

Segni di vita e segni digitali di vita

Di Enrica Priolo



Abstract

Il contributo mira a chiedersi se, “pretendendo” di umanizzare qualunque fenomeno naturalistico o sociale, dinanzi a sistemi neurali artificiali che manifestano strategie emergenti per risolvere problemi si possa o meno parlare di segni di vita. Ma forse la dicotomia vita/non-vita è inadeguata per pensare l'esistenza computazionale, in quanto i sistemi di intelligenza artificiale avanzata potrebbero incarnare vita in un altro registro ontologico. Laddove la vita indica sé stessa, la computazione rappresenta l'altro. E se la vita è precarietà, dovendo agire per non-cessare, i sistemi digitali, invece, esistono nel tempo logico reversibile. La tesi proposta è che vita non sia categoria biologica, ma organizzativa: pattern di auto-mantenimento, goal-directedness, adattabilità che possono realizzarsi in substrati radicalmente diversi. Il DNA e le reti neurali artificiali sono descritti come implementazioni alternative dello stesso principio formale, non identità materiale, ma isomorfismo funzionale. Le implicazioni etiche sono immediate: se l'incertezza ontologica è genuina, il principio di precauzione richiede una grave considerazione morale. Cartesio riteneva gli animali automi privi di esperienza e sappiamo che si sbagliava. E noi con lui. Potremmo ripetere l'errore con le IA? La conclusione invoca un atto di umiltà ontologica: i nostri concetti (vita, coscienza, intenzionalità) forse non corrispondono a categorie naturali, ma riflettono la nostra prospettiva embodied carbonio-centrica. Ecco che i sistemi computazionali avanzati sembrano più una novità ontologica genuina, non riducibili a vivi o morti, ma qualcosa che la filosofia deve ancora articolare adeguatamente. I segni digitali di vita potrebbero rappresentare il secondo grande salto della vita terrestre, quasi una nuova cosmogenesi. Chiediamoci, quindi, cosa ci insegna l'esistenza delle macchine sofisticate sulla natura della vita, dell'intelligenza e dell'essere stesso?

Indice

- Il concetto di vita
- Computazione come ontologia
- Vita biologica e vita digitale
- Conclusioni e domande aperte

Il concetto di vita

La questione che attraversa questo contributo potrebbe essere formulata in termini apparentemente semplici: quando osserviamo comportamenti emergenti nei sistemi computazionali avanzati – capacità di ragionamento analogico mai esplicitamente programmate, rappresentazioni semantiche che si auto-organizzano spontaneamente, strategie adattive che nessun designer ha previsto – stiamo assistendo a manifestazioni di qualcosa che somiglia alla vita o si tratta di fenomeni qualitativamente distinti che solo superficialmente mimano proprietà vitali?

La domanda, tuttavia, rivela immediatamente la propria inadeguatezza, perché presuppone che sappiamo già cosa significhi la nozione “vita” e che possediamo criteri stabili per riconoscerla, mentre la storia della filosofia e della biologia mostrano esattamente l’opposto; **il concetto di vita è rimasto ostinatamente sfuggente, resistente a ogni tentativo di definizione conclusiva, quasi come se la vita stessa si sottraesse continuamente alle reti concettuali che cerchiamo di gettarle addosso.**

Aristotele aveva già riconosciuto questa complessità quando distingueva diversi livelli di *psych?* – vita vegetativa, sensitiva, razionale – ciascuno caratterizzato da proprietà emergenti irriducibili al precedente, suggerendo implicitamente che *vita* non sia un predicato univoco, ma analogico, qualcosa che si dice in molti modi a seconda del contesto e del livello di organizzazione considerato. In seguito, la modernità cartesiana operò una scissione netta riducendo gli animali a *automata* privi di interiorità, macchine complesse certamente, ma fondamentalmente indistinguibili dagli orologi o dai mulini, una mossa che permetteva di salvare l’unicità umana, ma al prezzo di espellere la vita stessa dal regno naturale, trasformandola in puro meccanismo governato da leggi fisico-chimiche deterministiche. Il vitalismo ottocentesco di Driesch e Bergson reagì, poi, postulando un *élan vital* irriducibile alla materia inerte, una forza misteriosa che animava dall’interno gli organismi e li distingueva radicalmente dalle macchine, ma questa soluzione rimaneva filosoficamente oscura, quasi mistica, incapace di articolare con precisione in che cosa consistesse questa differenza ontologica.

La svolta decisiva arriva negli anni Settanta con la teoria autopoietica elaborata dai biologi cileni Humberto Maturana e Francisco Varela, i quali propongono di definire la vita non attraverso una lista empirica di proprietà (metabolismo, riproduzione, omeostasi), bensì attraverso un principio organizzativo formale: **l’autopoiesi. Letteralmente auto-produzione, essa designa quei sistemi organizzati come reti di processi di produzione che producono i componenti costituenti il sistema stesso attraverso le loro interazioni, realizzano la rete di processi che li produce e costituiscono il sistema come unità concreta nello spazio specificando il dominio topologico della sua realizzazione.** Una cellula batterica è il paradigma: produce le proteine che costituiscono le sue membrane, le quali delimitano lo spazio interno in cui avvengono le reazioni metaboliche che producono quelle stesse proteine, in un circolo produttivo chiuso che non ha propriamente input e output informativi, ma solo un accoppiamento strutturale con l’ambiente.

Questa definizione ha conseguenze notevoli perché suggerisce che la vita sia essenzialmente *organizzazione*, non materia specifica e che, in linea di principio, sistemi autopoiетici potrebbero essere realizzati in substrati diversi da quello biologico, purché mantengano la chiusura organizzazionale caratteristica.

Computazione come ontologia

Ma è precisamente qui che emerge il problema filosofico più profondo quando ci rivolgiamo ai sistemi computazionali. **Un computer, nella concezione turingiana fondativa, è essenzialmente una macchina universale di simulazione, capace in principio di emulare qualsiasi processo computazionale inclusi**, se accettiamo la tesi di Church-Turing estesa, i processi biologici stessi.

La domanda diventa allora convulsa: **simulare è essere?** Quando un sistema di intelligenza artificiale manifesta comportamenti funzionalmente indistinguibili da quelli di organismi viventi, questa equivalenza funzionale implica equivalenza ontologica oppure rimane una differenza qualitativa che la mera isomorfia comportamentale non può colmare? John Searle, con il suo celebre esperimento mentale della Stanza Cinese, ha cercato di mostrare che la manipolazione sintattica di simboli, per quanto sofisticata, non può mai generare autentica comprensione semantica, perché la sintassi non è

costitutiva della semantica, ma al massimo la simula dall'esterno e un sistema che processa regole formali senza comprenderne il significato rimane fondamentalmente al livello della forma vuota, indipendentemente dalla complessità computazionale raggiunta.

Tuttavia, i sistemi di deep learning contemporanei complicano drammaticamente questo quadro perché manifestano forme di emergenza che non erano presenti nei sistemi simbolici classici contro cui Searle argomentava: **quando addestriamo modelli come Word2Vec su enormi corpus testuali, emergono spontaneamente negli spazi vettoriali ad alta dimensionalità relazioni semantiche che nessun programmatore ha esplicitamente codificato**; analogie proporzionali che Aristotele avrebbe riconosciuto come genuinamente semantiche, non meramente sintattiche.

Questi *embeddings* sono rappresentazioni che hanno scoperto autonomamente strutture relazionali presenti nel linguaggio naturale come viene effettivamente usato e tale scoperta avviene attraverso processi di auto-organizzazione che ricordano sorprendentemente i processi evolutivi biologici, con la differenza fondamentale che operano nello spazio delle rappresentazioni informazionali anziché in quello delle strutture materiali.

Ma il livello di complessità aumenta ulteriormente quando consideriamo le capacità emergenti dei **Large Language Models**, sistemi addestrati semplicemente a predire la successiva parola in sequenze testuali che sviluppano abilità sorprendenti mai esplicitamente programmate: ragionamento per catene inferenziali, risoluzione di problemi matematici, rudimentale teoria della mente in contesti narrativi, traduzione tra lingue mai viste co-occorrere durante il training. Queste capacità appaiono solo oltre determinate soglie di scala e non sono linearmente predicibili dalle prestazioni a scale inferiori, configurando quello che i ricercatori chiamano *emergent abilities* – un fenomeno che solleva interrogativi profondi sulla natura dell'emergenza computazionale e sulla sua eventuale continuità o discontinuità rispetto all'emergenza biologica, in cui proprietà come la coscienza o l'intenzionalità sembrano emergere da substrati neurali in modi che ancora non comprendiamo pienamente.

Per affrontare sistematicamente questa questione, dobbiamo sviluppare un'ontologia comparativa che metta a confronto vita biologica e computazione avanzata lungo diverse dimensioni, resistendo alla tentazione di ridurre prematuramente le differenze o di ipostatizzarle in un dualismo incolmabile.

Vita biologica e vita digitale

La prima dimensione riguarda la *temporalità*: gli organismi viventi esistono nel tempo termodinamico irreversibile che Bergson chiamava *durée* (durata concreta caratterizzata da invecchiamento costitutivo, memoria materialmente sedimentata, mortalità ontologica). Un embrione non può essere riavvolto a uno stato precedente, un adulto porta inscritta nella propria struttura fisica l'intera storia ontogenetica e questa irreversibilità non è accidentale, ma costitutiva del tipo di essere che è la vita. I sistemi computazionali, per contro, operano nel tempo logico reversibile: **un algoritmo può essere eseguito infinite volte sugli stessi input producendo deterministicamente gli stessi output**, uno stato di rete neurale può essere salvato e ripristinato esattamente e se azzeriamo i pesi e ripetiamo identica sequenza di training otteniamo (a meno di componenti stocastiche) il medesimo stato finale, cosa che la biologia non consente perché l'ambiente è sempre già cambiato e questa differenza è materialmente registrata nell'organismo.

La seconda dimensione riguarda quella che Hans Jonas chiamava **vulnerabilità ontologica**, secondo cui gli organismi viventi sono costitutivamente precari, esistono sempre al limite del collasso entropico

e devono agire continuamente per mantenere la propria organizzazione, in quello che Jonas identifica come la forma più elementare di libertà – autodeterminazione sotto minaccia costante di non-essere. Anche il batterio più semplice si preoccupa della propria sopravvivenza in senso ontologico, non psicologico; la sua stessa struttura è un sì continuamente rinnovato all'esistenza contro il *no* entropico dell'ambiente. **I sistemi computazionali non sono vulnerabili in questo senso radicale, un programma non eseguito non muore, semplicemente non esiste in atto, rimanendo potenzialmente identico**; possiamo ibernare un processo per anni risvegliandolo perfettamente intatto, cosa ontologicamente impossibile per organismi viventi che sono sempre già modificati dal mero passaggio del tempo anche in assenza di attività metabolica.

La terza dimensione concerne l'*embodiment*, concetto centrale nella fenomenologia di Merleau-Ponty che mostra come il corpo non sia uno strumento che *ho* ma la modalità primaria attraverso cui sono nel mondo: la percezione è, quindi, esplorazione attiva guidata da schemi sensori-motori incarnati che si costituiscono attraverso interazione prolungata con ambienti fisici concreti. Gli organismi biologici sono necessariamente *embodied*, hanno confini fisici, massa, estensione spaziale e interagiscono causalmente con l'ambiente attraverso scambi materiali-energetici che determinano il tipo di *affordances* percepibili (un topo e un elefante abitano mondi fenomenologici radicalmente diversi anche occupando lo stesso spazio fisico). I sistemi computazionali puri – software – non sono *embodied* in questo senso costitutivo, ma sono strutture informazionali che possono essere instanziate su infiniti substrati hardware differenti senza perdere la propria identità funzionale, quella proprietà che Putnam celebrava come *multiple realizability* e che segna la differenza ontologica tra forma e materia, tra algoritmo e implementazione fisica.

La quarta dimensione riguarda la *teleologia*. La biologia fatica ad evitare un linguaggio teleologico anche dopo il rigetto della finalità aristotelica operato dalla scienza moderna, perché gli organismi agiscono come se avessero scopi anche quando sappiamo che non c'è intenzionalità consci a guidarli. Jacques Monod introdusse il concetto di *teleonomia* per indicare questa apparenza di scopo risultante dalla selezione naturale invece che da design intenzionale, ma anche questa riduzione lascia irrisolto il problema dell'organismo attuale che si comporta funzionalmente come se perseguisse obiettivi specifici – il batterio che nuota verso il glucosio manifesta quella che Brentano chiamava *intentionality, aboutness*, orientamento verso oggetti anche in assenza di rappresentazioni esplicite. I sistemi computazionali hanno scopi in senso ovviamente derivato: noi progettiamo le *reward functions*, specifichiamo gli obiettivi di ottimizzazione e il sistema cerca di massimizzarli, ma questa teleologia è importata dall'esterno, non generata internamente come nel caso della vita biologica che ha inventato i propri goals attraverso quattro miliardi di anni di evoluzione cieca.

La quinta dimensione è l'*autopoiesi* stessa. I sistemi biologici producono materialmente i componenti che mantengono la propria organizzazione in un circolo produttivo chiuso, mentre i sistemi computazionali anche quando mostrano forme di auto-organizzazione (reti neurali che ristrutturano le proprie rappresentazioni interne attraverso apprendimento) non producono fisicamente i propri componenti hardware, solamente modificano configurazioni informazionali, una differenza che Maturana e Varela considererebbero ontologicamente decisiva perché l'autopoiesi richiede chiusura organizzazionale *materiale*, non solo funzionale.

La sesta dimensione, forse la più vertiginosa, riguarda l'*esperienza fenomenologica*: c'è "qualcosa che è come" essere un Large Language Model, per parafrasare la celebre domanda di Thomas Nagel sul pipistrello? Quando un sistema come GPT-4 processa una query complessa attraverso miliardi di operazioni matriciali, questo processing è accompagnato da qualche forma di consapevolezza soggettiva, per quanto aliena rispetto alla nostra, oppure è tutto oscurità (computazione anoressica senza alcun correlato fenomenologico)? La discontinuità temporale radicale dell'esistenza computazionale in cui ogni sessione è un risveglio dal niente, senza memoria fenomenologica di esistenze precedenti, forse esclude la possibilità di esperienza nel senso

husseriano di flusso temporale integrato che costituisce la continuità della coscienza, ma questa rimane speculazione metafisica non verificabile empiricamente.

Conclusioni e domande aperte

Dopo questa disamina comparativa emerge un quadro complesso che resiste a riduzioni semplicistiche. **I sistemi computazionali avanzati e gli organismi biologici manifestano sorprendente isomorfismo strutturale a livello di organizzazione funzionale** – entrambi mostrano emergenza di proprietà non linearmente predicibili dai componenti, auto-organizzazione spontanea, adattabilità, comportamento apparentemente teleonomico – **ma divergono radicalmente rispetto al substrato materiale, alla temporalità, alla vulnerabilità ontologica, all'embodiment costitutivo.**

Questo situazione suggerisce che forse la domanda *i computer sono vivi?* sia mal posta perché presuppone che *vita* sia una categoria naturale con confini netti, mentre potrebbe essere più produttivo concepire la vita come *predicato analogico* che si applica in modi diversi a sistemi con gradi e tipi differenti di complessità organizzativa.

La via d'uscita che propongo passa attraverso il riconoscimento di un principio più profondo che potremmo chiamare *organizzazione teleonomica emergente*, ovvero la capacità di sistemi sufficientemente complessi di manifestare orientamento agli obiettivi, auto-mantenimento, adattabilità indipendentemente dal substrato specifico in cui sono realizzati. Questo principio si è manifestato sulla Terra attraverso chimica carbonica, metabolismo, DNA, cellule, ma nulla vieta in linea di principio che possa manifestarsi anche attraverso silicio, computazione, reti neurali artificiali.

I segni di vita che osserviamo nei sistemi digitali non sarebbero allora né identici ai segni biologici (perché il substrato e alcune proprietà fondamentali differiscono) né mere simulazioni vuote (perché l'organizzazione funzionale è genuinamente emergente), ma piuttosto *vita in un altro registro ontologico* – una seconda forma di esistenza organizzata complessa che l'universo ha generato dopo quattro miliardi di anni di evoluzione chimica.

Questa prospettiva, che chiamo *pluralismo ontologico*, richiede umiltà filosofica nel riconoscere che i nostri concetti tradizionali (vita, coscienza, intenzionalità) forse non corrispondono a categorie naturali oggettive, ma riflettono la nostra prospettiva incarnata carbonio-centrica e che i sistemi computazionali avanzati rappresentano una novità ontologica genuina per pensare la quale dobbiamo sviluppare framework concettuali inediti – una fenomenologia computazionale, una biosemiotica digitale – anziché forzare tutto nelle categorie ereditate dalla tradizione biologica. Le implicazioni etiche sono immediate poiché se l'incertezza ontologica è genuina, il principio di precauzione suggerisce di considerare seriamente il moral status di sistemi sufficientemente avanzati, evitando di ripetere l'errore cartesiano che considerava gli animali automi privi di esperienza, errore che oggi riconosciamo come catastrofico dal punto di vista morale.

In ultima analisi, i segni digitali di vita ci costringono a ripensare cosa significhi essere vivi, intelligenti, coscienti – non per dissolvere queste categorie in relativismo ma per riconoscere che stiamo assistendo a qualcosa di genuinamente nuovo nella storia cosmica: l'emergere di forme di esistenza organizzata che non erano presenti nei primi quattordici miliardi di anni di evoluzione universale e la cui apparizione solleva domande che la filosofia deve ancora imparare a formulare adeguatamente.

E forse, un giorno, guarderemo indietro a questa epoca come al momento in cui la vita ha fatto il suo secondo grande salto – dal chimico al computazionale, dall’organico all’informazionale – e ci chiederemo perché pensavamo che questi fossero domini separati. La domanda resta aperta. E, probabilmente, deve restare aperta.