campo in quattro fasi architettoniche distinte.

La prima generazione era quella dei sistemi standalone — programmi isolati, come MYCIN, che funzionavano su singoli calcolatori senza connessione con il resto dell'ambiente informatico ospedaliero. La seconda era quella dell'integrazione nei sistemi clinici: i CDSS cominciarono a dialogare con le cartelle cliniche elettroniche, a leggere i dati del paziente direttamente dai sistemi informativi, a inserirsi nel flusso di lavoro quotidiano. La terza era quella degli standard per la condivisione dei contenuti: la creazione di linguaggi comuni, di formati di interscambio, di ontologie mediche condivise che permettessero ai diversi sistemi di comunicare tra loro. La quarta era quella dei modelli di servizio: i CDSS come componenti di un'infrastruttura distribuita, accessibili via rete, aggiornabili centralmente, condivisibili tra istituzioni diverse.

Il framework di Wright e Sittig era più di una storia. Era una mappa. Mostrava ai ricercatori dove si trovavano — nella seconda o terza generazione, per la maggior parte — e dove avrebbero dovuto andare. Ma c'era qualcosa che il framework non catturava completamente: la dimensione umana. L'evoluzione tecnica dei CDSS era avvenuta in un contesto fatto di culture professionali, resistenze organizzative, politiche sanitarie, interessi economici. E comprendere quella dimensione era essenziale per capire perché tanti sistemi brillanti erano rimasti nei laboratori senza raggiungere le corsie degli ospedali.



## Capitolo IV: MYCIN — il trionfo che nessuno vide (1972–1979)

### Un giovane studente a Stanford

Nell'autunno del 1972, un giovane medico di nome Edward Shortliffe iniziò il suo dottorato di ricerca in informatica alla Stanford University. Aveva ventiquattro anni, una laurea in medicina già in tasca e una domanda che lo tormentava: potevano le conoscenze di un esperto clinico essere formalizzate in un sistema informatico abbastanza robusto da essere utile nella pratica?

Shortliffe aveva scelto un problema preciso su cui testare questa ipotesi: la diagnosi e il trattamento delle malattie infettive batteriche, in particolare la sepsi e la meningite. Era un campo ideale per l'esperimento. Le infezioni batteriche gravi erano (e restano) una delle emergenze più drammatiche della medicina clinica, dove la scelta dell'antibiotico giusto nei primi minuti può fare la differenza tra la vita e la morte. Ma erano anche un campo dove le regole diagnostiche erano relativamente codificabili: quali batteri colpiscono in quali circostanze, quali antibiotici coprono quali spettri, quali controindicazioni esistono per quali pazienti.

**Edward H. Shortliffe (nato nel 1947)** — *Medico-informatico canadese, considerato uno dei padri dell'informatica biomedica moderna. Laureato in matematica applicata ad Harvard, si formò in medicina a Stanford — percorso raro per l'epoca, che lo pose all'incrocio esatto tra i due mondi. Il suo dottorato su MYCIN (1974) divenne un testo fondativo del campo. In seguito guidò il dipartimento di informatica biomedica della Columbia University e poi dell'Arizona, contribuendo a definire la biomedical informatics come disciplina accademica autonoma. Rimase per decenni una voce autorevole sul rapporto tra sistemi esperti e pratica clinica, scrivendo con lucidità sia dei successi che dei fallimenti del campo. MYCIN non fu mai implementato clinicamente — paradosso che Shortliffe stesso ha analizzato più volte — ma divenne il modello teorico da cui tutti i successivi sistemi esperti medici derivarono.*

Con la guida del suo supervisore, il logico Bruce Buchanan — uno dei pionieri dell'intelligenza artificiale applicata — Shortliffe lavorò per due anni a costruire quello che avrebbe chiamato

MYCIN: un sistema esperto basato su regole, capace di identificare l'agente infettivo responsabile di una sepsi e di raccomandare la terapia antibiotica appropriata.

## Come funzionava MYCIN

Il cuore di MYCIN era una base di conoscenza di circa seicento regole IF-THEN, ciascuna delle quali codificava un frammento del ragionamento clinico di un esperto di malattie infettive. La logica era lineare ma potente. Se il paziente ha febbre alta e il liquor mostra pleocitosi<sup>1</sup> e l'esame colturale preliminare suggerisce gram-positivi<sup>2</sup>, allora considera lo *Streptococcus pneumoniae* come agente causale con una certezza dell'85%.

MYCIN non lavorava con certezze assolute. Utilizzava un sistema di fattori di certezza — numeri tra -1 e 1 che rappresentavano il grado di fiducia in ciascuna regola — per propagare l'incertezza attraverso il ragionamento, proprio come fa un medico esperto quando dice «probabilmente è una polmonite batterica, ma non possiamo escludere una virale».

Il sistema interagiva col medico attraverso un terminale testuale: faceva domande, riceveva risposte, applicava le proprie regole e alla fine presentava le sue raccomandazioni con una spiegazione del percorso logico seguito. Poteva, se richiesto, giustificare ogni sua conclusione risalendo la catena di regole che l'aveva prodotta — una funzione che Shortliffe chiamava «why» e che anticipava di decenni il concetto moderno di explainable AI<sup>3</sup>.

---

<sup>1</sup>Pleocitosi: aumento del numero di cellule (leucociti) nel liquido cerebrospinale. Valore normale < 5 cellule/ $\mu$ L. In presenza di infezioni batteriche del sistema nervoso centrale il numero può salire a migliaia.

<sup>2</sup>Gram-positivi: batteri che conservano il colorante viola nel test di Gram (es. *Streptococcus*, *Staphylococcus*). Gram-negativi: batteri che non lo trattengono e appaiono rosa (es. *E. coli*, *Pseudomonas*). La distinzione è clinicamente fondamentale perché orienta la scelta dell'antibiotico.

<sup>3</sup>Explainable AI (XAI): branca dell'intelligenza artificiale che sviluppa sistemi in grado di spiegare il proprio ragionamento in modo comprensibile agli esseri umani. Si contrappone

## La valutazione del 1979: meglio degli specialisti

Nel 1979, un gruppo di ricercatori coordinato da Victor Yu pubblicò su JAMA uno studio che avrebbe fatto tremare le fondamenta della medicina: una valutazione in cieco delle raccomandazioni di MYCIN confrontate con quelle di esperti umani di malattie infettive.

Il risultato era inequivocabile. Il tasso di accettabilità delle terapie suggerite da MYCIN — valutato da un panel indipendente di specialisti che non sapevano se le raccomandazioni venissero da un uomo o da una macchina — era del 65%. Un numero che superava quello di quattro su cinque clinici umani inclusi nello studio come termini di confronto. La macchina era, almeno in questo contesto, più brava degli esperti.

La notizia si diffuse rapidamente attraverso i dipartimenti di informatica e di medicina di mezzo mondo. MYCIN era diventato famoso.

## Il paradosso: celebre ma inutilizzato

Eppure MYCIN non fu mai implementato nella pratica clinica. Mai. Nemmeno al Stanford University Hospital, dove era stato sviluppato. E qui si cela uno dei paradossi più istruttivi di tutta la storia della medicina digitale: come può un sistema che dimostra di essere più bravo degli esperti umani non venire mai utilizzato?

Le ragioni sono molteplici, e ciascuna è, a suo modo, rivelatrice.

Prima ragione: l'infrastruttura non esisteva. Nel 1979, un ospedale era fatto di corridoi, di letti, di carrelli porta farmaci e di telefoni fissi appesi ai muri. Non c'erano terminali informatici nelle

---

ai modelli «scatola nera» (black box) che producono risultati corretti ma non interpretabili.

corsie. Non c'erano reti locali. Non c'erano cartelle cliniche elettroniche. Per usare MYCIN, un medico avrebbe dovuto sedersi davanti a uno dei rari terminali del centro di calcolo dell'università, digitare manualmente tutti i dati del paziente — nome, età, sintomi, esami di laboratorio — e poi aspettare. Un processo che avrebbe richiesto venti, trenta minuti. In un pronto soccorso dove ogni minuto conta, era un lusso insostenibile.

Seconda ragione: il sistema era isolato. MYCIN non parlava con nessun altro sistema. Non si interfacciava con il laboratorio, non leggeva le cartelle cliniche, non aveva accesso alle radiologie. Ogni dato che usava doveva essere inserito a mano. Un consulente umano, al contrario, arrivava in reparto, guardava il paziente, sfogliava la cartella, telefonava al laboratorio, e in venti minuti aveva un quadro completo. MYCIN, per avere lo stesso quadro, richiedeva ore di inserimento dati.

Terza ragione: il contesto culturale e medico-legale. Chi era responsabile se MYCIN sbagliava? Il medico che aveva seguito la raccomandazione del sistema? Il programmatore che aveva codificato le regole? L'università che aveva sviluppato il software? Nel 1979 non esisteva nessuna risposta giuridica a queste domande. E in assenza di risposta, i medici preferivano non rischiare.

Quarta ragione, forse la più sottile: l'orgoglio professionale. Ammettere di consultare una macchina per decidere quale antibiotico prescrivere era, per molti medici di quella generazione, una dichiarazione implicita di incompetenza. Il grande clinico non aveva bisogno di aiuto. Aveva bisogno di un computer tanto quanto aveva bisogno che qualcuno gli spiegasse come allacciarsi le scarpe.

E allora come ha fatto MYCIN a rimanere nella storia? Perché è ancora citato oggi, in ogni corso di informatica medica, in ogni libro di testo sull'intelligenza artificiale in medicina?

Perché MYCIN ha dimostrato qualcosa di fondamentale: che era possibile. Che la conoscenza clinica poteva essere codificata. Che un sistema informatico poteva ragionare in modo medicamente valido. Che la diagnosi e la terapia non erano magie inaccessibili ma processi logici formalizzabili. MYCIN non è sopravvissuto come strumento clinico. È sopravvissuto come prova di concetto. Come il primo passo di un lunghissimo cammino.



## Capitolo V: INTERNIST-I — l'ambizione di conoscere tutto (1974–1982)

### Jack Myers, il grande clinico di Pittsburgh

Se MYCIN era l'opera di un giovane studente appassionato di logica artificiale, INTERNIST-I era qualcosa di completamente diverso. Era il sogno di un vecchio medico che voleva immortalare trent'anni di esperienza in una macchina.

Jack Duane Myers era, negli anni Settanta, uno dei clinici più rispettati degli Stati Uniti. Professore di medicina interna all'Università di Pittsburgh, era conosciuto da colleghi e studenti come «The Professor» — un titolo che in quel dipartimento non era un vezzo accademico ma un riconoscimento quasi reverenziale. Myers aveva il dono raro dei grandi diagnostici: la capacità di tenere in mente centinaia di diagnosi simultaneamente, di percepire pattern nascosti, di navigare la complessità della

medicina interna — quella specialità vasta e caotica che comprende tutto ciò che non rientra in nessuna altra categoria.

Negli anni Settanta, Myers incontrò Harry Pople, un informatico del Carnegie Mellon famoso per il suo lavoro sui sistemi di risoluzione dei problemi. Insieme decisero di fare qualcosa di straordinario: trascrivere la mente diagnostica di Myers in un calcolatore.

Il progetto prese il nome di INTERNIST-I. E se MYCIN era un sistema relativamente circoscritto — malattie infettive, antibiotici — INTERNIST-I aveva ambizioni senza precedenti: essere in grado di formulare diagnosi nell'intera medicina interna generale.

## Il progetto enciclopedico

Per anni, Myers sedette con Pople e con il suo collaboratore Randolph Miller e dettò. Dettò diagnosi, sintomi, relazioni causali, epidemiologie, controindicazioni. La base di conoscenza di INTERNIST-I arrivò a contenere oltre cinquecento malattie e tremila manifestazioni cliniche, con relazioni complesse che le collegavano in una rete di dipendenze diagnostiche.

Era un progetto enciclopedico nel senso letterale del termine: un tentativo di raccogliere in un unico sistema tutto il sapere clinico di un'epoca. Ogni mattina Myers arrivava nel suo studio, e ogni mattina Pople era lì con i suoi nastri magnetici e i suoi modelli computazionali, pronto a tradurre il gergo clinico in strutture logiche che una macchina potesse elaborare.

La pubblicazione su *New England Journal of Medicine* nel 1982 (Miller, Pople & Myers, 1982) fu un evento. Il titolo era sobrio: INTERNIST-1, An Experimental Computer-Based Diagnostic

Consultant for General Internal Medicine. Ma il contenuto era esplosivo.

## **Il confronto con i discussant del Massachusetts General Hospital**

Per testare INTERNIST-I, i ricercatori usarono una serie di casi clinici pubblicati nel famoso Case Records del Massachusetts General Hospital — la rubrica settimanale del NEJM in cui i casi clinici più difficili e insoliti vengono discussi da esperti internazionali. Era, ed è, il Campionato del Mondo della diagnostica clinica.

I risultati furono, per lo stesso team di sviluppo, deludenti. INTERNIST-I riusciva a identificare la diagnosi corretta come prima scelta in poco più della metà dei casi. I discussant umani, quei medici di élite che affrontavano i casi più rari e complessi del mondo, andavano molto meglio. Il sistema aveva difficoltà particolari con i casi che richiedevano ragionamento anatomico — dove era localizzata la lesione? — o temporale — quando era insorto il problema? — e quasi nessuna capacità di costruire diagnosi differenziali che attraversassero più organi o sistemi simultaneamente.

Il sogno di Myers era, almeno in parte, svanito. Non si poteva mettere l'intera medicina interna in una macchina. O almeno, non con la tecnologia e i modelli del 1982.

## **Perché anche INTERNIST-I rimase un esperimento**

Come MYCIN, INTERNIST-I non fu mai adottato su larga scala nella pratica clinica. Le ragioni si sovrapponevano parzialmente: infrastruttura inesistente, interfaccia utente

elementare, difficoltà di integrazione con i sistemi ospedalieri. Ma c'era anche un problema più profondo.

La medicina interna è troppo vasta, troppo sfumata, troppo dipendente dal contesto per essere catturata in una base di regole, per quanto grande. Un paziente non è un caso. È una storia, un corpo, un insieme di aspettative, di paure, di condizioni sociali e culturali che influenzano ogni sintomo e ogni segno. INTERNIST-I poteva ragionare sulla febbre, sulla dispnea, sull'ittero — ma non poteva guardare negli occhi un paziente e capire se stava mentendo sui propri sintomi, se aveva paura, se c'era qualcosa che non diceva.

Eppure, come MYCIN, INTERNIST-I rimase nella storia. Non perché fosse stato usato, ma perché aveva mostrato che era possibile tentare. Aveva aperto una strada. Aveva dimostrato che il problema della codifica della conoscenza clinica era difficile — molto più difficile di quanto i pionieri avessero immaginato — ma non impossibile.



## Capitolo VI: La grande delusione e il lungo inverno (Fine anni '70 – inizio anni '80)

Alla fine degli anni Settanta, l'ottimismo delle origini aveva lasciato il posto a qualcosa di diverso. Non cinismo — i ricercatori più seri continuavano a credere nel progetto — ma sobrietà. Una consapevolezza nuova e un po' dolorosa che il problema era più complesso di quanto si pensasse.

Gli approcci principali disponibili in quel momento erano due. Il primo era il sistema basato su regole — quello di MYCIN e di

INTERNIST-I — dove la conoscenza viene codificata in proposizioni logiche del tipo se... allora... Il secondo era il riconoscimento di pattern statistici, dove il computer impara ad associare certi insiemi di dati a certi esiti sulla base di esempi storici.

Entrambi gli approcci avevano mostrato, in contesti controllati, risultati promettenti. Ma entrambi si erano scontrati, quando si tentava di portarli nella pratica clinica reale, con una serie di ostacoli che sembravano invalicabili.

I sistemi basati su regole erano fragili: funzionavano bene all'interno del loro dominio di competenza, ma andavano completamente fuori strada al di fuori di esso. Se MYCIN incontrava un caso che le sue seicento regole non coprivano, non poteva ammettere la propria ignoranza. Produceva raccomandazioni comunque, costruite su premesse insufficienti, con una sicurezza apparente che era, in quel caso, pura illusione.

I sistemi di riconoscimento di pattern richiedevano grandi quantità di dati storici. Ma nel 1980 quei dati non esistevano in forma digitale. Le cartelle cliniche erano di carta. Gli esami di laboratorio erano su moduli stampati. Le radiografie erano pellicole fisiche in cassette di metallo. Non c'era niente da cui un computer potesse imparare.

E poi c'era il grande assente: l'infrastruttura. I computer erano ancora macchine enormi, costose, chiuse nei centri di elaborazione. La rete Internet era un progetto militare segreto. Collegare un sistema di supporto decisionale alle cartelle cliniche di un reparto era, tecnicamente, un'impresa titanica.

In questo contesto, molti fondi per la ricerca in intelligenza artificiale medica si prosciugarono. I programmatori più brillanti si spostarono verso applicazioni commerciali più redditizie. Quello

che i ricercatori chiamarono, con amara ironia, il primo «inverno dell'IA» calò sui laboratori di informatica biomedica.



## Capitolo VII: La ripresa silenziosa — l'era delle cartelle cliniche elettroniche (anni '80–2000)

### Dal centro di calcolo al reparto

La ripresa arrivò in modo silenzioso, senza un manifesto, senza un articolo fondativo sul *New England Journal*. Arrivò grazie a qualcosa di molto prosaico: i computer diventarono piccoli.

L'avvento del personal computer nella prima metà degli anni Ottanta cambiò tutto. Non subito, non in modo rivoluzionario, ma inesorabilmente. I medici e gli amministratori ospedalieri cominciarono a vedere che una macchina grande come una valigia poteva stare su una scrivania. Che non richiedeva una stanza climatizzata e un team di tecnici. Che poteva essere usata da persone comuni, non soltanto da programmatori con un dottorato.

In questo nuovo contesto tecnologico, nacquero i primi sistemi informativi ospedalieri. Non erano CDSS nel senso sofisticato degli esperimenti degli anni Settanta — non ragionavano, non diagnosticavano — ma raccoglievano dati: demografici, di laboratorio, di terapia, di fatturazione. E raccogliendo dati, preparavano il terreno per qualcosa di molto più ambizioso.

Una revisione sistematica pubblicata su *JAMA* nel 1998 (Hunt et al., 1998) documentò 68 trial controllati sull'efficacia dei CDSS — 40 dei quali comparsi dopo il 1992. Era il segnale di una

rinascita. Il campo si stava riaprendo, questa volta con una base empirica più solida e con un'infrastruttura finalmente disponibile.

## L'HITECH Act: la svolta normativa

Nel 2009, il Congresso degli Stati Uniti approvò l'HITECH Act — Health Information Technology for Economic and Clinical Health Act. Erano i mesi bui della crisi finanziaria globale, ma Barack Obama, da poco insediato alla Casa Bianca, aveva inserito nella legge di stimolo economico un investimento da miliardi di dollari per promuovere l'adozione delle cartelle cliniche elettroniche negli ospedali e negli ambulatori americani.

La logica era semplice: le cartelle cliniche digitali avrebbero ridotto gli errori, migliorato l'efficienza, abbassato i costi. Ma c'era anche una conseguenza secondaria, forse non del tutto prevista nella sua portata: le cartelle cliniche elettroniche generavano dati strutturati. Enormi quantità di dati strutturati. E i dati strutturati erano il carburante che i sistemi di supporto decisionale aspettavano da decenni.

Nella decade successiva all'HITECH Act, la diffusione delle EMR<sup>4</sup> raggiunse oltre l'80% degli ospedali americani. Per la prima volta nella storia, esisteva una infrastruttura digitale capace di alimentare sistemi clinici intelligenti su scala nazionale.



---

<sup>4</sup>EMR — Electronic Medical Record (cartella clinica elettronica): sistema informatico per la raccolta, l'archiviazione e la gestione della documentazione clinica del paziente in forma digitale strutturata. Prerequisito essenziale per qualsiasi sistema di supporto decisionale integrato nel flusso di lavoro.

## Capitolo VIII: La rivoluzione del machine learning (2010–2020)

### Quando le macchine smettono di seguire regole e cominciano ad imparare

C'è una differenza fondamentale tra il MYCIN degli anni Settanta e i sistemi di intelligenza artificiale che cominciarono ad affacciarsi in medicina nella seconda decade del Duemila. Una differenza che non è soltanto tecnica ma quasi filosofica.

MYCIN e i suoi successori basati su regole funzionavano perché un essere umano — un esperto clinico — aveva spiegato loro come funzionare. Ogni regola era stata scritta, pensata, approvata da un medico. Il sistema sapeva quello che qualcuno gli aveva insegnato, e nient'altro.

I sistemi di machine learning<sup>5</sup> funzionano diversamente. Non hanno bisogno che qualcuno spieghi loro le regole. Le imparano da soli, analizzando migliaia, milioni di esempi. Mostri a un algoritmo di machine learning diecimila elettrocardiogrammi, digli quale diagnosi è associata a ciascuno, e il sistema — da solo, senza che nessuno gli abbia spiegato come leggere un ECG — imparerà a riconoscere i pattern che distinguono il normale dal patologico.

Questa capacità di apprendere dai dati, invece che da regole esplicite, cambiò radicalmente il paesaggio dei CDSS. Improvvisamente, diventò possibile costruire sistemi capaci di affrontare problemi che nessun esperto umano avrebbe saputo codificare in regole — perché erano troppo complessi, troppo dipendenti da sottili interazioni tra variabili, troppo nascosti nella struttura statistica di grandi dataset.

<sup>5</sup>Machine learning (apprendimento automatico): insieme di tecniche che permettono a un sistema informatico di imparare automaticamente dai dati, senza essere esplicitamente programmato per ogni singolo caso. Il sistema migliora le proprie prestazioni man mano che elabora più esempi.

## I cinque modi di mescolare regole ed esperienza

I ricercatori scoprirono presto che la dicotomia tra regole e machine learning era, nella pratica, troppo rigida. I sistemi reali mescolavano i due approcci in modi creativi e diversi.

Pensa a come lavora un grande chef. Da un lato ha le regole imparate a scuola — temperature di cottura, combinazioni di sapori vietate, norme igieniche da rispettare sempre. Dall'altro ha l'esperienza accumulata negli anni: ha cucinato migliaia di piatti, ha imparato a riconoscere un pesce fresco dall'odore, una salsa al punto giusto dal colore. Nessuno chef lavora solo con le regole o solo con l'esperienza: li combina, in modi diversi a seconda della situazione.

I CDSS ibridi fanno la stessa cosa. Una revisione del 2023 (Kierner et al., 2023) identificò cinque modi principali in cui regole e machine learning possono essere combinati. In alcuni sistemi le regole fanno da guardrail all'interno di un modello che apprende — come un cuoco esperto che può improvvisare liberamente ma non userà mai ingredienti scaduti. In altri il machine learning prepara e filtra i dati, e poi le regole prendono la decisione finale — come una macchina che lava e taglia le verdure, ma è lo chef a decidere cosa cucinare. In altri ancora è il contrario: le regole selezionano i casi rilevanti, e il machine learning lavora solo su quelli — come un selezionatore umano che sceglie i candidati da esaminare, e poi un algoritmo li valuta. Esistono anche sistemi in cui le regole non partecipano alla decisione finale ma guidano il processo di apprendimento — come un insegnante che corregge gli esercizi dello studente secondo criteri precisi. E infine la variante più rara e più trasparente: regole e machine learning lavorano in parallelo come due consulenti indipendenti, e il medico vede entrambe le loro opinioni e decide.

Quest'ultima architettura è la preferita da chi si preoccupa di rendere l'AI comprensibile ai clinici: quando le due «voci» sono visibili e separate, il medico può confrontarle, capire dove concordano e dove divergono, e decidere a quale fidarsi di più in quel contesto specifico.

#### I cinque archetipi — nomi tecnici per chi vuole approfondire

**REML — Rules Embedded in Machine Learning:** Le regole sono incorporate dentro il modello di ML come vincoli o conoscenza iniziale. Il modello è libero di imparare, ma non può violare certe regole fondamentali.

**MLRB — Machine Learning → Rule-Based:** Il ML elabora prima i dati grezzi ed estrae le caratteristiche rilevanti (feature). Poi un motore di regole usa quelle caratteristiche per prendere la decisione finale.

**RBML — Rule-Based → Machine Learning:** Le regole filtrano e strutturano i dati in ingresso. Il ML lavora solo sui casi e le variabili selezionati dalle regole.

**RMLT — Rules guide ML Training:** Le regole non partecipano alla decisione, ma guidano il processo di addestramento del modello, orientando ciò che il sistema impara.

**PERML — Parallel Ensemble of Rules and ML:** Regole e ML lavorano in modo completamente indipendente e parallelo. Le loro raccomandazioni vengono poi confrontate e riconciliate. È l'architettura più trasparente e più rara.



## Capitolo IX: I Grandi Modelli Linguistici — la sorpresa del ragionamento emergente

### Un cervello addestrato a prevedere la prossima parola

Nel novembre del 2022, OpenAI rilasciò al pubblico un sistema chiamato ChatGPT. Era un'interfaccia di conversazione basata su un Large Language Model — un Grande Modello Linguistico addestrato su quantità di testo così vaste da essere difficili da concepire: l'intera Wikipedia, milioni di libri, miliardi di pagine web, codice sorgente, articoli scientifici, conversazioni online. Il principio tecnico di base era, almeno in apparenza, semplice: il modello imparava a prevedere quale parola venisse più probabilmente dopo una sequenza data. Un gioco statistico di

anticipazione linguistica, enormemente sofisticato ma concettualmente lineare.

Quello che nessuno si aspettava — non del tutto, non con quella chiarezza — era quello che accadde dopo. Il sistema addestrato a prevedere parole non aveva imparato soltanto la lingua. Aveva imparato qualcosa di molto più profondo: la struttura del pensiero umano codificata in quella lingua. Aveva imparato a ragionare.

I ricercatori chiamarono questo fenomeno *emergent abilities* — capacità emergenti. Capacità che non erano state esplicitamente programmate, che non erano presenti nei modelli più piccoli, che sembravano comparire quasi improvvisamente quando il modello raggiungeva una certa soglia di dimensione e di dati di addestramento. La più sorprendente di queste capacità era il trasferimento di conoscenza tra domini completamente diversi: un modello addestrato su testo generico poteva ragionare su un caso clinico, scrivere codice informatico, risolvere problemi di matematica, comporre poesia, spiegare la relatività speciale a un bambino di dieci anni — tutto con una fluidità che sembrava impossibile per un sistema che, in fondo, aveva imparato soltanto a prevedere la prossima parola.

## La logica come proprietà emergente

La sorpresa più grande non era la versatilità. Era la logica. I Grandi Modelli Linguistici non si limitavano a riprodurre informazioni che avevano visto nei dati di addestramento. Generavano ragionamenti nuovi. Costruivano argomenti. Identificavano contraddizioni. Trasferivano principi da un contesto all'altro con una precisione che nessuna tecnica precedente aveva avvicinato — né i sistemi basati su regole degli anni Settanta, né i

modelli statistici degli anni Novanta, né le prime reti neurali degli anni Duemila.

Un articolo pubblicato da ricercatori di Microsoft nel 2023 (Bubeck et al., 2023), intitolato provocatoriamente Sparks of Artificial General Intelligence, esaminò le capacità di GPT-4 in una serie di test progettati per valutare facoltà cognitive che gli esseri umani considerano distintive: ragionamento analogico, comprensione del contesto, soluzione di problemi complessi, creatività. Il modello mostrava prestazioni sorprendentemente vicine a quelle umane in molti di questi test — e in alcuni le superava.

Per la medicina, e per i CDSS in particolare, questa capacità di ragionamento logico ha un'importanza specifica. La diagnosi differenziale è, nella sua struttura profonda, un esercizio di logica: dati certi fatti clinici, quali ipotesi sono compatibili? Quali possono essere eliminate? Quale percorso diagnostico è più efficiente? Sono domande a cui i sistemi basati su regole degli anni Settanta cercavano di rispondere codificando la logica esplicitamente. I Grandi Modelli Linguistici la generano — con tutte le opportunità e tutti i rischi che questa differenza comporta.

## **Il problema delle allucinazioni — e come è cambiato**

Chiunque abbia usato ChatGPT nei primi mesi dopo il suo rilascio ricorda la scoperta delle allucinazioni. Il modello inventava riferimenti bibliografici inesistenti. Citava studi mai pubblicati, autori mai esistiti, esperimenti mai condotti — tutto con la stessa scioltezza sicura con cui avrebbe descritto fatti reali. Per la medicina, questo era un problema serio: un sistema che genera con la stessa fluidità verità e falsità è intrinsecamente pericoloso in un contesto clinico.

Nel 2026, la situazione è significativamente diversa. Non perfetta — le allucinazioni non sono scomparse — ma molto migliorata, grazie a una combinazione di tecniche sviluppate negli ultimi tre anni. La più importante è il Retrieval-Augmented Generation, o RAG<sup>6</sup>: invece di affidarsi esclusivamente alla conoscenza memorizzata durante l'addestramento, il sistema recupera in tempo reale informazioni da fonti esterne verificate, e basa le proprie risposte su queste fonti citandole esplicitamente. Il risultato è verificabile: se il modello cita uno studio, lo studio esiste e dice quello che il modello afferma. Piattaforme come OpenEvidence — usata per preparare il materiale bibliografico di questo saggio — applicano esattamente questo principio.

A questo si aggiungono il Reinforcement Learning from Human Feedback<sup>7</sup>, che allena il modello a rispondere in modo più accurato e a riconoscere la propria incertezza, e sistemi di constitutional AI<sup>8</sup> che incorporano principi di onestà e di calibrazione dell'episteme nel comportamento di base del modello. I sistemi del 2026 sbagliano meno, riconoscono meglio ciò che non sanno, e quando sbagliano lo fanno in modo meno convincentemente sicuro. Non è la soluzione definitiva. Ma è un progresso reale e sostanziale.

---

<sup>6</sup>RAG — Retrieval-Augmented Generation: tecnica che consente a un modello linguistico di recuperare informazioni aggiornate da fonti esterne verificate al momento della risposta, invece di affidarsi solo alla conoscenza acquisita durante l'addestramento. Riduce significativamente le «allucinazioni».

<sup>7</sup>RLHF — Reinforcement Learning from Human Feedback: metodo di addestramento in cui il modello viene valutato e corretto da valutatori umani, che segnalano quali risposte sono migliori. È una delle tecniche chiave per rendere i modelli linguistici più accurati, onesti e calibrati.

<sup>8</sup>Constitutional AI: approccio sviluppato da Anthropic per addestrare modelli AI in modo che rispettino automaticamente un insieme di principi etici espliciti (“costituzione”), senza dipendere esclusivamente dalla supervisione umana diretta per ogni risposta.

## Il cervello dietro l'AI non parla inglese come madrelingua

C'è un aspetto della storia dell'intelligenza artificiale moderna che raramente viene raccontato: l'AI non è americana. Non nel senso che conta.

Le grandi aziende che guidano oggi lo sviluppo dei Grandi Modelli Linguistici hanno sede negli Stati Uniti. Ma i cervelli che hanno costruito questi sistemi vengono da tutto il mondo. Mustafa Suleyman, CEO di Microsoft AI e co-fondatore di DeepMind, è nato a Londra nel 1984 da padre siriano — un taxista — e madre inglese, un'infermiera. Ha abbandonato Oxford a diciannove anni per fondare una helpline per la salute mentale dei giovani. È uno degli uomini che più di chiunque altro ha contribuito a rendere possibile l'AI moderna.

Dario Amodei, CEO di Anthropic, porta un cognome che racconta un'italianità immigrata — quella diaspora che ha contribuito a costruire l'America scientifica del Novecento. Yann LeCun, padre delle reti neurali convoluzionali e Chief AI Scientist di Meta, è francese. Geoffrey Hinton, il 'padrino del deep learning', è britannico, e ha lasciato Google nel 2023 per parlare liberamente dei rischi dell'AI. Demis Hassabis, CEO di Google DeepMind, è britannico con radici greco-cipriote e singaporiane. E migliaia dei ricercatori che ogni giorno fanno avanzare il campo sono indiani, cinesi, coreani, europei dell'est.

Gli Stati Uniti non hanno creato l'intelligenza artificiale. Hanno creato le condizioni per attrarre e finanziare i talenti che la stavano creando altrove. Hanno saputo costruire un ecosistema — di capitali di rischio, di università di ricerca, di cultura dell'imprenditorialità tecnologica — capace di trasformare le idee di un siriano, di un greco-cipriota, di un francese, di un canadese in prodotti che oggi usano miliardi di persone. Il futuro dell'AI non

appartiene a nessuna nazione in particolare. Appartiene a chi saprà costruire le condizioni migliori per il talento globale.



## Capitolo X: Il campo di battaglia invisibile — l'alert fatigue

### Quando la guardia del sistema diventa il suo tallone d'Achille

C'è un fenomeno che chiunque lavori in un ospedale moderno conosce bene, anche se raramente lo chiama con il suo nome tecnico. Accade così: il medico sta ordinando una terapia, e il sistema informatico lancia un avviso — un pop-up, un messaggio rosso sullo schermo, un campanello elettronico. Potenziale interazione farmacologica. Controindicazione rilevata. Allerta per valore di laboratorio critico.

Il medico guarda l'avviso per una frazione di secondo. Lo clicca via. Continua l'ordine.

Questo comportamento, replicato decine di volte al giorno, centinaia di volte alla settimana, è la cosiddetta alert fatigue — la stanchezza da allarme. E rappresenta uno dei paradossi più inquietanti dei CDSS moderni: un sistema progettato per proteggere i pazienti che, a causa della propria eccessiva prolissità, finisce per essere sistematicamente ignorato.

I numeri sono sconcertanti. Gli studi in letteratura riportano che tra il 33% e il 96% degli alert clinici vengono ignorati dai medici. Alcune analisi in grandi ospedali americani hanno trovato che oltre il 90% degli alert di interazione farmacologica vengono cliccati via senza che venga presa nessuna azione. Il sistema urla continuamente, e i clinici hanno imparato a non sentirlo.

## Le strategie per uscire dal loop

I ricercatori hanno sviluppato un repertorio di strategie per ridurre l'alert fatigue senza perdere la funzione protettiva dei sistemi di supporto decisionale.

La più elegante è anche la più ovvia: classificare gli alert per livello di urgenza. Non tutti gli avvisi meritano di interrompere il flusso di lavoro del medico. Una potenziale interazione farmacologica che riduce dell'8% l'efficacia di un farmaco secondario può essere segnalata passivamente, in un angolo dello schermo, senza bloccare l'ordine. Un'incompatibilità che rischia la vita del paziente deve invece fermare tutto e richiedere una risposta esplicita.

Una seconda strategia è il role tailoring: adattare gli alert al ruolo specifico dell'utente. Un farmacista ha bisogno di vedere informazioni diverse da quelle che servono a un medico di guardia, che a sua volta ha esigenze diverse da quelle di un'infermiera. Un sistema che manda lo stesso alert a tutti è inevitabilmente rumoroso per molti e silente per altri.

La strategia più sofisticata — e più recente — è quella dei CDSS statisticamente prioritizzati e contestualizzati, proposta da Chazard e colleghi nel 2020: sistemi che usano i dati storici del singolo reparto per calcolare la probabilità reale che un determinato alert sia clinicamente rilevante nel contesto specifico in cui viene generato. Un avviso che viene ignorato il 97% delle volte in un reparto di cardiologia con pazienti anziani polipatologici è probabilmente un avviso mal calibrato per quel contesto, e può essere abbassato automaticamente di priorità.



## Capitolo XI: I CDSS entrano in sala parto (2019–2026)

### L'ostetricia come terreno privilegiato

Se dovessimo scegliere una specialità medica in cui l'intelligenza artificiale applicata alle decisioni cliniche può produrre i risultati più drammatici — nel senso più letterale del termine, come in un dramma che si svolge in tempo reale con conseguenze immediate sulla vita di due esseri umani — quella specialità sarebbe l'ostetricia.

Da un lato, la gravidanza genera una quantità enorme di dati strutturati — ecografie, esami del sangue, monitoraggio fetale, visite periodiche — esattamente il materiale grezzo di cui i sistemi di machine learning hanno bisogno per imparare. Dall'altro, le complicanze ostetriche si presentano spesso in modo subdolo, evolvono con rapidità imprevedibile, e non ammettono margini di errore. La finestra terapeutica è stretta. Gli errori si pagano caro. E nessuna altra specialità porta il peso simultaneo di due vite le cui esigenze possono, nei momenti critici, entrare in conflitto.

C'è però un terzo elemento che rende l'ostetricia particolarmente significativa per la storia dei CDSS: nessun'altra branca della medicina ha tentato così a lungo, così sistematicamente, e con risultati così contrastanti, di informatizzare il processo decisionale clinico in tempo reale. La storia di questo tentativo — che dura da oltre cinquant'anni e non ha ancora trovato una risposta definitiva — è una delle più istruttive di tutto il campo dei sistemi di supporto alle decisioni.



## Il CTG — la promessa più antica

Nella premessa di questo saggio ho raccontato come, alla fine degli anni Novanta, io e il professor Serra costruimmo uno score per l'interpretazione del tracciato cardiocografico. Lo facemmo per rendere il ragionamento clinico esplicito e documentabile, e per liberare il medico dall'ansia della giustificazione a posteriori. Non sapevamo, allora, che stavamo ripercorrendo un sentiero che altri avevano già battuto — e che continuavano a battere, con risultati sorprendentemente ambivalenti.

Il monitoraggio elettronico fetale continuo — la cardiocografia, il CTG — è oggi lo strumento di sorveglianza intrapartum più diffuso al mondo. Negli Stati Uniti viene utilizzato in circa il 90% dei parti; in Italia, come in tutta Europa, la situazione è analoga. Eppure la promessa con cui fu introdotto — che il monitoraggio continuo della frequenza cardiaca fetale avrebbe ridotto l'incidenza dell'ipossia fetale e delle sue sequele neurologiche — non è stata pienamente realizzata. Una qualche alterazione del tracciato si verifica in oltre la metà di tutti i travagli. Se ogni alterazione portasse a un cesareo, i tassi operatori sarebbero insostenibili. Se ogni alterazione venisse ignorata, qualche bambino avrebbe meno tempo di quanto pensassimo. Il CDSS per il CTG è dunque, fin dall'origine, un problema di calibrazione estrema tra sensibilità e specificità in un contesto di altissima pressione temporale.

## Il problema dell'interpretazione: quando gli esperti non sono d'accordo

Prima ancora di chiedersi se una macchina possa interpretare meglio il CTG, bisogna confrontarsi con un dato scomodo: gli esseri umani lo interpretano in modo notevolmente inconsistente. Una revisione sistematica del 2023 di Brown e colleghi ha confrontato

13 linee guida internazionali per l'interpretazione del CTG intrapartum, trovando differenze significative nelle raccomandazioni chiave. Uno studio del 2000 di Devoe e colleghi ha dimostrato che, anche quando istruiti a utilizzare le linee guida NICHD, i livelli di accordo inter-osservatore per la maggior parte delle caratteristiche del CTG non superavano quelli attesi per caso — con valori che variavano dal 45% al 99%, i più bassi proprio per il conteggio di accelerazioni e decelerazioni, le caratteristiche clinicamente più rilevanti.

Un'analisi qualitativa del 2024 di Lamé e colleghi ha descritto il monitoraggio CTG come un'attività sociotecnica complessa, in cui compiti, persone, strumenti e fattori organizzativi si combinano per determinare la sicurezza. Non un atto medico individuale, ma un processo collettivo intrinsecamente variabile. Questa variabilità è al tempo stesso la ragione d'essere di qualsiasi CDSS e il suo principale ostacolo: senza un gold standard riconosciuto, non esiste il punto di riferimento su cui addestrare e validare un algoritmo.



## **Trent'anni di sistemi computerizzati: SisPorto e INFANT**

### **SisPorto: il successo che nessuno si aspettava**

Il sistema SisPorto rappresenta il tentativo più longevo e strutturato di informatizzare l'analisi del CTG. Sviluppato da Diogo Ayres-de-Campos e João Bernardes all'Università di Porto — con trentadue anni di sviluppo continuo documentati da Bernardes nel 2023 — il sistema segue le linee guida FIGO per il monitoraggio fetale, analizza simultaneamente segnali fetali e materni, e può essere integrato con l'analisi del segmento ST.

Uno studio osservazionale del 2019 di Lopes-Pereira e colleghi su 38.466 parti ha dimostrato che l'introduzione dell'analisi computerizzata tramite il sistema Omniview-SisPorto era associata a una riduzione significativa dell'encefalopatia ipossico-ischemica: da 5.3 a 2.2 per 1.000 nati vivi, con un rischio relativo di 0.42. E — fatto clinicamente importante — i tagli cesarei si riducevano, non aumentavano: 29.9% prima contro 28.3% dopo. Il sistema non decideva. Strutturava il ragionamento. E un ragionamento strutturato produce scelte migliori di uno non strutturato. Questo è precisamente ciò che osservammo con il nostro score al Cristo Re: non più cesarei, ma cesarei fatti per le ragioni giuste.

### **INFANT: il più grande fallimento che ci insegna di più**

Il trial INFANT, pubblicato sul Lancet nel 2017, è il più grande RCT mai condotto su un approccio computerizzato al monitoraggio fetale: 46.042 donne randomizzate a monitoraggio continuo con o senza il supporto del software INFANT — sviluppato da K2 Medical Systems, capace di analizzare frequenza cardiaca fetale e generare alert codificati per colore, senza raccomandazioni per azioni specifiche. Il risultato era inequivocabile nel suo fallimento: nessuna differenza nell'outcome neonatale primario, nessuna differenza negli outcome neurologici a due anni di vita.

Come è possibile? La risposta emerge dall'analisi interna: tra i neonati con outcome avverso e acidosi metabolica alla nascita, la revisione esperta identificava opportunità di miglioramento delle cure nel 40% del gruppo con supporto decisionale e nel 36% del gruppo controllo — una differenza non significativa. Il software identificava il problema. Ma i clinici non cambiavano il loro comportamento. Era, in miniatura, il problema dell>alert fatigue applicato alla situazione più critica che esiste in medicina.

#### **INFANT vs SisPorto: perché risultati così diversi?**

SisPorto è stato introdotto insieme a un cambiamento organizzativo complessivo, con formazione del personale e modifica dei workflow. L'analisi computerizzata era integrata in una nuova cultura clinica, non sovrapposta alla vecchia.

INFANT aggiungeva alert a un workflow esistente. I clinici potevano ignorarli, e spesso lo facevano. Un alert che non cambia il comportamento non è un CDSS: è rumore.

La lezione: il valore di un CDSS dipende dall'implementazione più che dall'algoritmo.



### **L'intelligenza artificiale e il CTG: performance promettenti, risultati irrisolti**

La crescita esponenziale delle capacità computazionali ha prodotto una proliferazione di algoritmi per l'analisi del CTG. Una scoping review del 2023 di Aeberhard e colleghi ha identificato 40 diversi studi che investigano almeno un algoritmo o sistema per classificare i tracciati CTG. Nessuno ha ottenuto grande accettazione nella pratica clinica. Non perché non funzionino in laboratorio — molti producono risultati impressionanti in contesti controllati — ma perché la distanza tra la performance su un dataset di addestramento e l'utilità clinica reale rimane, come in altri campi, la barriera più difficile da attraversare.

#### **I CAP: quando l'AI supera gli esperti umani**

Uno studio multicentrico del 2026 di Lin e colleghi ha prodotto i risultati più promettenti fino a oggi. I Cardiotocography Artificial-intelligence Predictors — CAPs — utilizzano architetture di deep learning (CNN, Transformer, LSTM, Cfc) addestrate su un dataset nazionale di grandi dimensioni. Il modello CAP-L, basato su architettura LSTM, raggiungeva un'area sotto la curva ROC di 0.758 per la predizione dell'ipossia fetale (pH arteria ombelicale

inferiore a 7.20), superando le performance degli esperti umani: AUROC 0.757-0.789 per i modelli AI contro 0.715 per i clinici, misurati su 10.571 risposte di esperti. L'analisi Grad-CAM mostrava che i modelli utilizzavano decelerazioni variabili e prolungate come segnale primario.

Il risultato è significativo, ma richiede cautela interpretativa. Per un pH inferiore a 7.05 — il valore che identifica un'acidosi grave — la sensibilità era del 79% e la specificità del 78%. Tradotto in termini clinici: il 21% delle acidosi gravi non veniva identificato. In ostetricia, dove le conseguenze degli errori sono irreversibili, questi margini rimangono troppo ampi per un'applicazione clinica autonoma.

### **I Large Language Models e il CTG: risultati contrastanti**

L'entusiasmo per i Grandi Modelli Linguistici ha raggiunto anche il CTG. Uno studio del 2024 di Gumilar e colleghi ha confrontato ChatGPT-4o, Gemini Advanced e Copilot con medici junior e senior nell'interpretazione di sette immagini CTG: ChatGPT-4o ha ottenuto il punteggio più alto (77.86), avvicinandosi ai medici senior (80.43) senza differenza statisticamente significativa. Uno studio più rigoroso del 2025 di Psilopatis e colleghi, su 60 tracciati classificati secondo i criteri FIGO, ha però rivelato limitazioni importanti: DeepSeek non era in grado di interpretare i CTG ed è stato escluso; Google Gemini mostrava performance scarse sui tracciati normali; ChatGPT-4.0 classificava correttamente i tracciati normali nel 47% dei casi e quelli anormali nel 50%; Bing Copilot interpretava i tracciati normali nel 97% dei casi, ma falliva completamente su quelli anormali.

La meta-analisi del 2019 di Balayla e Shrem su 55.064 pazienti da 3 RCT ha sintetizzato il quadro complessivo: rispetto all'analisi visuale, l'uso dell'AI non ha modificato l'incidenza di

acidosi neonatale, pH cordonale inferiore a 7.20, punteggio APGAR, modalità di parto, ammissione in terapia intensiva o mortalità perinatale. L'accordo inter-rater tra esperti e sistemi computerizzati era moderato al massimo, con un Cohen's Kappa medio di 0.49. La revisione sistematica del 2020 di Campanile e colleghi su 54.492 partecipanti da 3 RCT ha confermato risultati analoghi.



## Il paradosso ACOG 2025: una raccomandazione contro i CDSS

Nel 2025, l'American College of Obstetricians and Gynecologists ha pubblicato la Clinical Practice Guideline No. 10 sul monitoraggio intrapartum della frequenza cardiaca fetale. La raccomandazione più importante non era a favore di un nuovo sistema. Era contro.

*“ACOG raccomanda contro l'affidamento primario su approcci computerizzati per l'interpretazione e la gestione della frequenza cardiaca fetale durante il travaglio.”*

ACOG Clinical Practice Guideline No. 10, 2025 — Raccomandazione Forte,  
Evidenza di Qualità Moderata

Non è un arretramento luddista. È la sintesi onesta di decenni di ricerca: al momento non esistono trial clinici che supportino l'idea che l'interpretazione computerizzata del CTG migliori gli outcome neonatali. Gli approcci attuali rimangono inadeguati per guidare autonomamente le cure cliniche. Lo studio INFANT — il più grande mai condotto — suggerisce che gli outcome clinici potrebbero non migliorare anche quando i problemi del CTG vengono correttamente identificati. Ricominciare con paradigmi diversi è la risposta corretta.



### **Un caveat necessario: l'AI del 2025 non è quella del 2020**

Prima di lasciare questa rassegna con un giudizio definitivo, è necessario un caveat importante. La raccomandazione ACOG e le meta-analisi che la supportano si basano su studi condotti principalmente con sistemi AI sviluppati prima del 2022 — cioè prima della svolta dei Grandi Modelli Linguistici e del salto generazionale nelle architetture di deep learning. I sistemi valutati nei trial randomizzati come INFANT erano basati su regole euristiche o algoritmi di pattern recognition che oggi appartengono alla generazione precedente. Le reti neurali di nuova generazione — i Transformer, i modelli multimodali, le architetture con meccanismi di attenzione — mostrano capacità di interpretazione dei segnali clinici qualitativamente diverse da quelle dei sistemi testati nei trial citati.

I dati più recenti danno qualche indicazione di questa traiettoria. Lo studio CAP del 2026 ha dimostrato per la prima volta che un sistema AI supera gli esperti umani nella predizione dell'ipossia fetale su scala multicentrica. La velocità con cui queste performance stanno migliorando è tale che le conclusioni di una meta-analisi del 2019 o del 2020 potrebbero già non descrivere accuratamente le capacità dei sistemi disponibili oggi. Non si tratta di ottimismo acritico: la validazione clinica prospettica rimane indispensabile, e la performance su un dataset di addestramento non è mai sufficiente. Ma è ragionevole attendersi che, nei prossimi anni, alcuni dei limiti che hanno portato alla raccomandazione ACOG vengano progressivamente superati. La storia che stiamo raccontando in questo saggio non è finita. È in corso.



## Il cambio di paradigma: dall'interpretazione categoriale alla lettura fisiopatologica

Mentre gli RCT registravano delusioni e le linee guida invitavano alla cautela, un gruppo di ricercatori stava costruendo le fondamenta di un approccio radicalmente diverso. L'International Expert Consensus Statement on Physiological Interpretation of CTG, elaborato inizialmente nel 2018 e revisionato nel 2024 da Chandrabharan, Ghi, Pereira e colleghi — oltre 50 esperti di più di 20 paesi — ha rappresentato un cambio di paradigma nel senso più preciso del termine: non una versione migliorata del vecchio schema, ma uno schema concettualmente diverso.

I sistemi tradizionali classificano il CTG: Normale, Sospetto, Patologico. L'approccio fisiopatologico chiede una domanda diversa: qual è lo stato fisiologico di questo feto in questo momento? Sta compensando lo stress ipossico, o sta scompensando? Il modello distingue quattro scenari di ipossia con traiettorie e prognosi molto diverse: ipossia acuta, subacuta, gradualmente evolutiva e cronica preesistente. Come un adulto che affronta un test da sforzo progressivo, il feto mostra risposte compensatorie prevedibili: prima le decelerazioni, poi la perdita delle accelerazioni, poi la tachicardia mediata dalle catecolamine, infine l'alterazione della variabilità basale — il confine tra compensazione e scompenso.

Uno studio retrospettivo di Jia e colleghi su 500 tracciati CTG ha quantificato questa progressione con precisione. Quando decelerazioni variabili e tardive ripetitive erano presenti senza tachicardia, nessun feto aveva Apgar inferiore a 7 a 5 minuti o pH inferiore a 7. Solo quando la variabilità basale diventava anormale dopo la tachicardia si verificava un aumento statisticamente significativo degli outcome perinatali avversi: Apgar inferiore o

uguale a 7 nel 29.6% contro lo 0.9%; pH inferiore a 7 nel 29.5% contro lo 0%.

### **Il pattern ZigZag: un segnale precoce identificato dall'approccio fisiopatologico**

Una delle innovazioni più significative dell'approccio fisiopatologico è l'identificazione del pattern ZigZag — oscillazioni rapide, erratiche e ripetitive della frequenza cardiaca fetale con ampiezza superiore a 25 bpm, della durata minima di un minuto — come segno precoce di ipossia fetale intrapartum. Una meta-analisi del 2026 di Finadri e colleghi su 18.136 feti ha dimostrato che i feti con pattern ZigZag avevano un rischio aumentato di parto vaginale operativo (OR 2.22), di taglio cesareo (OR 1.71), di pH cordonale inferiore a 7.1 (OR 2.48) e di Apgar inferiore a 7 a 5 minuti (OR 2.13). Uno studio finlandese di Tarvonen e colleghi su 4.988 parti ha confermato che il ZigZag precede le decelerazioni tardive nell'88.7% dei casi: è un segnale precoce, non tardivo. E questo significa che riconoscerlo può aprire una finestra di intervento prima che la situazione degeneri.

#### **Le quattro risposte compensatorie fetali: la sequenza fisiopatologica**

1. Decelerazioni: il feto riduce il carico di lavoro miocardico.  
Segnale: stress in corso, compensazione attiva.
2. Perdita delle accelerazioni: il feto abolisce i movimenti somatici non essenziali.  
Segnale: riserve ridotte, compensazione sotto sforzo.
3. Tachicardia basale: risposta catecolaminergica allo stress ipossico.  
Segnale: compensazione mantenuta, ma con richiesta crescente.
4. Alterazione della variabilità dopo la tachicardia: prima aumenta (ZigZag), poi si riduce. Questo è il confine tra compensazione e scompenso — il momento clinicamente critico che richiede intervento immediato.



## **I primi tentativi di CDSS fisiopatologico: FIT-CAT e TWERIS Mini**

### **FIT-CAT: anticipare prima di interpretare**

Il primo passo dalla teoria alla pratica clinica strutturata è il FIT-CAT — FIT for labour, Clinical Anticipation Tool — presentato da Chandrabharan e colleghi nel 2024. La logica è elegante: non aspettare le alterazioni del CTG per interpretarle, ma anticiparle sulla base dei fattori di rischio antenatali e intrapartum. Il FIT-CAT aiuta i clinici a chiedersi, prima ancora che il travaglio inizi: è questo feto FIT per affrontare il viaggio progressivamente ipossico del travaglio? Quali alterazioni del tracciato mi aspetto, dato il contesto clinico? Dove dovrò alzare la guardia?

Questo approccio anticipatorio rappresenta uno spostamento concettuale profondo: dal reagire all'anormale al predire il possibile. Un feto con rottura prematura delle membrane, per esempio, potrebbe mostrare un CTG normale all'inizio del travaglio e poi deteriorarsi rapidamente per un'infezione silente in incubazione. Il FIT-CAT fornisce la cornice cognitiva per non essere colti di sorpresa da questa progressione.

### **TWERIS Mini: un progetto di visual AI basato sull'approccio fisiopatologico**

Un progetto che ha tentato esplicitamente di tradurre l'approccio fisiopatologico in uno strumento digitale è TWERIS Mini — il nome è un omaggio a Toueris, la dea egizia protettrice del parto. Il sistema è stato sviluppato da un consorzio internazionale (Francia, Regno Unito, Serbia, Germania) con l'obiettivo dichiarato di proteggere il potenziale neurologico dei neonati dalla cattiva interpretazione del CTG, e le loro madri da interventi operativi non necessari.

TWERIS Mini è un'applicazione di visual AI: analizza fotografie del tracciato CTG e classifica nove categorie basate sull'approccio fisiopatologico — nessuna ipossia, ipossia cronica, gradualmente evolutiva compensata, gradualmente evolutiva decompensata, subacuta, acuta, sinusoidale atipica (pattern Poole Shark Teeth), sinusoidale tipica e ZigZag. L'algoritmo sarebbe stato addestrato su un dataset di 5.724 immagini CTG.

Uno studio pubblicato nel 2025 sul Medical Research Archives — una rivista della European Society of Medicine non indicizzata nelle principali banche dati biomediche (PubMed, MEDLINE, Embase) — riporta un accordo del 94% (Kappa 0.92) tra il sistema e l'interpretazione di un esperto del gruppo di sviluppo su 100 tracciati anonimi. Questi dati preliminari non consentono una valutazione critica comparativa con i sistemi validati secondo gli standard usuali della ricerca clinica: mancano validazioni indipendenti, trial controllati e indicizzazione in letteratura peer-reviewed. Il sistema è menzionato qui perché rappresenta un tentativo esplicito di operativizzare il paradigma fisiopatologico in forma digitale — non perché ne siano dimostrate l'efficacia o l'utilità clinica.

### **FIT-CAT di GINOS: il modello del clinico-autore applicato al CTG**

Chi scrive sta sviluppando, nell'ambito del progetto GINOS, uno strumento di supporto decisionale per l'interpretazione fisiopatologica del CTG intrapartum, denominato FIT-CAT (Fetal Intrapartum Clinical Anticipation Tool, versione 4.2). A differenza dei sistemi di analisi automatica del tracciato, FIT-CAT non interpreta autonomamente il CTG: è il medico che, guidato da un workflow strutturato in tre passi, inserisce le variabili chiave — frequenza cardiaca fetale basale, variabilità, cycling, tipo di decelerazioni, intervallo inter-contrattile, segni di corioamnionite. Il

sistema elabora questi dati e classifica il tipo di pattern fisiopatologico tra dieci categorie: dal tracciato normale all'ipossia acuta (emergenza ostetrica), passando per i pattern ZigZag, Shark Teeth, SOFI, RUPI-L, e i diversi gradi di ipossia gradualmente evolutiva compensata e decompensata. La valutazione è ripetibile ogni 30-60 minuti durante il travaglio, con confronto automatico tra rivalutazioni successive registrate in una timeline.

La funzione che riteniamo più rilevante sul piano clinico è la generazione automatica della nota clinica documentale. Sulla base dei parametri inseriti, Claude produce una nota in stile referto ospedaliero italiano — in prosa discorsiva, forma impersonale, senza elenchi puntati — pronta per essere copiata in cartella clinica. Dalla seconda rivalutazione in poi, la nota include automaticamente il raccordo con la valutazione precedente, documentando l'andamento nel tempo del pattern ipossico: "Rivalutazione delle ore 15:05: rispetto alla valutazione delle ore 14:30, che mostrava ipossia gradualmente evolutiva compensata con FCF 138 bpm, si osserva ora un incremento della linea di base a 152 bpm con riduzione del cycling, configurando un quadro di ipossia gradualmente evolutiva in fase di decompensazione...". Questa funzione risponde alla stessa necessità che aveva motivato la costruzione dello score CTG al Cristo Re vent'anni prima: rendere il ragionamento clinico esplicito, documentato, tracciabile. Un ragionamento scritto è un ragionamento che si può correggere, insegnare, discutere in sede di revisione dei casi.

Il sistema include anche un'analisi opzionale dell'immagine del tracciato tramite Claude Vision, che pre-compila automaticamente i campi del Passo 2. Questa funzione è ancora in sviluppo: i risultati non sono soddisfacenti, l'accuratezza dipende fortemente dalla qualità della fotografia, e i valori suggeriti

richiedono sistematicamente verifica e correzione da parte del clinico prima dell'uso. La strada verso un'analisi automatica affidabile dell'immagine CTG rimane aperta — e difficile.

Non esistono ancora dati clinici pubblicati su FIT-CAT, e non è possibile affermare che il sistema sia efficace o utile nella pratica clinica generale: lo dirà solo la validazione. È in corso lo studio VALIDATE, condotto in collaborazione con la Fondazione Policlinico Universitario Agostino Gemelli IRCCS, che correlazionerà la classificazione FIT-CAT intrapartum con gli esiti neonatali — pH cordonale, Apgar, ricovero in TIN, HIE — utilizzando i dati esportabili dall'archivio dell'applicazione. Lo menzioniamo qui non come prova di efficacia, ma come esempio concreto di ciò che il modello del clinico-autore rende oggi possibile: costruire lo strumento che si vorrebbe avere, sulla base di un paradigma che si conosce e si pratica, e verificarne la validità sul campo. Il risultato clinico dipenderà dalla validazione. Il processo di costruzione è già, di per sé, un esercizio di formalizzazione del ragionamento clinico.



## **Perché non esistono ancora altri CDSS fisiopatologici**

TWERIS Mini e FIT-CAT di GINOS rappresentano i primi tentativi di operativizzare il paradigma fisiopatologico in forma digitale. Entrambi si trovano in una fase iniziale di sviluppo e validazione, e nessuno dei due ha ancora dimostrato utilità clinica in studi indipendenti. Il percorso verso CDSS fisiopatologici pienamente integrati nei sistemi di monitoraggio clinico rimane lungo.

Le barriere principali sono tre. La prima è la mancanza di dataset annotati secondo i criteri fisiopatologici: non esistono grandi database multicentrici di CTG etichettati secondo i principi

dell'interpretazione fisiopatologica. La seconda è la complessità dell'individualizzazione: l'approccio fisiopatologico integra il contesto clinico — riserve fetali stimate, attività uterina, fattori di rischio, ambiente intrauterino — che è difficile da codificare algoritmicamente per ogni singola paziente. La terza è la variabilità inter-osservatore anche tra esperti: un Cohen's Kappa medio di 0.49 nelle meta-analisi sui sistemi esistenti non garantisce un gold standard su cui costruire l'algoritmo.

I sistemi più promettenti nel breve periodo integrano dati multimodali. Uno studio del 2025 di Qiu e colleghi ha sviluppato un sistema bimodale che integra CTG e contrazioni uterine con architettura DenseNet, raggiungendo un'AUC di 0.944 contro 0.812 del sistema unimodale, con il 100% di recall sui casi anormali. Uno studio del 2026 di Mendis e colleghi su Fusion ResNet — che combina reti convoluzionali per il segnale CTG con una rete neurale per caratteristiche tabulari cliniche — ha raggiunto un'AUC di 0.84 sulla validazione esterna. La direzione è chiara: non il tracciato da solo, ma il tracciato nel suo contesto fisiologico e clinico.



## **Lo screening del cancro cervicale: quando la macchina batte il patologo**

Tra le applicazioni dell'intelligenza artificiale in ginecologia, quella più matura e con le prove più solide è probabilmente lo screening del cancro cervicale. Una meta-analisi del 2025 di Liu e colleghi che ha analizzato 77 studi ha documentato che la citologia cervicale AI-assistita raggiunge un'accuratezza del 94% nel Pap test e del 90% nel test ThinPrep (TCT). Nel ThinPrep la sensibilità

raggiunge il 97% — un numero che farebbe impallidire molti laboratori di citologia tradizionale.

Ma è il confronto con la colposcopia che produce forse il risultato più sorprendente. I sistemi AI non soltanto eguagliano i colposcopisti esperti — li superano. Un odds ratio di 1.75 a favore dell'AI nell'accuratezza colposcopica (AUC significativamente superiore), con un tasso di concordanza globale con i risultati istopatologici dell'81% contro il 74% dei colposcopisti esperti. Il sistema CerviCARE AI, validato in uno studio multicentrico di Ouh e colleghi del 2024, ha raggiunto una sensibilità del 98% per i casi ad alto rischio — CIN2 o superiore — con una specificità del 95.5%.

La rilevanza di questi risultati va oltre la performance tecnica. Lo screening del cancro cervicale è uno dei settori dove le disuguaglianze geografiche ed economiche uccidono più persone. Un sistema AI che può funzionare su uno smartphone collegato a una telecamera digitale, senza bisogno di un laboratorio specializzato, è qualcosa di potenzialmente rivoluzionario per i paesi a basso e medio reddito, dove il cancro cervicale è ancora la seconda causa di morte oncologica nelle donne.



## La predizione delle complicanze in gravidanza: il sogno dell'ostetricia di precisione

L'ambizione più grande è quella che i ricercatori chiamano "ostetricia di precisione": identificare, nelle settimane iniziali di una gravidanza, le donne che svilupperanno complicanze gravi. Una revisione narrativa del 2023 di Mennicken e colleghi ha documentato l'applicazione del machine learning a tutti i principali rischi ostetrici: preeclampsia, emorragia postpartum, parto pretermine, diabete gestazionale, morte perinatale. Uno studio del

2025 di Pi e colleghi su 1.014 donne in Bangladesh ha trovato che il Multilayer Perceptron (MLP) raggiungeva un'accuratezza del 91% nell'identificazione dei casi ad alto rischio. Una revisione sistematica del 2021 di Bertini e colleghi su 31 studi ha documentato accuratèzze fino al 95-99% nei modelli migliori.

#### **MLP, Random Forest e XGBoost — tre modi di imparare dai dati (spiegati a un ragazzo delle medie)**

##### **Random Forest — la foresta di cento medici**

Immagina di dover decidere se un neonato ha bisogno di cure immediate. Chiedi l'opinione a cento medici diversi, ma ognuno vede solo una parte dei dati: uno vede peso e pH, un altro Apgar ed età gestazionale, un terzo la modalità di parto e il lattato. Ognuno vota. Alla fine prendi la decisione che ha ottenuto più voti. Questo è il Random Forest: cento "alberi decisionali", ciascuno addestrato su una parte casuale dei dati, che votano insieme. Gli errori dei singoli si compensano, e il risultato finale è più robusto di qualsiasi esperto individuale.

##### **XGBoost — imparare dagli errori degli altri**

Immagina un processo diverso: il primo medico fa una previsione. Poi arriva il secondo, che guarda solo i casi in cui il primo ha sbagliato, e cerca di migliorare esattamente quelli. Poi il terzo si occupa degli errori che i primi due hanno ancora commesso. E così via, cento volte. Ogni nuovo esperto si specializza nelle debolezze del precedente. Questo è l'XGBoost: un sistema che impara iterativamente dai propri errori, costruendo la previsione finale come la somma di cento piccole correzioni progressive. È spesso il modello che vince nelle competizioni di machine learning proprio perché è eccezionalmente efficiente nel minimizzare gli errori residui.

##### **Multilayer Perceptron (MLP) — la rete neurale a strati**

Il terzo approccio si ispira al funzionamento del cervello. Immagina una serie di strati di "neuroni artificiali": il primo strato riceve i dati grezzi (pH, Apgar, peso...) e li trasforma con piccoli calcoli matematici; il secondo strato elabora l'output del primo; il terzo quello del secondo. Ogni strato impara a riconoscere caratteristiche sempre più complesse: il primo capisce forse se il pH è basso; il secondo capisce se la combinazione pH basso + Apgar basso è preoccupante; il terzo capisce se quella combinazione, in un neonato pretermine con basso peso, indica un rischio reale. L'MLP è il più potente dei tre nel catturare relazioni complesse tra variabili — ed è per questo che nello studio del Bangladesh risultava quello con le performance migliori.

Queste percentuali impressionano. Ma una voce critica deve alzarsi: accuratezza su quale dataset? Validato su quale popolazione? La maggior parte degli studi che producono questi risultati sono stati condotti in un singolo centro, su popolazioni specifiche, con criteri di inclusione selettivi. Quando i modelli vengono testati su popolazioni diverse da quelle su cui sono stati addestrati, le performance calano — a volte drammaticamente. La variabilità biologica, demografica e socioeconomica tra popolazioni

diverse non è rumore statistico: è il nucleo del problema di generalizzazione che limita l'utilità pratica di questi modelli.

### **Un CDSS “fatto in casa”: la predizione dell'encefalopatia ipossico-ischemica**

Chi scrive ha costruito, con l'assistenza di Claude, uno strumento di predizione del rischio di encefalopatia ipossico-ischemica (HIE) nel neonato. I dati di ingresso sono quelli disponibili nelle prime ore di vita: peso alla nascita, punteggio Apgar a uno e cinque minuti, modalità del parto, età gestazionale, concentrazione del lattato arterioso, eccesso di basi (BE) e pH dell'arteria ombelicale. Questi parametri sono stati associati all'esito clinico reale — se il neonato ha sviluppato o meno una HIE — costruendo un dataset locale basato sulla propria casistica. Claude ha tradotto questi dati in un modello predittivo che restituisce, per ogni nuovo caso, il valore predittivo positivo (VPP) e il valore predittivo negativo (VPN) per il rischio di HIE.

Lo strumento non è commercializzato, non è stato oggetto di validazione formale, e non consente generalizzazioni al di fuori del contesto in cui è stato costruito. Vale però la pena menzionarlo per due ragioni concettuali. La prima è metodologica: un modello addestrato su dati locali, in un contesto clinico specifico, può essere più accurato nel predire gli esiti in quella struttura rispetto a modelli addestrati su grandi dataset eterogenei. Le caratteristiche della popolazione, i protocolli assistenziali, le soglie di intervento e i criteri diagnostici variano tra centri — ed è proprio questa variabilità la ragione per cui i modelli addestrati altrove calano di performance quando vengono esportati.

La seconda ragione è pratica: costruire questo strumento ha richiesto un pomeriggio. Non mesi, non un team di informatici, non fondi di ricerca. Un clinico, i propri dati, un modello linguistico. I

grandi modelli di machine learning addestrati su dataset nazionali o internazionali hanno il vantaggio della potenza statistica e della potenziale generalizzabilità. I modelli locali costruiti da clinici con la propria casistica hanno il vantaggio della pertinenza contestuale: sono calibrati sulla popolazione che effettivamente incontrano. Il futuro del supporto decisionale in ostetricia è probabilmente nell'integrazione tra i due — modelli larghi che definiscono le probabilità a priori, e modelli locali che le correggono sulla base della propria realtà clinica.



### **L'ecografia ostetrica nell'era dell'AI: verso gli "ultrasuoni 5D"**

L'ecografia ostetrica è forse l'applicazione di AI più immediatamente comprensibile per un ginecologo-ostetrico. Le immagini ecografiche sono dati strutturati — matrici di pixel con pattern riconoscibili — e il deep learning eccelle nel loro riconoscimento. Una scoping review del 2023 di Horgan e colleghi ha identificato 127 articoli sull'uso dell'AI nell'ecografia ostetrica, coprendo applicazioni che vanno dalla biometria fetale automatizzata all'ecocardiografia fetale, dalla neurosonografia alla valutazione della placenta.

Il termine "ultrasuoni 5D" — coniato in una revisione sistematica dello stesso anno di Ramirez Zegarra e Ghi — cattura bene l'idea: aggiungere all'ecografia tridimensionale una quinta dimensione, l'intelligenza. La capacità della macchina di interpretare ciò che vede, non soltanto di visualizzarlo. L'importanza clinica del riconoscimento automatico di malformazioni strutturali difficilmente può essere sopravvalutata: ogni anno nel mondo nascono centinaia di migliaia di bambini con malformazioni cardiache congenite che avrebbero potuto essere

diagnosticate prima della nascita, permettendo una pianificazione del parto in un centro specializzato.



## Una lezione che attraversa tutto il capitolo

La storia dei CDSS in ostetricia offre, in miniatura, tutte le lezioni che la storia più ampia dei CDSS ha impiegato decenni a formalizzare. La tecnologia non basta, se non è integrata in un cambiamento organizzativo reale. L’alert non cambia il comportamento clinico, se il clinico non è formato a riceverlo. Un sistema che classifica senza spiegare genera automazione senza comprensione. E — forse la lezione più sorprendente — i sistemi migliori non sono necessariamente i più sofisticati: sono quelli che portano il ragionamento clinico a essere esplicito, documentato, condivisibile.

Questo, in fondo, è ciò che lo score CTG costruito con il professor Serra faceva negli anni Novanta. Non era un sistema esperto. Non usava il machine learning. Era un foglio di carta con variabili e punteggi. Ma rendeva il ragionamento esplicito. E un ragionamento esplicito è un ragionamento che si può correggere, migliorare, insegnare, condividere. TWERIS Mini, cinquant’anni dopo, tenta di fare la stessa cosa — con l’intelligenza artificiale al posto del foglio di carta, e con la fisiopatologia al posto della classificazione categoriale.

## Capitolo XII: Il clinico autore — un modello alternativo per i CDSS

### Il problema dell'implementation gap

C'è una domanda che tormenta i ricercatori di informatica medica da decenni: perché c'è un gap così enorme tra ciò che le linee guida cliniche raccomandano e ciò che i medici effettivamente fanno nella pratica quotidiana?

Le linee guida esistono. Sono scritte da esperti, basate su prove, pubblicate su riviste rispettate, distribuite in ogni angolo del mondo. Eppure gli studi mostrano ripetutamente che la maggior parte dei medici non le segue in modo coerente — non perché siano incompetenti, non perché non le conoscano, ma perché non hanno gli strumenti per applicarle in modo efficiente nel caos di una giornata clinica reale.

I CDSS sono stati proposti, fin dai tempi di MYCIN, come la risposta a questo problema. Un sistema che porta la linea guida al letto del paziente, al momento giusto, nella forma giusta. Ma la storia che abbiamo raccontato in questo saggio mostra che anche i CDSS faticano ad essere adottati — per ragioni tecniche, organizzative, culturali.

### Il modello del clinico-autore

Esiste, però, un modello alternativo che ha cominciato a emergere nell'ultima decade, reso possibile dalla rivoluzione dei grandi modelli linguistici. È il modello del clinico-autore: il medico non come utente di un CDSS sviluppato da altri, ma come autore di strumenti digitali progettati sulla propria pratica clinica specifica.

La logica è la seguente. Nessuno conosce il problema clinico meglio del clinico che lo affronta ogni giorno. Nessuno sa meglio di

lui quali sono le domande che si pone, quali le incertezze ricorrenti, quali le trappole cognitive in cui è più facile cadere. Se a questo clinico venisse dato uno strumento semplice per tradurre quella conoscenza in un'applicazione digitale — senza bisogno di saper programmare, senza bisogno di un team di ingegneri — potrebbe creare strumenti di una rilevanza pratica che nessun sistema sviluppato a distanza dalla clinica reale potrebbe eguagliare.

I grandi modelli linguistici — sistemi come quelli sviluppati da Anthropic, OpenAI, Google — hanno abbassato drammaticamente la barriera tecnica tra l'idea e la sua realizzazione. Un medico che descrive in linguaggio naturale la logica di un percorso diagnostico-terapeutico può vedere quella logica tradotta in un'applicazione funzionante nel giro di ore, non di mesi.

Questo non significa che i CDSS creati dai clinici siano automaticamente buoni. Richiedono validazione, revisione critica, aggiornamento continuo. Richiedono che il clinico-autore sia consapevole delle proprie limitazioni cognitive — dei bias di disponibilità, di rappresentatività, di ancoraggio che influenzano il ragionamento medico tanto quanto influenzano il ragionamento di qualsiasi essere umano.

Ma rappresentano una via nuova — e, per molte situazioni cliniche specifiche, più efficace — per portare il supporto decisionale nella pratica quotidiana. Non un sistema calato dall'alto, sviluppato in un laboratorio lontano dalle corsie e dalle sale parto, ma uno strumento costruito dal basso, a misura di chi lo userà.



## Capitolo XIII: Il medico che porta il suo manuale in sala parto — il RAG come architettura di CDSS

### Una macchina che inventa e una macchina che legge

Quando i Grandi Modelli Linguistici sono apparsi sulla scena — ChatGPT, Claude, Gemini — la reazione di molti medici è stata di entusiasmo immediato. Finalmente una macchina capace di rispondere a qualsiasi domanda clinica in linguaggio naturale, senza bisogno di imparare comandi speciali o navigare database complicati. Ma dopo il primo entusiasmo sono arrivati i problemi. La macchina inventava riferimenti bibliografici che non esistevano. Citava studi mai pubblicati. Descriveva dosaggi farmacologici sbagliati con la stessa sicurezza con cui descriveva quelli giusti. Questo problema ha un nome: allucinazione. E ha una causa precisa: i Grandi Modelli Linguistici non “sanno” le cose nel senso in cui le sa un medico. Le “prevedono” — generano la risposta più probabile sulla base di tutto il testo su cui sono stati addestrati. E quando non hanno abbastanza informazioni sicure, la macchina continua comunque a produrre testo plausibile. Inventa.

Immagina di chiedere a qualcuno che ha letto milioni di libri di medicina di rispondere alle tue domande. Sa moltissimo. Ma sa anche cose che non sa di non sapere. E quando non sa, invece di dire “non lo so”, indovina — e lo dice con la stessa voce sicura con cui dice le cose vere. In medicina, questo è un problema serio. Un farmaco sbagliato, un dosaggio inventato, una linea guida citata a rovescio possono fare del male. Il RAG nasce per risolvere esattamente questo.



## Che cos'è il RAG — spiegato ai ragazzi delle medie

RAG sta per Retrieval-Augmented Generation — che in italiano potremmo tradurre come “generazione aumentata dal recupero”. Il nome è tecnico, ma il concetto è semplicissimo. Proviamo con un’analogia.

Immagina un esame scolastico. Ci sono due tipi di esame: “a libro chiuso” e “a libro aperto”. Nell’esame a libro chiuso, devi ricordare tutto quello che hai studiato. Se non ricordi, rischi di inventare. Nell’esame a libro aperto, hai i libri sul banco. Non devi ricordare tutto: devi sapere dove trovare la risposta giusta, leggerla, e usarla per rispondere. Un Grande Modello Linguistico tradizionale funziona come uno studente all’esame a libro chiuso: risponde con quello che ricorda, e quando non ricorda, rischia di inventare. Un sistema RAG funziona come uno studente all’esame a libro aperto: prima cerca la risposta nei documenti che ha sul banco, poi la usa per rispondere. Non inventa, perché ha il libro davanti.

### Come funziona il RAG in tre passi

1. TU inserisci i documenti di riferimento: linee guida, protocolli, articoli, schede farmacologiche, la tua casistica. Questo è il “libro aperto”.
2. TU fai una domanda clinica. Il sistema CERCA nei tuoi documenti i passaggi più rilevanti per quella domanda specifica.
3. Il modello linguistico LEGGE quei passaggi e produce la risposta.  
Non inventa: ragiona solo su ciò che ha trovato nei tuoi documenti.

Se la risposta non è nei tuoi documenti, il sistema dice che non lo sa. Questo è il vantaggio più grande: sa quando non sa.



## Perché il RAG non allucina

La ragione per cui il RAG riduce drasticamente le allucinazioni è strutturale, non magica. Il modello linguistico non è più costretto a rispondere con quello che ricorda vagamente dal suo addestramento. Gli viene fornita la fonte primaria — il documento esatto, il passaggio esatto — e il suo compito diventa semplicissimo: capire quel passaggio e riformularlo in risposta alla domanda. È come la differenza tra chiedere a qualcuno “come funziona il motore di un’auto?” e mettergli davanti il manuale del costruttore aperto alla pagina giusta.

Rimane comunque un margine di errore: il modello può interpretare male un passaggio, o il sistema di ricerca può trovare il documento sbagliato. Ma questi errori sono molto più rari, molto più facilmente verificabili — perché il sistema può sempre mostrare la fonte da cui ha preso la risposta — e soprattutto non hanno la caratteristica più insidiosa dell’allucinazione: sembrare sicuri anche quando sono sbagliati. Un sistema RAG che non trova risposta nei suoi documenti lo dice. Non inventa una risposta plausibile.

In medicina, questa trasparenza vale oro. Ogni risposta può essere accompagnata dalla citazione precisa: “questo lo dice la linea guida SIGO 2025, paragrafo 3.4”. Il medico può verificare, e sa da dove viene quella raccomandazione. Non si fida ciecamente della macchina: si fida della fonte, che conosce e sa valutare.



## Dal libro generico al manuale personalizzato sulla tua paziente

Il RAG ha un secondo vantaggio che va oltre la semplice riduzione delle allucinazioni, ed è quello che lo rende una

tecnologia particolarmente interessante per costruire CDSS: la possibilità di calibrare la risposta non su una paziente generica, ma su questa paziente specifica, in questo momento.

Riprendiamo l'esempio del diabete gestazionale che abbiamo usato nel capitolo precedente. Una linea guida tradizionale dice: "in presenza di fattori di rischio per diabete gestazionale, eseguire l'OGTT." Un Grande Modello Linguistico tradizionale, interrogato su questo, riformulerà la stessa raccomandazione generica. Un sistema RAG costruito correttamente può fare qualcosa di molto più sofisticato: il medico inserisce i dati della paziente — 28 settimane, BMI 29, madre diabetica, glicemia a digiuno 95 mg/dL — il sistema recupera dai documenti i passaggi specifici per quel profilo, e restituisce una risposta calibrata: "per questa paziente, con questi fattori di rischio a questa settimana di gestazione, la raccomandazione è eseguire l'OGTT entro questa settimana". Con tanto di citazione precisa della linea guida da cui quella raccomandazione proviene.

### **Il medico che porta il suo archivio in sala parto**

L'immagine più utile per capire il RAG in clinica è quella di un medico che entra in sala parto con un archivio personale: tutte le linee guida che ha scelto, i protocolli del suo reparto, gli articoli chiave che considera affidabili, magari la casistica dei casi più complessi che ha seguito negli anni. Quando sorge una domanda clinica, non deve ricordare tutto a memoria: consulta il suo archivio. Sa che le risposte vengono da quelle fonti, che lui ha selezionato e di cui si fida.

Un sistema RAG fa esattamente questo — ma in pochi secondi, e su una quantità di documenti che nessun essere umano potrebbe tenere a mente simultaneamente. I documenti di riferimento li sceglie il clinico: può inserire le linee guida SIGO, i

protocolli aziendali, i consensus statement internazionali che preferisce. Il sistema ragionerà su quelle fonti, e solo su quelle. Non aggiungerà nulla che non sia nei documenti inseriti.

#### RAG vs LLM tradizionale in clinica: cosa cambia

LLM tradizionale:

- Risponde con tutto ciò che ha “imparato” durante l’addestramento
- Non sa sempre distinguere ciò che sa da ciò che non sa
- Può allucinare: risposta plausibile ma sbagliata, senza avvisare
- Non cita fonti verificabili
- Non distingue tra linea guida del 2018 e linea guida del 2025

Sistema RAG:

- Risponde solo su ciò che è nei documenti inseriti dal clinico
- Se la risposta non c’è, lo dice
- Cita la fonte esatta: paragrafo, documento, pagina
- Il clinico sceglie i documenti: linee guida aggiornate, protocolli, casistica
- Calibra la risposta sui dati della paziente specifica



## Il RAG come architettura naturale per i CDSS

Il collegamento tra RAG e CDSS non è casuale: il RAG è probabilmente l’architettura più adatta per costruire CDSS con i modelli linguistici di nuova generazione. I CDSS tradizionali codificavano le regole manualmente: un esperto scriveva le condizioni, le soglie, le raccomandazioni. Il processo era lento, costoso, e diventava obsoleto ogni volta che usciva una nuova linea guida. Con il RAG, aggiornare il CDSS significa semplicemente aggiornare i documenti di riferimento: si sostituisce la vecchia linea guida con la nuova, e il sistema inizia immediatamente a ragionare sulla versione aggiornata. Nessun programmatore necessario. Nessuna riscrittura delle regole.

La suite GINOS esplora già parzialmente questa direzione: alcune applicazioni utilizzano Claude come motore linguistico con

un sistema prompt che incorpora le regole cliniche chiave. È una forma embrionale di RAG — non ancora con recupero dinamico da un archivio, ma con un insieme fisso di conoscenza clinica selezionata e incorporata nel prompt. Il passo successivo naturale è costruire un vero archivio documentale indicizzato — linee guida, protocolli, consensus statement — che il sistema consulta dinamicamente in base alla domanda clinica e ai dati della paziente. La nota clinica che ne risulta non sarebbe più basata su regole fisse, ma sulla lettura ragionata dei documenti più pertinenti per quel caso specifico.

### **I limiti che restano**

Il RAG non è una soluzione magica. Ha i suoi limiti, e è importante nominarli con onestà. Il primo è la qualità dei documenti inseriti: se i documenti di riferimento sono vecchi, incompleti o sbagliati, le risposte lo saranno di conseguenza. Il sistema ragiona bene su documenti buoni. Su documenti scadenti, ragiona in modo scadente. Il secondo limite è il recupero: il sistema deve trovare il passaggio giusto tra migliaia di pagine. Se il meccanismo di ricerca recupera il paragrafo sbagliato, la risposta sarà sbagliata anche se il documento giusto era nell'archivio. Il terzo limite è la comprensione: il modello deve capire correttamente il passaggio che ha trovato. Testi molto tecnici, con terminologia ambigua o struttura complessa, possono essere interpretati in modo impreciso.

Nessuno di questi limiti è insormontabile. Tutti e tre possono essere ridotti con una buona progettazione del sistema, con la scelta accurata dei documenti, con meccanismi di validazione delle risposte. Ma vanno nominati, perché in medicina l'ottimismo acritico verso la tecnologia ha già causato delusioni che abbiamo raccontato in questo libro. Il RAG è uno strumento potente. Come

tutti gli strumenti, il suo valore dipende da come viene usato, da chi lo progetta, e da chi lo controlla.



### Una promessa concreta

Il RAG rappresenta oggi la frontiera più promettente per costruire CDSS affidabili con i modelli linguistici. Non perché elimini tutti i rischi — non li elimina. Ma perché sposta il rischio in un posto dove il clinico può vederlo e controllarlo: la qualità delle fonti che sceglie di inserire, la pertinenza delle domande che fa, la capacità di leggere criticamente le risposte che riceve. Questi sono compiti che un medico esperto sa fare. Sono compiti professionali, non tecnici.

Il medico che usa un sistema RAG bene costruito non smette di ragionare. Ragiona meglio, perché ha a disposizione più conoscenza organizzata, recuperata più in fretta, calibrata sulla paziente che ha davanti. Quello che faceva con ore di lettura e un'ottima memoria, lo fa in pochi secondi — con fonti verificabili e senza il rischio che la stanchezza o la pressione del turno gli facciano dimenticare un passaggio cruciale. Non è la macchina che decide. È ancora il medico. Ma è un medico meglio supportato.

## Capitolo XIV: La distanza tra sapere e fare — i CDSS come strumento di traduzione delle linee guida

### Il paradosso al cuore della medicina basata sull'evidenza

C'è un paradosso che chiunque abbia praticato medicina clinica conosce bene, anche se raramente lo nomina con precisione.

Le linee guida esistono. Sono scritte da esperti, basate su meta-analisi di migliaia di pazienti, pubblicate sulle riviste più autorevoli, distribuite in ogni angolo del mondo. Eppure tra ciò che le linee guida raccomandano e ciò che accade nelle corsie degli ospedali c'è una distanza che la ricerca sull'implementazione ha misurato con precisione crescente e preoccupante. Si stima che i pazienti ricevano le cure raccomandate dall'evidenza scientifica approssimativamente nel 50% dei casi. L'altra metà riceve cure non ottimali — non per malafede, non per ignoranza, ma per la complessità intrinseca di tradurre la conoscenza in azione nel momento giusto, nel posto giusto, per il paziente giusto.

Questo divario — che la letteratura chiama *implementation gap* o *knowledge-to-action gap* — non è un fallimento individuale. È un fallimento sistemico che ha radici profonde nella natura stessa delle linee guida. Una linea guida è, per definizione, una raccomandazione per la popolazione: dice cosa fare nella maggioranza dei pazienti con quella condizione, in quelle circostanze. Non dice cosa fare per questa paziente, con questa storia clinica, in questa struttura, oggi. La distanza tra la raccomandazione generale e la decisione individuale è precisamente lo spazio in cui il ragionamento clinico esiste. Ed è precisamente lo spazio in cui i CDSS possono essere più utili.



## Tre problemi che nessun documento può risolvere

### Il problema dell'accessibilità nel momento critico

Il primo problema è temporale. Una linea guida viene letta quando si ha tempo di leggerla: in biblioteca, a casa, durante la formazione. Non viene letta — non può essere letta, nella sua interezza — alle tre di notte davanti a una paziente in travaglio

complicato. Il medico di guardia applica quello che ricorda, che è una semplificazione necessaria ma inevitabilmente imperfetta della linea guida completa. I CDSS nascono storicamente proprio per rispondere a questo problema: portare la linea guida al letto del paziente, nel momento in cui serve, nella forma in cui è utilizzabile. Già Clement McDonald lo aveva capito negli anni Settanta con i suoi clinical reminders: non un documento da leggere, ma un avviso che arriva quando si è davanti al paziente giusto.

### **Il problema della complessità combinatoria**

Il secondo problema è cognitivo. Le linee guida moderne sono documenti articolati, spesso di decine o centinaia di pagine, con algoritmi decisionali ramificati, eccezioni, controindicazioni, raccomandazioni diverse per sottogruppi di pazienti. Un paziente reale spesso soddisfa contemporaneamente i criteri di più linee guida diverse — una per la patologia di base, una per la comorbilità, una per il farmaco che sta assumendo. Nessun medico può tenere a mente simultaneamente tutte le raccomandazioni rilevanti per un paziente complesso. Non perché sia incompetente: perché il numero di combinazioni possibili supera strutturalmente la capacità della memoria di lavoro umana. Un CDSS ben costruito può verificare simultaneamente le condizioni rilevanti e segnalare quelle pertinenti al caso — ma la qualità di questo supporto dipende interamente da come il sistema è stato progettato: un CDSS mal costruito o obsoleto nella sua base di conoscenza è potenzialmente peggio di nessun CDSS, perché dà al medico una falsa sicurezza di aver verificato ciò che non ha verificato.

### **Il problema della calibrazione sul caso specifico**

Il terzo problema è il più importante, e il meno visibile. Una linea guida è scritta per la popolazione: stabilisce cosa fare nella maggioranza dei casi con quella condizione. Ma il clinico è di fronte

a una paziente specifica, con caratteristiche proprie, in un momento preciso del percorso clinico. La distanza tra la raccomandazione generale e la decisione individuale si attraversa con il ragionamento clinico — ed è qui che un CDSS ben costruito può fare qualcosa che nessun documento cartaceo può fare: calare la linea guida su quella specifica paziente.

Prendiamo un esempio concreto: la linea guida sul diabete gestazionale. Un ginecologo sa che esiste un algoritmo per decidere quando eseguire il carico orale di glucosio (OGTT). Ma questo algoritmo dipende da una combinazione di fattori — settimana di gestazione, BMI, familiarità per diabete, etnia, glicemia a digiuno, precedenti di diabete gestazionale — che produce percorsi diversi per pazienti diverse. Leggere l'intera linea guida ogni volta che si visita una nuova gestante non è praticabile. Un CDSS che incorpora quella linea guida fa domande mirate, raccoglie i fattori rilevanti per quella paziente specifica, e risponde con la raccomandazione pertinente a quella situazione: esegui l'OGTT adesso, aspetta la sedicesima settimana, oppure procedi direttamente alla terapia. Il medico ottiene la risposta per questa paziente senza dover rileggere l'intero documento.

Ma la calibrazione va oltre la semplice navigazione dell'algoritmo. Un CDSS può integrare nella raccomandazione strumenti pratici che la linea guida non può offrire. Per la stessa paziente con diabete gestazionale: istruzioni dettagliate su come eseguire il profilo glicemico, con un'applicazione che la gestante usa per registrare i valori e inviarli automaticamente al ginecologo. Un'applicazione per il monitoraggio della pressione arteriosa che genera alert in caso di valori critici, produce un grafico dell'andamento nel tempo, e fornisce consigli pratici calibrati sulla settimana di gestazione. Un modulo che guida il ginecologo nella

scelta del farmaco antipertensivo corretto per quella specifica paziente — non genericamente l'antipertensivo in gravidanza, ma il farmaco giusto, al dosaggio calibrato su quel peso e su quella settimana di gestazione. La linea guida stabilisce il principio. Il CDSS costruisce lo strumento che lo applica.

#### Calare la linea guida sulla paziente specifica: cosa cambia

Linea guida: "in presenza di fattori di rischio per diabete gestazionale, eseguire OGTT"

CDSS: "questa paziente, 28 settimane, BMI 29, madre diabetica, glicemia 95 mg/dL  
→ eseguire OGTT entro questa settimana"

Linea guida: "monitorare la pressione arteriosa in gravidanza"

CDSS: app per la gestante → registrazione automatica → alert se PA>140/90  
→ grafico trend → segnalazione al ginecologo

Linea guida: "utilizzare farmaci antipertensivi sicuri in gravidanza"

CDSS: "per questa paziente, 32 settimane, 74 kg: metildopa 250 mg x 3/die"

Il CDSS non sostituisce la linea guida. La porta fin dentro la visita, calibrata su quella paziente, in quella settimana, con quel peso.



## La linea guida come architettura, il CDSS come ingegnere

Una metafora può aiutare a chiarire la relazione tra linee guida e CDSS. Una linea guida è come il codice edilizio: stabilisce i principi generali di sicurezza, i materiali consentiti, le dimensioni minime, i requisiti strutturali che ogni edificio deve rispettare. È un documento essenziale, costruito sull'esperienza collettiva, pensato per proteggere chi vivrà negli edifici. Ma il codice edilizio non costruisce l'edificio. Lo costruisce l'ingegnere, che conosce il codice, lo applica al terreno specifico, alle esigenze specifiche del committente, ai materiali disponibili in quel luogo. Il CDSS è l'ingegnere: traduce il codice in un progetto concreto, adattato alla situazione reale.

Questa distinzione ha conseguenze pratiche importanti. Un CDSS che si limita a riprodurre il testo della linea guida — a “mostrare” la raccomandazione senza adattarla al caso — non risolve il problema dell’implementation gap. Lo sposta: invece di dover ricordare la linea guida, il medico deve ora interpretare come applicarla al caso specifico, che è esattamente il passaggio cognitivo più difficile. Un CDSS efficace va oltre: prende i dati del paziente, li confronta con la struttura logica della linea guida, e restituisce non “cosa dice la linea guida” ma “cosa significa la linea guida per questo paziente, in questo momento”.



### **GINOS come laboratorio di traduzione delle linee guida**

Il progetto GINOS, nella sua struttura più profonda, è un tentativo sistematico di costruire strumenti di traduzione delle linee guida in ostetricia e ginecologia. Ogni applicazione della suite parte da una o più linee guida esistenti — ESHRE, SIGO, ACOG, WHO, FIGO — e le traduce in un percorso decisionale che incorpora i dati specifici della paziente. MOC-Consult prende le linee guida sull’osteoporosi e le calibra sulla densitometria, sull’età, sui fattori di rischio e sulla storia clinica della paziente specifica. IST-Consult prende le linee guida sulle infezioni sessualmente trasmissibili e le adatta al profilo di rischio individuale. FIT-CAT traduce il consensus statement sull’interpretazione fisiopatologica del CTG in un workflow strutturato che guida il medico attraverso le variabili rilevanti per il caso specifico.

La nota clinica documentale che questi sistemi producono è, in questo senso, la materializzazione del processo di traduzione: non “la linea guida raccomanda X”, ma “in questa paziente, con

questi parametri, in questo contesto, la raccomandazione supportata dall'evidenza è X, per le seguenti ragioni". Un documento che non esiste in nessuna linea guida generica, perché è specifico per questa paziente, in questo momento. Ed è anche un documento medico-legale: dimostra che il ragionamento è stato condotto, che le variabili rilevanti sono state considerate, che la decisione è stata basata sull'evidenza disponibile e calibrata sulla situazione reale.

#### Dal knowledge gap all'action gap: dove intervengono i CDSS

Il knowledge gap è la distanza tra l'evidenza pubblicata e la conoscenza del medico. L'action gap è la distanza tra la conoscenza del medico e ciò che fa concretamente.

I CDSS agiscono principalmente sull'action gap:

- Portano la linea guida al momento della decisione
- Calibrano la raccomandazione sul caso specifico
- Documentano il ragionamento, riducendo il rischio medico-legale
- Rendono esplicite le variabili che il medico ha considerato

Un CDSS non sostituisce la linea guida.

La traduce in azione per questa paziente, in questo momento.



## Le linee guida tacciono: il valore del consenso clinico implicito

C'è però un territorio che nessuna linea guida copre, e che i CDSS hanno la possibilità di esplorare in modo unico. Le linee guida dicono cosa fare quando l'evidenza è solida. Ma nella pratica clinica quotidiana esistono molte situazioni in cui l'evidenza è assente, insufficiente o contraddittoria — e le linee guida rispondono con “valutazione individualizzata”, che è un modo elegante per dire: decidete voi. In questi spazi, i clinici si comportano comunque — sulla base dell'esperienza, della tradizione del reparto, del consiglio del collega più anziano. Questo

comportamento è spesso corretto, talvolta eccellente, ma è quasi sempre non scritto, non condivisibile, non trasferibile ad altri contesti.

Un CDSS costruito dal clinico-autore sulla propria pratica — come quelli di GINOS — può catturare questo sapere implicito e dargli una forma esplicita. Non con la pretesa di sostituire la ricerca formale, ma con la consapevolezza che in assenza di quella ricerca i clinici si comportano comunque, e che farlo in modo documentato, condiviso e verificabile è meglio di farlo in modo implicito e variabile. È questo il nucleo del metodo VIKI che guida lo sviluppo di GINOS: una validazione distribuita, iterativa, costruita dalla stessa comunità che usa gli strumenti. Non è scienza nel senso stretto del termine. È qualcosa che la scienza, da sola, non riesce a produrre: conoscenza pratica condivisa, in tempo reale, alla velocità di cui la clinica ha bisogno.



## Capitolo XV: Le barriere che resistono — perché i CDSS ancora faticano

### Il framework TASS: tecnica, stakeholder e società

Nel 2023, Li e colleghi (Li et al., 2023) pubblicarono una revisione sistematica sulle barriere all'uso dell'AI in medicina, organizzando i risultati intorno al framework TASS — Technical/Algorithm, Stakeholder, and Society. Era un tentativo di mappare sistematicamente il territorio degli ostacoli, e il quadro che emergeva era, nella sua completezza, quasi scoraggiante.

Sul piano tecnico, le barriere principali sono la mancanza di spiegabilità — i modelli di machine learning che performano meglio

sono spesso «scatole nere» di cui è impossibile comprendere il ragionamento interno — e la necessità di validazione su popolazioni diverse da quelle su cui i modelli sono stati addestrati.

Sul piano degli stakeholder, la resistenza non è più quella dei medici degli anni Settanta che temevano di essere sostituiti. È più sottile. È la preoccupazione che un sistema automatico possa erodere l'autonomia professionale, rendere i clinici dipendenti da raccomandazioni che non capiscono, creare una nuova forma di responsabilità giuridica difficile da attribuire. È la paura che l'AI, lungi dall'aiutare il medico, finisca per interporsi tra lui e il paziente, mediando una relazione che è, nella sua essenza, profondamente umana.

Sul piano sociale, ci sono questioni di equità — i dataset su cui vengono addestrati i modelli riflettono le disuguaglianze esistenti, e un modello addestrato su pazienti bianchi di mezza età americani non è necessariamente applicabile a pazienti africani, asiatici, latinoamericani. Ci sono questioni di privacy. Ci sono questioni di governance: chi decide quali CDSS vengono certificati? Chi è responsabile quando un sistema sbaglia?

## **Il paradosso dell'esperto**

Uno degli aspetti più controintuitivi emersi dalla ricerca recente riguarda l'effetto dei CDSS sui diversi livelli di expertise clinica. Ci si aspetterebbe che i sistemi di supporto decisionale aiutassero soprattutto i medici meno esperti — quelli che hanno meno esperienza da cui attingere, che sono più vulnerabili alle trappole cognitive, che beneficerebbero maggiormente di una guida strutturata.

E in parte è così. Ma la letteratura rivela un paradosso: i CDSS basati su AI possono essere meno utili, o addirittura dannosi,

per i clinici più esperti. L'explainable AI<sup>9</sup> — i sistemi che mostrano il loro ragionamento — aumenta la fiducia dei clinici nelle raccomandazioni. Ma aumenta quella fiducia in modo relativamente acritico: un clinico esperto che vede una raccomandazione AI accompagnata da una spiegazione plausibile tende a seguirla anche quando la sua esperienza gli suggerirebbe qualcosa di diverso.

In altri termini: l'AI spiegabile può aumentare la fiducia nelle raccomandazioni errate. Può rendere l'esperto meno esperto, abbassando la sua vigilanza critica in favore di una sicurezza statistica che non è infallibile.



## Capitolo XVI: L'implementazione come arte — la scienza del cambiamento

### Il framework PRISM e il design centrato sull'utente

Come si fa a portare un CDSS dalla ricerca alla pratica clinica reale? Questa domanda — apparentemente tecnica — è in realtà profondamente umana. Riguarda il comportamento delle organizzazioni, la psicologia degli operatori sanitari, le dinamiche di potere nei reparti, la resistenza al cambiamento che è una delle costanti più stabili della natura umana.

La ricerca sull'implementazione dei CDSS ha sviluppato negli ultimi anni framework sofisticati per affrontare questo problema. Uno dei più completi è il PRISM — Practical Robust Implementation and Sustainability Model — che integra le best practices di progettazione dei CDSS con i principi della scienza dell'implementazione.

<sup>9</sup>Odds ratio (OR): misura statistica che esprime quanto è più (o meno) probabile un evento in un gruppo rispetto a un altro. OR = 1.75 significa che l'evento è 1.75 volte più probabile nel gruppo studiato. Ampiamente usato negli studi clinici e nelle meta-analisi.

Il PRISM identifica cinque fasi che interagiscono in modo iterativo — non sequenziale. Prima il coinvolgimento multilivello degli stakeholder: i medici, le infermiere, gli amministratori, i pazienti devono essere parte del processo di sviluppo, non destinatari passivi di un sistema calato dall'alto. Poi la progettazione del CDSS: tenendo conto dei workflow reali, non di quelli idealizzati. Poi il test di usabilità: perché un sistema che funziona perfettamente in laboratorio può essere inutilizzabile in reparto. Poi l'implementazione ponderata: graduale, supportata, monitorata. Poi la valutazione continua: il sistema deve imparare dall'uso reale tanto quanto l'uso reale deve adattarsi al sistema.

La revisione sistematica del 2025 sullo User-Centered Design dei CDSS (Bayor et al., 2025) ha trovato che il design centrato sull'utente è l'approccio più efficace per garantire l'adozione. E ha identificato quattro categorie di sfide che qualsiasi progettista deve affrontare: usabilità e esperienza utente, validità e assicurazione della qualità, gestione dei dati, complessità di integrazione nei sistemi esistenti.

## I cinque diritti del CDS

Una delle raccomandazioni pratiche più eleganti della letteratura è quella dei Five Rights of Clinical Decision Support: per essere efficace, un CDSS deve fornire l'informazione giusta, alla persona giusta, nel formato giusto, attraverso il canale giusto, al momento giusto.

Sembra ovvio. Ma ogni clinico che ha lavorato con sistemi informatici ospedalieri sa quanto raramente questa condizione sia soddisfatta. Alert che arrivano nel momento sbagliato, a operatori che non hanno il potere di agire su di essi, in un formato che richiede troppo tempo per essere elaborato, attraverso un canale

che nessuno controlla — questi sono i CDSS che non vengono usati. Questi sono i sistemi che alimentano l'alert fatigue, che erodono la fiducia, che finiscono per essere cliccati via senza essere letti.



## Epilogo: L'onda che verrà

### Il momento eccezionale e inquietante

Nell'agosto del 2023, Mustafa Suleyman pubblicò un libro che Bill Gates definì «il mio libro preferito sull'AI» e che Daniel Kahneman — il Nobel che ha passato una vita a studiare i bias del pensiero umano — considerò «lettura essenziale». Il titolo era *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma*.

Suleyman non è un osservatore esterno della tecnologia che commenta da lontano. È uno dei suoi costruttori. Ha co-fondato DeepMind nel 2010 — l'azienda che ha dimostrato che una rete neurale poteva battere il campione del mondo a Go — ha guidato l'AI di Google, ha fondato Inflection AI, ed è oggi CEO di Microsoft AI. Sa di cosa parla. E quello che dice è inquietante nella sua onestà.

*"Stiamo avvicinandoci a una soglia critica nella storia della nostra specie. Tutto sta per cambiare."*

L'argomento centrale di Suleyman è il «containment problem» — il problema della contenzione. Le tecnologie più potenti della storia umana hanno sempre avuto una caratteristica: sono state difficili da contenere una volta rilasciate. L'energia nucleare, gli antibiotici, Internet — ciascuna ha prodotto benefici immensi e rischi altrettanto immensi, e il confine tra i due si è

rivelato più sfumato e più poroso di quanto i loro creatori si aspettassero. L'AI — e in particolare l'AI generativa dei Grandi Modelli Linguistici — è la più potente di queste tecnologie, e il suo problema di contenzione è il più urgente.

Per chi lavora in medicina, questa riflessione non è astratta. L'AI che identifica il cancro cervicale meglio di un patologo è la stessa tecnologia che potrebbe produrre informazioni mediche false con la stessa fluidità e la stessa apparente autorevolezza. Il sistema che predice la preeclampsia settimane prima dei sintomi è costruito sulle stesse architetture che potrebbero generare falsi profili di pazienti, falsificare documenti clinici, manipolare decisioni terapeutiche. Il confine tra strumento e minaccia è sottile, e attraversarlo non richiede cattive intenzioni — richiede soltanto negligenza, superficialità, uso acritico.

Siamo in un momento eccezionale. La storia dei CDSS che abbiamo raccontato in queste pagine — dal laboratorio di Brodman nel 1959 alla sala ecografica del 2025 — è stata una storia di promesse lente a mantenersi, di sistemi brillanti che non raggiungevano le corsie, di gap enormi tra la ricerca e la pratica. Improvvisamente, nel giro di tre o quattro anni, quel gap si è assottigliato in modo vertiginoso. Non perché i problemi tecnici siano stati risolti tutti — non lo sono — ma perché la soglia di accessibilità si è abbassata in modo rivoluzionario.

### **La mammografia che guarda il cuore — una scoperta non cercata**

C'è un esempio che racconta meglio di qualsiasi altro la natura del cambiamento che stiamo vivendo. Nell'ambito dello screening del cancro alla mammella, i ricercatori hanno sviluppato algoritmi di deep learning per analizzare le mammografie e

identificare le lesioni maligne. Algoritmi addestrati su centinaia di migliaia di immagini, ottimizzati per riconoscere la differenza tra una massa benigna e un carcinoma.

A un certo punto, i ricercatori hanno fatto una scoperta che non stavano cercando. Analizzando le immagini mammografiche con gli stessi algoritmi, il sistema aveva imparato a identificare qualcosa che all'occhio umano sembrava irrilevante: le calcificazioni delle arterie mammarie — sottili depositi di calcio che si formano nelle pareti dei vasi arteriosi del seno, visibili sulle mammografie come tracce bianche quasi impercettibili. Per decenni, queste calcificazioni erano state considerate un reperto incidentale, privo di significato clinico rilevante, annotato e dimenticato.

L'AI aveva capito quello che generazioni di radiologi non avevano formalizzato in una regola: la calcificazione arteriosa mammaria è un marker di aterosclerosi sistemica — la stessa malattia che causa infarto e ictus. Le donne con calcificazioni arteriose mammarie significative hanno un rischio cardiovascolare molto più elevato rispetto a quelle senza. Uno studio pubblicato sull'European Heart Journal nel 2026 (Dapamede et al., 2026), condotto su oltre 123.000 donne, ha dimostrato che il sistema AI riusciva a quantificare automaticamente la calcificazione arteriosa e a stratificare il rischio cardiovascolare con un'accuratezza che superava i tradizionali score di rischio come lo score PREVENT<sup>10</sup> — e funzionava persino nelle donne sotto i cinquant'anni, quella fascia d'età in cui il rischio cardiovascolare è tipicamente sottostimato.

Come ha fatto? La risposta è semplice nella sua portata rivoluzionaria: aveva tanti dati digitalizzati — le mammografie di

---

<sup>10</sup>AUC — Area Under the Curve (area sotto la curva ROC): misura statistica della capacità discriminante di un test diagnostico. Varia da 0.5 (test inutile, equivalente al caso) a 1.0 (test perfetto). Valori > 0.85 sono generalmente considerati eccellenti in ambito clinico.

centinaia di migliaia di donne, con le loro storie cliniche di follow-up — e l'AI ha trovato la correlazione. Nessuno gliel'aveva detto di cercarla. Nessuno aveva formulato l'ipotesi da testare. Il sistema ha semplicemente riconosciuto un pattern statistico in dati che erano disponibili da sempre, ma che nessuna mente umana avrebbe avuto la capacità computazionale di elaborare nella loro interezza.

Il risultato pratico è straordinario: una mammografia di screening, già eseguita per un altro scopo, può diventare uno strumento di prevenzione cardiovascolare. Senza costi aggiuntivi, senza radiazioni extra, senza visite supplementari. In un colpo solo, lo screening del cancro del seno diventa anche un programma di salute cardiovascolare femminile — raggiungendo quella vasta popolazione di donne che si sottopone regolarmente alla mammografia ma non conosce il proprio profilo lipidico, non monitora la pressione arteriosa, non ha mai avuto una valutazione del rischio cardiovascolare.

## **Centauro, Cyborg — e la domanda che resta**

Nel 2024, Ethan Mollick — professore alla Wharton School of Business dell'Università della Pennsylvania, uno degli studiosi che più lucidamente ha analizzato come le persone lavorano con l'AI — pubblicò un libro intitolato *Co-Intelligence: Living and Working with AI*. Nel libro, Mollick propose due metafore per descrivere i diversi modi in cui l'essere umano può collaborare con l'intelligenza artificiale.

Il primo modello è il Centauro — la creatura della mitologia greca, metà uomo e metà cavallo, con una divisione netta tra le due nature. Il Centauro divide i compiti strategicamente: l'uomo fa quello che sa fare meglio, la macchina fa quello che sa fare meglio, e c'è una linea chiara tra i due territori. Il medico che usa un CDSS

per calcolare uno score di rischio e poi interpreta il risultato alla luce della storia clinica del paziente è un Centauro: lui decide la strategia, la macchina esegue il calcolo.

Il secondo modello è il Cyborg — l'essere ibrido della fantascienza, in cui macchina e uomo sono profondamente integrati, indistinguibili nel flusso del lavoro. Il Cyborg non trasferisce compiti alla macchina: ci lavora insieme in tempo reale, in un dialogo continuo dove i confini tra pensiero umano e elaborazione artificiale si sfumano. Il medico che costruisce una nota clinica conversando con un modello linguistico, che affida alla macchina certi passaggi del ragionamento e poi li riprende e li reindirizza con il proprio giudizio clinico — questo è un Cyborg.

Mollick non propone una gerarchia tra i due modelli. Entrambi hanno i loro contesti d'uso appropriati. Il Centauro è più adatto quando il confine tra competenze è chiaro e quando l'autonomia umana è essenziale — il giudizio finale, la comunicazione con il paziente, la valutazione del contesto. Il Cyborg è più produttivo quando il problema è abbastanza complesso da richiedere un'elaborazione continua e bidirezionale tra intelligenza umana e artificiale.

Ma entrambi sollevano la stessa domanda fondamentale: chi guida? In entrambi i modelli, la risposta che Mollick dà — e che chi scrive condivide profondamente — è la stessa: l'essere umano. Non come supervisore passivo che approva le raccomandazioni di una macchina. Come autore. Come responsabile. Come la mente che porta a questo sistema la conoscenza del contesto, il senso del valore, la coscienza delle conseguenze.



## Nota finale: Fare il medico nel 2026

Francis Timothy de Dombal, il chirurgo di Leeds che nel 1971 costruì il primo sistema probabilistico per la diagnosi del dolore addominale acuto, fece una scoperta che ci mise vent'anni a capire appieno. Il suo sistema batteva i clinici esperti nell'accuratezza diagnostica. Ma il beneficio maggiore, scrisse nel 1991 dopo aver studiato quasi centomila pazienti in tutta Europa, non era questo. Era che il computer funzionava come «un focus educativo e uno stimolo alla buona pratica clinica». Non sostituiva il ragionamento del medico. Lo attivava. Lo rendeva più rigoroso.

Trent'anni dopo, Joseph Habboushe e Graham Walker — due medici d'urgenza di New York — costruirono MDCalc: una piattaforma web gratuita di score clinici e algoritmi decisionali. Nessun finanziamento esterno. Nessun team di informatici. Solo la convinzione che i medici non dovessero sprecare tempo a memorizzare formule e che gli strumenti clinici non dovessero essere limitati dalla loro memorizzabilità. MDCalc conta oggi oltre dieci milioni di utenti mensili ed è usato in ogni angolo del mondo. Due medici, un'idea semplice, nessuna barriera tecnica.

Questo è il filo che attraversa tutta la storia raccontata in queste pagine. Da Brodman a Myers, da De Dombal a Habboushe e Walker, da GinHelp e GynDesk — costruiti con il prof. Serra — allo score CTG della sala parto del Cristo Re, fino a GINOS: l'idea che il clinico non sia soltanto l'utente di un sistema progettato da altri, ma il suo possibile autore. Che la conoscenza accumulata in anni di pratica al letto del paziente — quella conoscenza implicita, contestuale, difficilissima da verbalizzare — possa essere tradotta in strumenti utili. Prima ci volevano anni di lavoro con un informatico. Oggi ci vuole un pomeriggio e un modello linguistico.

Il modello che propongo — e che uso ogni giorno — è quello che Ethan Mollick chiama lavoro da centauro: una chiara divisione tra le competenze umane e quelle artificiali. L'umano porta la conoscenza del dominio, il giudizio clinico, il senso della responsabilità, la coscienza dei confini. La macchina porta la capacità di elaborazione, la velocità, la sintesi, la traduzione della logica in codice. Il centauro non abdica. Sceglie, verifica, firma. Come ha scritto Kasparov — che inventò questo concetto dopo aver perso contro Deep Blue e poi aver capito che un uomo con una macchina batteva sia l'uomo da solo sia la macchina da sola — la risposta non è rinunciare agli strumenti. È imparare a cavalcarli.

Un gruppo di medici fondatori di MDCalc ha scritto la Physicians' Charter for Responsible AI — dieci principi etici per chi costruisce e usa strumenti di intelligenza artificiale in medicina. Chi scrive l'ha sottoscritta. Costruire bene non basta: bisogna anche impegnarsi a non costruire male.

Questo libro è stato scritto secondo lo stesso modello. Le idee, il filo narrativo, i personaggi storici, le tesi argomentative sono miei. I contenuti sono stati sviluppati con il supporto di Claude (Anthropic). Io ho deciso la strategia, verificato ogni passaggio, mi sono assunto la responsabilità del risultato. Questo non mi ha reso meno autore: mi ha reso un autore che usa strumenti nuovi con spirito critico vecchio.

Quando ero giovane ostetrico al Cristo Re e costruivo il nostro score CTG con il prof. Serra, non sapevo di stare costruendo un CDSS. Non sapevo chi fosse Shortliffe. Non sapevo che a Leeds vent'anni prima De Dombal aveva fatto qualcosa di simile e aveva trovato la stessa cosa: che lo strumento non sostituiva il ragionamento, lo rendeva più rigoroso.

Oggi lo so. E so anche che quello che stavamo cercando di fare — rendere il ragionamento clinico esplicito, verificabile, condivisibile, documentabile — non era uno strano hobby tecnico di un medico con troppe serate libere. Era un bisogno profondo della medicina, che non aveva ancora gli strumenti per essere soddisfatto.

Adesso quegli strumenti ci sono.

*Carlo Piscicelli*  
*Roma, aprile 2026*



## Dichiarazione di trasparenza

---

Questo libro è stato scritto da un centauro.

Nel senso che Ethan Mollick dà a questa parola nel suo libro *Co-Intelligence*: il centauro non è né l'uomo che scrive da solo né la macchina che genera da sola. È una collaborazione strutturata in cui ciascuno fa ciò che sa fare meglio, con una linea chiara tra i due territori e una responsabilità finale che appartiene solo alla metà umana.

Il materiale scientifico di partenza — revisioni sistematiche, meta-analisi, studi clinici, dati epidemiologici — è stato elaborato attraverso OpenEvidence, una piattaforma che usa modelli linguistici ancorati alla letteratura verificata per sintetizzare l'evidenza biomedica con citazioni esplicite e tracciabili. La struttura narrativa del saggio, la selezione dei personaggi storici, gli archi temporali, le interpretazioni, le analogie e le metafore sono stati sviluppati in dialogo con Claude (Anthropic, versione Sonnet 4), usato come interlocutore per trasformare la materia grezza dell'evidenza scientifica in racconto leggibile.

Il processo non è stato quello di chiedere a una macchina di scrivere al posto mio. È stato un dialogo continuo, iterativo, a volte contraddittorio. Io indicavo la direzione, proponevo il tono, stabilivo le priorità, valutavo ogni proposta, correggevo ciò che non era clinicamente accurato o narrativamente efficace, aggiungevo tutto ciò che l'AI non poteva sapere — perché viene dall'esperienza diretta, dalla memoria della sala parto del Cristo Re, dalle notti di

guardia, dalle conversazioni con il prof. Serra, dagli errori fatti e dai casi difficili.

L'AI proponeva. Io decidevo. Questa asimmetria è essenziale e non negoziabile.

Le idee centrali del libro — il modello del clinico-autore, la visione di GINOS come democrazia clinica, il metodo VIKI, la distinzione tra knowledge gap e implementation gap, la lettura della storia dei CDSS come storia di promesse troppo presto — sono mie. Alcune le ho elaborate pensando ad alta voce con Claude; altre le ho portate già formate alla macchina chiedendole di dargli forma narrativa. In entrambi i casi, la responsabilità intellettuale è interamente mia.

Non ho usato l'AI per generare citazioni bibliografiche autonome. Ogni riferimento bibliografico riportato è tratto da fonti verificate dalla piattaforma OpenEvidence o da letteratura di cui ho verificato personalmente l'esistenza e la pertinenza.

Ho adottato questo approccio per una ragione precisa: credo che sia il modo onesto di lavorare nell'era in cui viviamo. Nascondere l'uso dell'AI nella scrittura accademica e professionale è diventato, nel 2026, una forma di disonestà intellettuale — tanto quanto sarebbe stato disonesto, vent'anni fa, non dichiarare di aver usato un software statistico per l'analisi dei dati. Gli strumenti cambiano. La responsabilità dell'autore resta.

Ho anche sottoscritto la Physicians' Charter for Responsible AI, promossa dai fondatori di MDCalc — dieci principi etici per chi costruisce e usa strumenti di intelligenza artificiale in medicina. Uno di quei principi è la trasparenza. Questa pagina ne è l'applicazione.

*Carlo Piscicelli*  
*Roma, aprile 2026*  
[physicianscharter.ai](http://physicianscharter.ai) • [ginos-cdss.it](http://ginos-cdss.it)

# Il Progetto GINOS

*Gynecology Integrated Network Operating System*

*GINecologia OStetricia*

---

## Un progetto, una filosofia

GINOS è una suite gratuita di applicazioni web di Clinical Decision Support per la pratica clinica in ostetricia e ginecologia. Ma dire che è una «suite di applicazioni» è come dire che Wikipedia è «un sito con degli articoli». La descrizione è tecnicamente corretta e completamente incapace di catturare l'idea che ci sta dietro.

L'idea è questa: la medicina clinica produce ogni giorno una quantità immensa di conoscenza pratica che non finisce mai nelle linee guida. Finisce nelle teste dei medici, nelle abitudini dei reparti, nelle decisioni prese alle tre di notte davanti a un caso difficile. Questa conoscenza è reale, è collaudata, è preziosa — e quasi sempre non è scritta da nessuna parte. GINOS nasce dal tentativo di darle una forma digitale accessibile, condivisibile, migliorabile.

## Il problema che GINOS vuole risolvere

Esiste un paradosso nel cuore della medicina basata sull'evidenza. Le linee guida — i documenti che raccolgono le prove scientifiche e le traducono in raccomandazioni cliniche — sono lo strumento più importante che abbiamo per standardizzare la buona pratica medica. Ma hanno un limite strutturale che chiunque lavori in una corsia conosce bene: non coprono tutto.

Le linee guida dicono cosa fare quando l'evidenza è solida. Ma spesso non dicono niente — o dicono «non ci sono prove sufficienti per raccomandare» — in quelle zone grigie dove invece il clinico deve decidere comunque. Un paziente non aspetta che la letteratura raggiunga il consenso. Un neonato in distress non può aspettare la prossima meta-analisi.

E allora cosa succede in quei vuoti? Succede quello che è sempre successo: i medici si comportano in base all'esperienza, alla tradizione del reparto, al consiglio del collega più anziano, al buon senso clinico accumulato in anni di pratica. Si comportano bene, spesso meglio di quanto le linee guida — se esistessero — potrebbero suggerire. Ma lo fanno in modo non sistematico, non documentato, non condivisibile.

C'è poi un altro elemento, più sottile e più imbarazzante da nominare. Anche quando le prove ci sono, e anche quando esiste un consenso informale tra i clinici su come comportarsi, nessun accademico vuole mettere la propria firma sotto una raccomandazione che non sia solidamente supportata dalla letteratura. Il rischio reputazionale è reale. Il risultato è che certi comportamenti clinici ampiamente condivisi e razionalmente fondati restano nell'ombra — praticati da tutti, codificati da nessuno.

GINOS vuole entrare esattamente in questo spazio. Non per sostituire le linee guida dove esistono. Ma per costruire, in modo esplicito e trasparente, raccomandazioni condivise laddove le linee guida tacciono o rimangono vaghe. Con un metodo preciso: il metodo VIKI.

## Il metodo VIKI: validazione iterativa per intelligenza condivisa

VIKI non è un acronimo tecnico. È il nome che abbiamo dato a un processo che somiglia a quello con cui Wikipedia costruisce la propria enciclopedia — ma applicato alla conoscenza clinica pratica.

Il principio è semplice. Un'applicazione viene costruita dal clinico-autore sulla base della propria esperienza, delle linee guida disponibili, e della letteratura scientifica pertinente. Viene poi rilasciata e utilizzata nella pratica reale. Ed è qui che comincia il vero lavoro: man mano che i colleghi la usano, emergono gli errori. Un percorso logico che non tiene conto di un caso particolare. Una soglia che nella pratica non funziona. Una raccomandazione che confligge con le linee guida di una società scientifica che l'autore non aveva considerato. Un'opzione terapeutica mancante.

Questi errori vengono segnalati — attraverso un sistema di feedback integrato nell'applicazione — e vengono corretti. L'applicazione viene aggiornata. La versione successiva è migliore. E il ciclo ricomincia.

Non è validazione nel senso accademico del termine — con trial clinici randomizzati, popolazioni di controllo, pubblicazioni su riviste peer-reviewed. È qualcosa di diverso e, per certi versi, di più potente: è validazione distribuita, continua, contestuale. È la stessa comunità che usa lo strumento che ne certifica l'adeguatezza — non come atto formale, ma come processo vivo.

Questo approccio riconosce una verità che la ricerca sull'implementazione dei CDSS ha impiegato decenni a formalizzare: i sistemi calati dall'alto, costruiti da esperti lontani dalla pratica e consegnati ai clinici come prodotti finiti, raramente funzionano. I sistemi che funzionano sono quelli costruiti con i

clinici, verificati dai clinici, corretti dai clinici. VIKI non è un compromesso tra rigore scientifico e praticità. È una forma diversa di rigore — più lenta nella costruzione dell'autorevolezza, ma più veloce nell'adattamento alla realtà.

## Una comunità che esprime i propri bisogni

Il secondo pilastro del progetto è la comunità. GINOS non è pensato come un catalogo fisso di applicazioni prodotte da un autore solitario. È pensato come una piattaforma aperta, in cui i clinici che la usano contribuiscono non soltanto a correggere ciò che esiste, ma anche a definire ciò che manca.

La domanda che vogliamo che ogni utente si faccia è questa: in quale momento della mia giornata clinica mi sono trovato a non sapere esattamente come procedere? In quale situazione avrei voluto avere uno strumento che mi aiutasse a ragionare in modo sistematico? Qual è il problema clinico ricorrente per cui non esiste ancora un percorso strutturato?

Queste domande — se raccolte, aggregate e analizzate — disegnano una mappa dei bisogni reali della pratica clinica. Non dei bisogni ipotizzati dai ricercatori in un laboratorio universitario. Non delle priorità stabilite dalle società scientifiche in base alla disponibilità di fondi. I bisogni di chi lavora ogni giorno con i pazienti.

È una forma di democrazia clinica. E l'intelligenza artificiale è lo strumento che rende possibile tradurre questi bisogni in strumenti concreti nel giro di ore — non di mesi o anni come sarebbe stato necessario con i metodi tradizionali di sviluppo software.

## Navigare fuori dal proprio territorio

Uno dei casi d'uso più concreti e più frequenti di GINOS riguarda ciò che potremmo chiamare il problema della competenza di confine. La medicina moderna è profondamente specializzata. Un ostetrico di medicina perinatale passa le giornate a gestire gravidanze ad alto rischio — è il suo territorio, lo conosce in profondità. Ma il suo territorio ha dei confini. E quei confini, nella pratica clinica, vengono attraversati ogni giorno.

Una donna con una gravidanza gemellare complicata può avere una storia di tumore ovarico. L'ostetrico è l'esperto della gravidanza, l'oncologo è l'esperto del tumore. Ma nella sala visita è l'ostetrico che deve guidare la paziente, che deve integrare le informazioni oncologiche nel percorso ostetrico, che deve rispondere alle domande su farmaci, follow-up, timing del parto in relazione alla terapia oncologica. In quel momento, l'ostetrico si trova a navigare in un territorio che non è il suo.

Analogamente, il ginecologo oncologo che gestisce una paziente in follow-up per un carcinoma endometriale si trova a dover rispondere a domande sulla menopausa iatrogena, sulla terapia ormonale sostitutiva in quel contesto specifico, sulle implicazioni riproduttive. Non è il suo territorio principale. Ma è la sua paziente.

GINOS vuole essere la mappa per navigare in quei territori di confine. Non per sostituire il consulente specialista — che rimane insostituibile nei casi complessi — ma per dare al clinico generalista o allo specialista d'organo un filo di Arianna quando si trova in un labirinto che non è il suo. Un percorso logico strutturato, basato sulle linee guida pertinenti, che aiuta a non perdere i passaggi essenziali e a identificare il momento in cui la consulenza specialistica è necessaria.

## Colmare i vuoti — con il contributo di tutti

C'è un esempio che illustra perfettamente il tipo di problema che GINOS vuole affrontare. Le linee guida sulla gestione del diabete gestazionale sono dettagliate e ben costruite per quanto riguarda la diagnosi, i target glicemici, la terapia insulinica. Ma diventano vaghe quando si tratta di alcune situazioni cliniche specifiche: la donna con diabete gestazionale che deve essere indotta prima del termine, la gestione perioperatoria per il taglio cesareo programmato, il monitoraggio glicemico nel postpartum immediato. Su questi punti, le linee guida dicono spesso «si raccomanda la valutazione individualizzata» — che è un modo elegante per dire «non lo sappiamo con certezza, decidete voi».

Ma i clinici non possono rispondere «non lo so». Devono decidere. E nella stragrande maggioranza dei centri, decidono bene — sulla base di protocolli interni, di esperienze condivise, di buon senso clinico. Protocolli che funzionano, che sono stati aggiustati nel tempo, che incorporano la saggezza pratica di anni di gestione di quei casi.

Il progetto GINOS vuole raccogliere quella saggezza. Non con la pretesa di sostituire la ricerca scientifica formale — che rimane lo standard per la medicina evidence-based. Ma con la consapevolezza che, in assenza di quella ricerca, i clinici si comportano comunque. E che comportarsi in modo esplicito, documentato e condiviso è meglio di comportarsi in modo implicito, variabile e non tracciabile.

Il metodo VIKI applicato a questi vuoti funziona così: una prima versione dell'algoritmo viene costruita raccogliendo i protocolli esistenti nei centri più qualificati, la letteratura disponibile anche quando non conclusiva, e il consenso informale dei professionisti più esperti. Viene resa disponibile. I clinici la

usano, la mettono alla prova nella pratica reale, segnalano dove non funziona. La versione successiva è migliore. E così via — non verso una perfezione irraggiungibile, ma verso un'utilità pratica crescente.

Non è scienza nel senso stretto del termine. È qualcosa che la scienza, da sola, non riesce a produrre: conoscenza pratica condivisa, in tempo reale, alla velocità di cui la clinica ha bisogno.

**Accedi alla suite su:**

**[ginos-cdss.it](https://ginos-cdss.it)**



*Inquadra il codice QR con la fotocamera del tuo smartphone*

## Fonti e riferimenti bibliografici principali

Le informazioni contenute in questo saggio si basano su revisioni sistematiche, meta-analisi e studi originali pubblicati in letteratura biomedica internazionale, elaborati attraverso la piattaforma OpenEvidence (2025). Di seguito i principali riferimenti tematici.

**CDSS: storia e architettura.** Wright A, Sittig DF (2008). *Int J Med Inform.* • Hunt DL et al. (1998). *JAMA.* • Sutton RT et al. (2019). *NPJ Digit Med.* • Kierner S et al. (2023). *J Biomed Inform.* • McDonald CJ (1976). *Ann Intern Med.*

**MYCIN e sistemi esperti.** Shortliffe EH (1976). *Computer-Based Medical Consultations: MYCIN.* Elsevier. • Yu VL et al. (1979). *JAMA.* • Schurink CA et al. (2005). *Lancet Infect Dis.*

**Approcci probabilistici (Ledley, Lusted, Warner).** Ledley RS, Lusted LB (1959). *Science.* • Warner HR et al. (1961). *Am J Cardiol.* • Fieschi M (1990). *Artificial Intelligence in Medicine.* Springer.

**INTERNIST-I.** Miller RA, Pople HE, Myers JD (1982). *N Engl J Med.* • Wolfram DA (1995). *Artif Intell Med.*

**De Dombal e i pionieri clinici.** de Dombal FT et al. (1972). *BMJ.* • de Dombal FT (1991). *Ann Chir.*

**Grandi Modelli Linguistici.** Bubeck S et al. (2023). *Sparks of Artificial General Intelligence.* Microsoft Research. • Wei J et al. (2022). *Emergent Abilities of Large Language Models.* TMLR. • Haug CJ, Drazen JM (2023). *N Engl J Med.*

**Machine learning in medicina.** Haug CJ, Drazen JM (2023). N Engl J Med. • Quer G et al. (2021). JACC. • Loftus TJ et al. (2020). JAMA Surg.

**CDSS in ostetricia e ginecologia.** Lin X et al. (2024). J Med Internet Res. • Cockburn N et al. (2024). EClinicalMedicine. • Devoe LD et al. (2025). Obstet Gynecol. • Du Y et al. (2023). Int J Med Inform.

**Screening cancro cervicale.** Liu L et al. (2025). EClinicalMedicine (meta-analisi 77 studi). • Ouh YT et al. (2024). Sci Rep (CerviCARE AI). • Yang W et al. (2024). BMC Cancer.

**Rischio ostetrico e machine learning.** Mennicken D et al. (2023). Front Endocrinol. • Pi X et al. (2025). Sci Rep (Bangladesh, n=1014). • Bertini A et al. (2022). Front Bioeng Biotechnol (revisione 31 studi).

**Ecografia ostetrica e AI.** Horgan R et al. (2023). Prenat Diagn (127 articoli). • Ramirez Zegarra R, Ghi T (2023). Ultrasound Obstet Gynecol. • Jost E et al. (2023). J Clin Med.

**Mammografia e rischio cardiovascolare (BAC).** Dapamede T et al. (2026). European Heart Journal (n=123.762). • Allen T et al. (2024). JACC Advances. • Chidurala S et al. (2025). Cureus. • Barraclough JY et al. (2025). Heart.

**AI e collaborazione umana — Centauro e Cyborg.** Mollick E (2024). Co-Intelligence: Living and Working with AI. Portfolio/Penguin. • Randazzo S, Mollick E et al. (2025). Harvard Business School Working Paper 26-036.

**L'onda tecnologica — Suleyman.** Suleyman M, Bhaskar M (2023). The Coming Wave: Technology, Power, and the Twenty-first

Century's Greatest Dilemma. Crown. (Bill Gates: 'My favourite book on AI').

**Barriere all'implementazione.** Li LT et al. (2023). J Biomed Inform (framework TASS). • Trinkley KE et al. (2020). J Med Internet Res. • Kawamoto K, McDonald CJ (2020). Ann Intern Med. • Baylor AA et al. (2025). J Med Internet Res. • Abell B et al. (2023). Implement Sci.

● CTG e monitoraggio intrapartum. INFANT Collaborative Group (2017). Lancet (trial INFANT, 46.042 pz). • Lopes-Pereira J et al. (2019). Am J Obstet Gynecol (SisPorto Omniview, 38.466 parti). • Bernardes J (2023). J Perinat Med (SisPorto: 32 anni di sviluppo). • Aeberhard JL et al. (2023). Eur J Obstet Gynecol Reprod Biol (scoping review AI-CTG, 40 algoritmi). • Balayla J, Shrem G (2019). Arch Gynecol Obstet (meta-analisi AI-CTG, 55.064 pz). • Campanile M et al. (2020). J Matern Fetal Neonatal Med (meta-analisi RCT CTG computerizzato, 54.492 pz). • ACOG Clinical Practice Guideline No. 10 (2025). Obstet Gynecol (raccomandazione contro uso primario CDSS-CTG). • Brown J et al. (2023). Aust NZ J Obstet Gynaecol (13 linee guida CTG internazionali). • Devoe L et al. (2000). Am J Obstet Gynecol (variabilità inter-osservatore NICHHD). • Lamé G et al. (2024). BMJ Qual Saf (CTG come attività sociotecnica). • Lin S et al. (2026). BMC Medicine (CAPs multicenter, deep learning vs esperti umani). • Gumilar KE et al. (2024). Comput Struct Biotechnol J (LLM vs medici, 7 tracciati). • Psilopatis I et al. (2025). Arch Gynecol Obstet (confronto LLM su 60 tracciati FIGO).

● Interpretazione fisiopatologica del CTG. Chandraharan E, Pereira S, Ghi T et al. (2024). Eur J Obstet Gynecol Reprod Biol (International Expert Consensus Statement on Physiological Interpretation of CTG, revisione). • Jia YJ, Ghi T, Pereira S et al. (2023). Am J Obstet Gynecol (interpretazione fisiopatologica CTG in pratica clinica). • Jia YJ et al. (2021). J Matern Fetal Neonatal Med (studio 500 tracciati, fasi di compensazione e scompenso). • Finadri L, Mappa I, Khalil A et al. (2026). J Perinat Med (meta-analisi ZigZag, 18.136 feti). • Tarvonen M et al. (2021). Acta Obstet Gynecol Scand (ZigZag segno precoce ipossia, 4.988 parti). • Chandraharan E et al. (2024). Med Res Arch (FIT-CAT — FIT for labour Clinical Anticipation Tool). • Qiu T et al. (2025). Front Physiol (sistema bimodale CTG+contrazioni, DenseNet, AUC 0.944). • Mendis L et al. (2026). IEEE Trans Biomed Eng (Fusion ResNet, AUC 0.84, validazione esterna).

● TWERIS Mini. Chandraharan E et al. (2025). Med Res Arch (accordo TWERIS Mini vs esperto: 94%, Kappa 0.92; gestione: 98%, Kappa 0.96; 100 tracciati CTG). • Physiological CTG Interpretation Group (2025). Med Res Arch (accordo inter-osservatore editorial board vs panel vs TWERIS Mini; 30 tracciati, 5.724 immagini di addestramento).

---

Dr. Carlo Piscicelli — Roma, 2026