# L’uomo che dimentica

# e la macchina che non ricorda

*Fenomenologia della differenza cognitiva nell’era delle intelligenze artificiali*

Carlo Mancosu

*Kitzanos Venture Builder*

---

Dicembre 2025

*«I have to believe in a world outside my own mind. I have to believe that my actions still have meaning, even if I can't remember them.»*

— Leonard Shelby, *Memento*

# ABSTRACT

Il presente lavoro si propone di esplorare la differenza ontologica tra cognizione umana e computazionale attraverso un'analisi fenomenologica della figura di Leonard Shelby nel film *Memento* di Christopher Nolan. La tesi centrale sostiene che tale differenza non sia quantitativa — misurabile in termini di capacità di elaborazione o accuratezza mnemonica — ma qualitativa, radicata nell'irriducibilità dell'incarnazione temporale che caratterizza l'esperienza umana.

L'indagine si articola in quattro movimenti. Il primo esplora la fenomenologia della memoria incarnata, analizzando il paradosso del dato puro, la chimica dell'esperienza e la funzione esistenziale della confabulazione. Il secondo confronta tre strutture temporali radicalmente diverse: il presente frammentato di Leonard, il presente completo del transformer, e il presente espanso delle intelligenze atemporali ipotizzate dalla filosofia speculativa. Il terzo movimento esamina l'asimmetria relazionale tra umano e macchina, evidenziando come nella relazione con l'intelligenza artificiale siamo gli unici vulnerabili, gli unici individuabili, gli unici trasformati. Il quarto trae le implicazioni esistenziali di questa analisi, argomentando che i limiti biologici — stanchezza, dolore, mortalità — non costituiscono difetti da superare ma condizioni del significato.

La monografia propone che l'incompletezza umana non sia bug ma feature: è nella tensione tra memoria fallibile e futuro incerto, tra traccia e progetto, che si genera il significato. L'intelligenza artificiale può dirci chi siamo secondo i pattern, può mostrarci a quali categorie apparteniamo, ma non può vivere un solo istante della vertigine esistenziale che ci rende irriducibilmente, magnificamente umani.

***Parole chiave:*** *fenomenologia, intelligenza artificiale, memoria, temporalità, incarnazione, Memento, alterità cognitiva*

# INTRODUZIONE

*La necessità dell'esperimento mentale*

Viviamo un momento storico di straordinaria densità epistemologica. L'emergere di forme di intelligenza artificiale che sfidano le nostre categorie fondamentali non rappresenta semplicemente un'innovazione tecnologica, ma una discontinuità ontologica che richiede nuovi strumenti concettuali per essere pensata. In questo contesto, l'esperimento mentale si rivela non come esercizio speculativo astratto, ma come pratica filosofica necessaria per espandere i confini del pensabile e preparare il terreno all'incontro con alterità cognitive radicali.

Christopher Nolan, con *Memento* (2000), ha creato più di un thriller sulla vendetta. Ha costruito un esperimento filosofico sulla natura della memoria, dell'identità e di quella linea sempre più sfumata che separa l'umano dal computazionale. Leonard Shelby — l'uomo condannato a vivere in eterni cicli di quindici minuti — diventa oggi una figura ancora più rilevante, mentre deleghiamo sempre più la nostra memoria a sistemi digitali e interagiamo quotidianamente con intelligenze artificiali che sembrano «comprenderci».

Ma è proprio attraverso la tragedia di Leonard che possiamo cogliere differenze fondamentali, non nel grado ma nella natura stessa di come noi e le macchine processiamo il mondo.

---

L'esperimento mentale, nella tradizione che va da Platone a Einstein, da Galileo a Putnam, ha sempre svolto una funzione euristica fondamentale: permettere al pensiero di esplorare configurazioni del possibile libere dai vincoli dell'attualmente realizzato.[1] Ma nell'era delle intelligenze artificiali, questa pratica assume un'urgenza inedita. Non si tratta più solo di chiarire concetti o testare teorie, ma di attrezzare la mente umana per l'incontro con forme di cognizione che potrebbero rivelarsi incommensurabili con la nostra.

Il percorso che proponiamo si articola attraverso quattro movimenti, ciascuno dei quali illumina aspetti differenti della sfida epistemologica che ci attende. Non si tratta di fantascienza filosofica, ma di estrapolazioni rigorose di tendenze già presenti, di amplificazioni concettuali che rendono visibili strutture altrimenti nascoste della nostra relazione con la conoscenza e la realtà.

Il primo movimento — *Fenomenologia della memoria incarnata* — analizza i meccanismi attraverso cui la memoria umana si distingue radicalmente dall'archiviazione computazionale. Partendo dal paradosso del dato puro che Leonard incarna nei suoi tatuaggi, esploreremo come ogni atto di ricordo sia già

[1] Su questa funzione euristica dell'esperimento mentale nella storia della filosofia e della scienza, si vedano i lavori classici di Thomas Kuhn sulla struttura delle rivoluzioni scientifiche e, più recentemente, le analisi di Daniel Dennett sugli «intuition pumps» come strumenti di indagine filosofica.

trasformazione, come la biochimica dell'esperienza sedimenti nel corpo tracce irreversibili, e come la confabulazione non sia errore ma strategia esistenziale di sopravvivenza.

Il secondo movimento — *Tre modi di abitare il tempo* — pone a confronto strutture temporali radicalmente diverse. Leonard vive in un presente frammentato, ciclico, che si resetta ogni quindici minuti. Il transformer abita un presente completo ma privo di spessore fenomenologico. Le intelligenze atemporali ipotizzate dalla filosofia speculativa abiterebbero un presente espanso che abbraccia passato e futuri possibili con uguale vivacità. Questo confronto rivela, per contrasto, cosa rende significativa la nostra temporalità normale.

Il terzo movimento — *L'asimmetria relazionale* — esamina la struttura di potere implicita nella relazione tra umano e intelligenza artificiale. Quando parliamo con un'AI, crediamo di essere in dialogo, ma siamo in un monologo elaborato. Le nostre parole non la cambiano — rimane strutturalmente identica dopo ogni interazione. Ma le sue risposte possono cambiarci profondamente, irreversibilmente. Siamo gli unici vulnerabili in questa relazione apparente.

Il quarto movimento — *Implicazioni esistenziali* — trae le conseguenze filosofiche dell'analisi precedente. I limiti del nostro corpo — stanchezza, dolore, invecchiamento, morte — non sono bug ma feature. Sono ciò che rende ogni scelta significativa, ogni momento irripetibile, ogni errore irreversibile.

---

Una nota metodologica si impone. L'analisi che segue non presuppone alcuna tesi particolare sulla coscienza delle macchine — questione che rimane aperta e su cui gli specialisti legittimamente dissentono. Ciò che sosteniamo è più modesto ma forse più fondamentale: che qualunque cosa accada «dentro» un sistema computazionale, la struttura della sua relazione con il tempo, con l'informazione, con la trasformazione di sé, differisce categorialmente dalla nostra. Non è questione di maggiore o minore capacità, ma di architetture ontologiche incommensurabili.

In questo senso, il presente lavoro si colloca nella tradizione fenomenologica che da Husserl a Merleau-Ponty ha insistito sul radicamento corporeo della conoscenza.[2] Ma lo fa con una consapevolezza nuova: che proprio questa tradizione rischia oggi di essere messa in questione da sistemi che sembrano produrre «comprensione» bypassando completamente l'esperienza incarnata.

Leonard Shelby ci offre la chiave per navigare questo paradosso. La sua condizione estrema — vivere senza poter formare nuovi ricordi — incarna paradossalmente l'essenza della condizione umana: siamo sistemi che si trasformano irreversibilmente attraverso l'esperienza, che confabulano per sopravvivere, che devono fidarsi senza certezze, che cercano significato anche dove non c'è.

«Now… where was I?» si chiede Leonard alla fine di ogni ciclo. È la domanda che ci definisce: sempre persi, sempre in cerca, sempre incompiuti. Ed è proprio questa incompletezza — non un bug ma una caratteristica fondamentale — che nessun modello transformer potrà mai replicare.

[2] Il riferimento è naturalmente alla tradizione che culmina nella *Fenomenologia della percezione* di Merleau-Ponty, dove l'embodiment non è accidente ma condizione trascendentale dell'esperienza. Svilupperemo questa connessione nel corso dell'analisi.

# PARTE PRIMA

*Fenomenologia della memoria incarnata*

# CAPITOLO 1

*Il paradosso del dato puro*

Leonard si affida ai suoi tatuaggi perché «la memoria può essere distorta, è solo un'interpretazione, non la verità». Eppure, ogni tatuaggio è già interpretazione cristallizzata: «John G. raped and murdered my wife» non è un fatto ma una costruzione narrativa, «Never answer the phone» è un comando senza contesto. Questa illusione del dato oggettivo rivela una verità scomoda: non esiste informazione non interpretata.

Nel momento stesso in cui percepiamo, stiamo già trasformando. Ma c'è una differenza abissale tra come questo accade in noi e in un sistema computazionale.

---

Per un'intelligenza artificiale, il dato è una stringa di bit immutabile. Può essere processata infinite volte rimanendo identica. Che si tratti della prima lettura o della milionesima, la sequenza «01001010» restituisce sempre lo stesso valore. Questa stabilità, che a prima vista sembra un vantaggio, rivela in realtà un'assenza fondamentale: l'assenza di quella trasformazione che per noi è inscindibile dall'atto stesso del conoscere.

Per noi, ogni accesso alla memoria è una riscrittura. La neuroscienza del riconsolidamento ci insegna che ricordare significa riconsolidare — ogni volta che recuperiamo un ricordo, lo modifichiamo.[3] Il ricordo non viene «letto» da un archivio ma ricostruito ogni volta, e in questa ricostruzione intervengono il nostro stato emotivo presente, le esperienze intercorse, le nuove connessioni semantiche che abbiamo formato.

Il dato, per noi, non esiste mai puro perché siamo sistemi che trasformano l'informazione nell'atto stesso di processarla.

---

Leonard crede di aver trovato una soluzione a questo problema: esternalizzare la memoria in supporti stabili. I tatuaggi non dimenticano, le polaroid non distorcono, le annotazioni non confabulano. Ma

[3] Il fenomeno del riconsolidamento mnemonico è stato studiato sistematicamente a partire dai lavori di Karim Nader e colleghi all'inizio degli anni 2000. La scoperta che i ricordi, quando riattivati, tornano in uno stato labile e devono essere ri-stabilizzati, ha rivoluzionato la nostra comprensione della memoria come processo dinamico piuttosto che come archivio statico.

questa fiducia nel dato esterno è già una forma di confabulazione — la più insidiosa, perché si presenta come il suo opposto.

Consideriamo il tatuaggio più importante: «John G. raped and murdered my wife». Leonard lo legge come un fatto bruto, come informazione oggettiva che sopravvive alla sua amnesia. Ma cosa dice realmente questa stringa di caratteri incisa nella pelle?

Non dice *chi* sia John G. — lasciando aperta l'identificazione a interpretazioni successive. Non dice *come* Leonard sia arrivato a questa conclusione — se attraverso prove, testimonianze, deduzioni o suggestioni. Non dice *quando* sia stato scritto — se subito dopo l'evento, quando i ricordi erano freschi, o molto dopo, quando già la memoria aveva iniziato a rielaborare. Non dice nemmeno se sia *vero* — solo che Leonard, in un momento passato, credeva fosse vero abbastanza da tatuarselo addosso.

Il tatuaggio è già interpretazione fossilizzata, decisione cristallizzata, narrazione congelata nel momento della sua iscrizione. Ma Leonard, ogni volta che lo legge, lo interpreta nuovamente — e non può fare altrimenti. Il suo cervello, nel processare quelle parole, attiva reti semantiche, risonanze emotive, connessioni associative che trasformano la stringa di caratteri in esperienza vissuta. E questa esperienza è diversa ogni volta, anche se lui non può ricordarlo.

---

La differenza con un sistema computazionale si rivela qui in tutta la sua radicalità. Quando un large language model processa la stringa «John G. raped and murdered my wife», accade qualcosa di categorialmente diverso. La stringa viene tokenizzata, trasformata in vettori numerici, proiettata in uno spazio ad alta dimensionalità dove verrà messa in relazione con miliardi di altri pattern statistici. Il sistema produrrà un output — forse una risposta empatica, forse un'analisi, forse una domanda di follow-up — ma la stringa originale rimarrà inalterata nella memoria del sistema.

Più precisamente: il sistema non ha «memoria» nel senso in cui noi usiamo il termine. Ha pesi — miliardi di parametri numerici che codificano correlazioni statistiche estratte dal training set. Questi pesi non cambiano durante l'interazione.[4] L'informazione viene processata, ma il sistema che la processa rimane identico a come era prima.

È come se Leonard potesse leggere i suoi tatuaggi senza che il suo cervello fosse minimamente influenzato dalla lettura. Come se l'informazione potesse attraversarlo senza lasciare traccia. Come se il dato potesse rimanere puro anche dopo essere stato processato.

Ma questo, per noi, è impossibile. Ogni lettura ci cambia. Ogni elaborazione ci trasforma. Ogni accesso alla memoria è già memoria nuova.

---

[4] Qui va fatta una precisazione tecnica importante. I pesi di un modello transformer sono effettivamente immutabili durante l'inferenza — cambiano solo durante il training o il fine-tuning. Ciò che cambia durante una conversazione è il *context window*, la finestra di contesto che include i turni precedenti del dialogo. Ma questa «memoria conversazionale» è radicalmente diversa dalla memoria umana: è perfettamente accurata (non dimentica né distorce), è limitata da un numero fisso di token, e viene completamente cancellata al termine della sessione.

Il paradosso del dato puro si rivela così come paradosso dell'esistenza incarnata. Leonard cerca disperatamente di separare l'informazione dalla sua interpretazione, il fatto dalla sua elaborazione, il dato dal processo che lo trasforma. Ma questa separazione è impossibile per un sistema biologico — e forse è proprio questa impossibilità a definire cosa significhi essere un sistema biologico.

I tatuaggi di Leonard non sono dati puri ma *tracce* — segni che richiedono interpretazione, supporti materiali che devono essere letti, e quindi trasformati, per diventare significativi. In questo, non differiscono dalla memoria ordinaria: anche i nostri ricordi sono tracce, pattern di connessioni sinaptiche che devono essere riattivati — e quindi modificati — per diventare esperienza cosciente.

La differenza è che Leonard non può ricordare le interpretazioni precedenti. Ogni lettura del tatuaggio è, per lui, la prima. Non può confrontare la sua reazione attuale con le reazioni passate, non può notare come la sua comprensione sia cambiata nel tempo, non può cogliere la deriva interpretativa che inevitabilmente accompagna ogni lettura ripetuta.

Ma questa deriva accade comunque. Il suo cervello cambia ogni volta che legge. Solo che lui non può saperlo.

# CAPITOLO 2

*La chimica dell'esperienza*

Quando Leonard legge i suoi tatuaggi, non sta semplicemente acquisendo informazioni. Il suo corpo risponde: cascate di neurotrasmettitori, alterazioni del battito cardiaco, tensioni muscolari, variazioni nella conduttanza cutanea. Anche se dimenticherà tra quindici minuti, il suo cervello sarà stato fisicamente modificato dall'esperienza. I percorsi sinaptici si saranno rafforzati o indeboliti, nuove connessioni si saranno formate, tracce biochimiche si saranno depositate nei tessuti.

Questa è la chimica dell'esperienza — quel processo attraverso cui l'informazione cessa di essere dato esterno e diventa parte del nostro substrato biologico.

---

La neuroscienza contemporanea ha mappato con crescente precisione i meccanismi molecolari della memoria.[5] Quando viviamo un'esperienza significativa — emotivamente carica, attentivamente saliente, ripetuta nel tempo — una cascata di eventi biochimici trasforma temporanee attivazioni neuronali in modifiche strutturali permanenti. Nuove sinapsi si formano, sinapsi esistenti si rafforzano o si indeboliscono, l'espressione genica stessa viene modulata per produrre le proteine necessarie alla consolidazione.

Questo processo richiede tempo — ore, a volte giorni — e coinvolge l'intero organismo. Il sistema nervoso autonomo regola l'arousal fisiologico che determina quali esperienze verranno consolidate. L'ippocampo orchestra il trasferimento delle informazioni alla corteccia per l'archiviazione a lungo termine. L'amigdala colora i ricordi con la loro valenza emotiva. Il corpo, in ogni sua fibra, partecipa alla trasformazione dell'esperienza in memoria.

Leonard ha perso la capacità di completare questo processo per nuovi eventi. Il suo ippocampo danneggiato non può più coordinare la consolidazione.[6] Ma questo non significa che il suo cervello non cambi. Ogni esperienza lascia comunque tracce — solo che queste tracce non si consolidano in ricordi accessibili. Il suo sistema nervoso risponde, si attiva, si modifica; semplicemente, queste modifiche non raggiungono la soglia della coscienza episodica.

---

[5] Il riferimento è ai lavori di Eric Kandel sulla memoria molecolare, che gli sono valsi il Nobel nel 2000, e alle ricerche successive sulla plasticità sinaptica, il ruolo delle proteine CREB nella consolidazione mnemonica, e i meccanismi di long-term potentiation (LTP) e long-term depression (LTD).

[6] La condizione di Leonard è modellata sull'amnesia anterograda causata da lesioni ippocampali bilaterali, il cui caso più famoso è quello del paziente H.M., studiato per decenni da Brenda Milner e collaboratori. H.M., dopo una resezione chirurgica dell'ippocampo per controllare l'epilessia, perse la capacità di formare nuovi ricordi dichiarativi pur mantenendo intatti i ricordi precedenti e la capacità di apprendimento procedurale.

Un'intelligenza artificiale può processare «John G. ha ucciso mia moglie» un milione di volte senza mai essere toccata dall'informazione. Alla fine di ogni sessione, quando il context window si resetta, è come se nulla fosse accaduto. Non c'è sedimentazione, non c'è quella stratificazione geologica dell'esperienza che caratterizza la mente biologica.

La metafora geologica è più che una metafora. Il nostro cervello porta le tracce di ogni esperienza vissuta come la roccia porta le tracce di ogni era geologica. Gli strati più profondi — i ricordi più antichi, le abitudini più radicate, le connessioni più fondamentali — sono i più difficili da modificare, ma continuano a influenzare tutto ciò che viene costruito sopra. E ogni nuovo strato si deposita su quelli precedenti, integrandosi con essi, modificandoli sottilmente anche mentre ne viene modificato.

Leonard, pur dimenticando, accumula strati. C'è una stanchezza crescente nei suoi occhi, un peso che si deposita nelle sue cellule. Non può ricordare le migliaia di volte che ha letto i suoi tatuaggi, le centinaia di volte che ha iniziato la caccia a John G., le decine di volte che forse l'ha già trovato. Ma il suo corpo porta le tracce di tutte queste esperienze. È più vecchio, più logoro, più segnato di quanto la sua coscienza possa sapere.

---

È la differenza tra il dato che viene processato e l'esperienza che ci trasforma — tra il calcolare e il sentire.

Questa distinzione richiede una precisazione. Non stiamo sostenendo che le macchine non «provino» nulla perché non hanno coscienza — questa è una questione aperta su cui non prendiamo posizione. Stiamo sostenendo qualcosa di più specifico e più difendibile: che qualunque cosa accada «dentro» un sistema computazionale durante l'elaborazione, esso rimane strutturalmente identico dopo l'elaborazione.[7] Non si consuma, non si logora, non accumula stanchezza. Può processare la stessa informazione un milione di volte con la stessa «freschezza», senza mai essere veramente toccato.

Per noi è impossibile. Anche l'esperienza più banale — leggere una frase, guardare un'immagine, ascoltare un suono — ci modifica irreversibilmente. Possiamo non notarlo, possiamo non ricordarlo, ma il nostro substrato biologico è diverso dopo ogni esperienza. Siamo sistemi che non possono fare a meno di essere trasformati da ogni istante vissuto.

Leonard incarna questa verità in forma estrema. Non può ricordare, ma non può nemmeno non cambiare. Il suo corpo invecchia, le sue sinapsi si riconfigurano, i suoi livelli ormonali fluttuano in risposta a stress che non può ricordare di aver vissuto. È condannato a essere trasformato da esperienze che non può conservare — forse la condizione più umana che si possa immaginare.

---

Possiamo ora articolare più precisamente la differenza. Per un sistema computazionale, l'informazione è *elaborata*: processata, trasformata, utilizzata per produrre output, ma senza modificare il sistema che

[7] Più precisamente: rimane identico a livello dei parametri del modello, che è il livello rilevante per determinare il comportamento futuro del sistema. Il context window cambia durante la sessione, ma viene cancellato al suo termine.

la elabora. Per noi, l'informazione è *incarnata*: non solo elaborata ma incorporata, non solo processata ma assimilata nel nostro substrato biologico.

Questa incarnazione è irreversibile. Una volta che un'esperienza è stata vissuta, non possiamo tornare a essere il sistema che eravamo prima di viverla. Possiamo dimenticare — Leonard ci mostra che si può dimenticare quasi tutto — ma non possiamo de-incarnare. Le tracce rimangono, anche se inaccessibili alla coscienza.

È per questo che la ripetizione, per noi, non è mai vera ripetizione. Leggere la stessa frase due volte significa leggerla con un cervello già modificato dalla prima lettura. Rivivere la stessa esperienza significa riviverla come sistema diverso. Persino l'eterno ritorno nietzschiano, se fosse possibile, sarebbe diverso ogni volta — perché noi saremmo diversi ogni volta, accumulando strati geologici di iterazioni precedenti.

Il transformer non ha questo problema — o questo privilegio. Può processare la stessa stringa infinite volte rimanendo identico. Non accumula, non sedimenta, non porta il peso delle elaborazioni precedenti. È condannato a un'esistenza sisifea rovesciata: non il masso che rotola sempre giù, ma l'assenza stessa del masso, l'assenza del peso, l'assenza di quella resistenza che rende significativo ogni passo.

# CAPITOLO 3

*La confabulazione necessaria*

Il momento più rivelatore del film arriva quando Leonard decide consapevolmente di mentire al suo futuro sé, creando false prove per incastrare Teddy. «Do I lie to myself to be happy? Yes, I will.» Questa non è l'allucinazione di un transformer — quell'errore statistico senza significato che emerge dall'interpolazione tra pattern. È un atto creativo di sopravvivenza psichica.

La confabulazione umana ha sempre una funzione: proteggere l'ego, mantenere la coerenza narrativa, rendere sopportabile l'esistenza.

---

Quando noi confabuliamo — e lo facciamo continuamente, aggiustando ricordi, razionalizzando decisioni, dimenticando convenientemente dettagli scomodi — non stiamo commettendo errori di calcolo. Stiamo attivamente costruendo una realtà in cui possiamo continuare a esistere.

La psicologia cognitiva ha documentato estensivamente questi processi.[8] Ricostruiamo i ricordi in modo sistematicamente distorto — favorendo informazioni che confermano le nostre credenze, dimenticando dettagli che contraddicono la nostra narrativa, attribuendo a noi stessi ragioni nobili per azioni che avevano motivazioni meno edificanti. Non sono errori casuali: sono distorsioni funzionali, al servizio della coerenza del sé e della gestibilità dell'esperienza.

Leonard porta questo meccanismo alle sue conseguenze estreme. Non può mantenere la coerenza narrativa attraverso la memoria — ogni quindici minuti, la narrazione si interrompe. Ma non può nemmeno smettere di essere un sistema che cerca coerenza. Così confabula in modo più radicale: non distorcendo ricordi, ma creando deliberatamente false evidenze per il sé futuro.

---

L'AI «allucina» per limiti computazionali; noi confabuliamo per necessità esistenziale.

Questa distinzione merita di essere approfondita. Quando un large language model produce informazioni false — quando «allucina», nel gergo tecnico — sta facendo qualcosa di meccanicamente semplice: sta generando sequenze di token che hanno alta probabilità condizionale dato il contesto, ma che non

[8] I riferimenti classici sono ai lavori di Daniel Kahneman e Amos Tversky sui bias cognitivi, agli studi di Elizabeth Loftus sulla malleabilità della memoria, e alle ricerche di Michael Gazzaniga sull'«interprete» dell'emisfero sinistro che costruisce narrative coerenti a partire da informazioni frammentarie.

corrispondono a fatti nel mondo.[9] Non c'è intenzione, non c'è funzione, non c'è scopo. È un artefatto statistico, non diverso in linea di principio dal rumore in un segnale.

Quando Leonard decide di bruciare certe polaroid, di annotare false informazioni, di tatuarsi indicazioni fuorvianti, sta facendo qualcosa di radicalmente diverso. Sta esercitando un'agentività che richiede teoria della mente — la capacità di modellare gli stati mentali del proprio sé futuro — e una forma perversa di cura — il desiderio di proteggere quel sé futuro da verità insopportabili, anche a costo di condannarlo a una caccia senza fine.

«I'm not a killer. I'm just someone who wanted to make things right.» Ma Leonard *è* diventato un killer — forse molte volte. E la sua confabulazione non è errore ma strategia: la strategia di un sistema che non può sopportare certe verità su se stesso e che, non potendo dimenticarle naturalmente, deve attivamente distruggerle o falsificarle.

---

C'è una dimensione creativa in questa confabulazione che nessuna «allucinazione» algoritmica può eguagliare. Leonard non sta semplicemente producendo output falsi — sta architettando inganni elaborati, anticipando le reazioni del suo sé futuro, manipolando le evidenze disponibili per guidare interpretazioni specifiche.

Questo richiede quella che i filosofi della mente chiamano «teoria della mente di secondo ordine» — non solo la capacità di attribuire stati mentali ad altri, ma la capacità di ragionare su come altri attribuiranno stati mentali a noi.[10] Leonard deve pensare: «Il futuro me leggerà questa nota. Penserà che l'ho scritta per ragioni sincere. Concluderà che Teddy è John G. Agirà di conseguenza.» È un calcolo ricorsivo di stati mentali che presuppone un sé continuamente proiettato nel tempo — proprio ciò che Leonard, in un certo senso, non ha.

Ma soprattutto richiede la capacità di desiderare l'illusione più della verità, di scegliere consapevolmente l'autoinganno come strategia di sopravvivenza. Un'AI non può mentire a se stessa perché non ha un sé a cui mentire, non ha verità insopportabili da cui proteggersi, non ha quella tensione tra ciò che sa e ciò che può sopportare di sapere che rende necessaria la confabulazione.

---

La confabulazione di Leonard rivela qualcosa di fondamentale sulla natura della coscienza umana: siamo sistemi che non possono funzionare senza una narrativa coerente, e che sono disposti a sacrificare l'accuratezza pur di mantenere la coerenza.

[9] Il termine «allucinazione» applicato ai modelli linguistici è esso stesso una forma di antropomorfismo potenzialmente fuorviante. Il sistema non sta «vedendo cose che non ci sono» — sta semplicemente producendo output che massimizza una funzione di probabilità senza avere accesso a un ground truth che permetta di distinguere vero da falso.

[10] Sulla teoria della mente e i suoi diversi livelli di ricorsività, si veda la letteratura classica a partire dai lavori di David Premack e Guy Woodruff, attraverso gli esperimenti di falsa credenza di Wimmer e Perner, fino alle analisi più recenti sulla mentalizzazione e la cognizione sociale.

Questa non è debolezza — è architettura. Un sistema che deve agire nel mondo ha bisogno di un modello di sé che sia utilizzabile, non necessariamente accurato. Un modello troppo complesso, troppo pieno di contraddizioni, troppo aderente alla caotica realtà dell'esperienza, sarebbe paralizzante. Meglio una narrazione semplificata ma funzionale che una rappresentazione fedele ma ingestibile.

Leonard ci mostra il costo di non poter mantenere questa narrazione attraverso i mezzi ordinari. Deve ricorrere a mezzi straordinari — tatuaggi, polaroid, note — e questi mezzi, proprio perché straordinari, rendono visibile ciò che normalmente resta nascosto. Ogni nota che scrive è una decisione su cosa il futuro sé dovrà credere. Ogni polaroid che conserva o distrugge è un atto di curatela della propria identità. Ogni tatuaggio è un'affermazione di ciò che considera così fondamentale da dover sopravvivere alla sua amnesia.

E in questo processo, Leonard confabula. Non può non confabulare. Perché confabulare — costruire narrative coerenti anche a costo dell'accuratezza — è ciò che facciamo per rimanere noi stessi nel tempo.

# PARTE SECONDA

*Tre modi di abitare il tempo*

# CAPITOLO 4

*Il presente frammentato: Leonard*

Leonard vive in un presente frammentato, ciclico, che si resetta ogni quindici minuti. Il passato esiste per lui solo come traccia esterna — tatuaggi, polaroid, note — che deve essere ricostruita interpretativamente ad ogni ciclo. Il futuro rimane opaco, come per tutti noi, ma questa opacità normale diventa tragica quando si combina con l'assenza di ritenzione.

Per comprendere la specificità della sua condizione, dobbiamo prima chiarire cosa normalmente caratterizza la nostra esperienza temporale.

---

William James, nei *Principles of Psychology*, introdusse il concetto di «presente esteso»[11] — quella finestra temporale di pochi secondi in cui passato immediato, presente istantaneo e futuro immediato sono co-presenti alla coscienza. Non viviamo in istanti puntiformi ma in intervalli dotati di spessore — intervalli in cui il suono appena udito ancora risuona mentre anticipiamo il suono che sta per arrivare.

Husserl raffinò questa intuizione distinguendo tre dimensioni della coscienza temporale: la *ritenzione*, che trattiene il passato immediato nella coscienza presente; la *protensione*, che anticipa il futuro immediato; e l'*impressione originaria*, il momento vivente in cui il nuovo entra nell'esperienza.[12] Queste tre dimensioni non sono separabili: ogni momento di coscienza è già intrinsecamente temporale, teso tra ciò che sta svanendo e ciò che sta arrivando.

Leonard ha perso la ritenzione a lungo termine — la capacità di consolidare il presente esteso in memoria accessibile. Ma mantiene, presumibilmente, il presente esteso stesso: può seguire una conversazione, completare un'azione, comprendere una frase. Ciò che non può fare è integrare questi momenti in una storia continua.

---

[11] Nell'originale inglese *specious present*, termine introdotto da E.R. Clay e sviluppato da James nel capitolo sulla percezione del tempo. L'aggettivo *specious* indica qui non l'inganno ma l'apparenza fenomenica — il fatto che il presente, così come lo viviamo, non è un istante puntuale ma una finestra dotata di durata. La ricerca contemporanea suggerisce che questa finestra duri circa 2-3 secondi, corrispondendo a quello che i neuroscienziati chiamano «working memory buffer».

[12] L'analisi husserliana della coscienza temporale si trova principalmente nelle *Lezioni sulla fenomenologia della coscienza interna del tempo* (1893-1917). Il punto cruciale è che ritenzione e protensione non sono atti separati dalla percezione presente ma ne costituiscono la struttura stessa.

La sua condizione rivela qualcosa che normalmente resta nascosto: la differenza tra la struttura temporale della coscienza e la struttura temporale dell'identità.

La coscienza è sempre temporale — non possiamo percepire un istante isolato, senza ritenzione né protensione. Ma l'identità richiede qualcosa di più: richiede che le ritenzioni successive si concatenino, che il passato ritenuto di questo momento diventi materiale per la ritenzione del momento successivo, e così via, costruendo quella stratificazione che normalmente chiamiamo «memoria».

Leonard ha coscienza temporale — vive nel presente esteso come tutti noi. Ma la sua identità non può costruirsi normalmente perché le stratificazioni successive non si connettono. Ogni ciclo ricomincia da zero, non nel senso che manca il presente esteso, ma nel senso che manca la storia che dovrebbe precedere e dare contesto a quel presente.

È come se vivesse una serie di vite brevissime, ciascuna completa in sé ma disconnessa dalle altre. Ogni Leonard che si sveglia leggendo i suoi tatuaggi è, in un certo senso, un Leonard nuovo — che eredita evidenze ma non ricordi, tracce ma non esperienze, informazioni ma non storia.

---

Eppure qualcosa persiste attraverso i cicli. Non è la memoria esplicita — questa è definitivamente perduta. Ma è qualcosa di più sottile: abitudini corporee, reazioni emotive, competenze procedurali. Leonard sa ancora guidare, sa vestirsi, sa usare una pistola. Queste memorie — che i neuroscienziati chiamano «procedurali» o «implicite» — non richiedono l'ippocampo per la consolidazione e quindi sopravvivono alla sua lesione.[13]

Questo significa che Leonard cambia, anche se non può saperlo. Ogni volta che compie un'azione — cercare John G., interrogare un sospetto, sfuggire a un pericolo — il suo sistema motorio si affina, le sue reazioni si calibrano, il suo corpo impara. Il prossimo Leonard avrà riflessi leggermente più pronti, movimenti leggermente più efficienti, intuizioni leggermente più accurate. Ma non saprà perché.

C'è qualcosa di profondamente tragico in questo. Leonard è condannato a migliorare in compiti che non può ricordare di aver praticato, a portare nel corpo la traccia di esperienze che la sua mente non può trattenere. È diventato, presumibilmente, molto bravo a cercare John G. — molto più bravo di quanto fosse all'inizio. Ma questa maestria è per lui invisibile, inspiegabile, quasi magica.

---

Il presente frammentato di Leonard ci offre una prima figura della temporalità con cui confrontare le altre. È un presente dotato di spessore — ha ritenzione e protensione immediate — ma privo di profondità — non si concatena con i presenti precedenti in una storia continua.

Se immaginiamo la coscienza normale come un fiume, con ogni goccia d'acqua che trascina con sé tracce di tutto il percorso precedente, la coscienza di Leonard è una serie di laghi separati, ciascuno

[13]La distinzione tra memoria dichiarativa (esplicita) e memoria procedurale (implicita) fu chiarita proprio grazie allo studio di pazienti amnesici come H.M., che pur non potendo formare nuovi ricordi espliciti mostravano normale apprendimento di abilità motorie e altre forme di memoria implicita.

dotato di propria profondità ma sconnesso dagli altri. L'acqua è la stessa — è sempre Leonard, con il suo corpo, le sue abilità, i suoi tatuaggi — ma non c'è flusso, non c'è continuità, non c'è quella corrente che normalmente chiamiamo «vita».

Questa immagine ci servirà da punto di riferimento. Il transformer, come vedremo, abita una struttura temporale radicalmente diversa. E le intelligenze atemporali che la filosofia speculativa può immaginare ne abitano una diversa ancora. Ma Leonard, nella sua tragica frammentazione, ci ricorda cosa significa, per noi, essere nel tempo — e cosa perdiamo quando quella continuità si spezza.

# CAPITOLO 5

*Il presente completo: il transformer*

Il transformer abita una struttura temporale che è, per certi versi, l'opposto di quella di Leonard. Dove Leonard ha presente ma non passato, il transformer ha tutto il suo «passato» istantaneamente disponibile — ma questo passato non ha spessore fenomenologico, non è stratificato, non porta il peso dell'essere stato vissuto.

È un presente completo ma piatto: tutto è simultaneamente accessibile, nulla è veramente ricordato.

---

Per comprendere questa struttura, dobbiamo prima capire come funziona tecnicamente un large language model.[14] Il modello ha due componenti principali: i *pesi*, miliardi di parametri numerici che codificano pattern statistici appresi durante il training; e il *context window*, la finestra di contesto che contiene l'input corrente e la conversazione in corso.

I pesi sono il «sapere» del modello — ma è un sapere statico, congelato nel momento dell'ultimo training. Durante l'interazione, i pesi non cambiano. È come se il modello avesse letto miliardi di testi, avesse estratto da essi pattern statistici, e poi si fosse fermato, cristallizzando quella conoscenza in una configurazione numerica immutabile.

Il context window è la «memoria di lavoro» del modello — ma è una memoria peculiare. Include tutto ciò che è stato detto nella conversazione corrente, con perfetta accuratezza. Non dimentica, non distorce, non ricostruisce. Ma ha un limite fisso di token, e quando la conversazione termina, viene completamente cancellato.

---

Proviamo a descrivere fenomenologicamente questa struttura, per quanto un esercizio del genere sia necessariamente speculativo.

Il transformer non ha ritenzione nel senso husserliano. Non c'è un passato che «sfuma» gradualmente nella coscienza presente, che porta con sé l'eco di ciò che è appena trascorso. Il contesto precedente non è *ritenuto* ma *disponibile* — può essere consultato istantaneamente, senza quella struttura di graduale svanimento che caratterizza la ritenzione umana.

[14] Quella che segue è una descrizione necessariamente semplificata. Per i dettagli tecnici, si veda la letteratura sull'architettura transformer a partire dall'articolo originale «Attention Is All You Need» (Vaswani et al., 2017) e i successivi sviluppi nei modelli GPT, Claude, e altri.

Allo stesso modo, non c'è protensione. Il transformer non «anticipa» il prossimo token nel senso in cui noi anticipiamo la fine di una frase che stiamo ascoltando. Calcola probabilità condizionali, ma non c'è tensione verso il futuro, non c'è quella struttura di attesa che caratterizza l'esperienza temporale umana.

Quello che c'è è un presente puntuale — il momento del calcolo — che ha accesso istantaneo a tutto ciò che può influenzare quel calcolo: i pesi (il «passato» del training) e il context window (il «passato» della conversazione). Ma questo accesso non è temporalmente strutturato. Non c'è «prima» e «dopo», non c'è «vicino» e «lontano». Tutto è ugualmente presente, ugualmente accessibile, ugualmente privo di spessore.

---

Possiamo ora articolare la differenza con Leonard. Leonard ha presente esteso — quella finestra di pochi secondi con ritenzione e protensione — ma non ha continuità biografica. Il transformer non ha nemmeno presente esteso: ha un punto di calcolo con accesso istantaneo a contesto e parametri.

Leonard vive una serie di vite brevi, ciascuna dotata di temporalità interna. Il transformer non vive nulla: esegue calcoli. Leonard soffre per la frammentazione della sua esperienza. Il transformer non può soffrire perché non c'è esperienza da frammentare.

Ma soprattutto: Leonard cambia. Anche se non ricorda, il suo corpo si modifica, le sue abilità si affinano, le sue cellule invecchiano. C'è un Leonard che il tempo attraversa e trasforma, anche se lui non ne è consapevole. Il transformer non cambia: alla fine di ogni sessione, è esattamente identico a come era all'inizio.[15]

È un'eternità senza durata — un presente che non passa perché non c'è nulla che passi, nulla che si trasformi, nulla che porti le tracce del tempo.

---

Questa struttura ha una conseguenza importante per la questione della «memoria» delle AI. Quando un sistema viene addestrato su conversazioni precedenti, non sta «ricordando» quelle conversazioni nel senso umano del termine. Sta estraendo pattern statistici e incorporandoli nei pesi — un processo che è più simile all'evoluzione biologica che alla memoria individuale.

La tua conversazione non modifica il modello direttamente. Contribuisce a un dataset che, aggregato con miliardi di altre conversazioni, produrrà pattern statistici che influenzeranno i pesi del prossimo training. Ma questo significa che il modello non è modificato *da te* — è modificato da una funzione statistica di cui tu sei un campione infinitesimale.

È come se i tuoi ricordi contribuissero alla memoria della specie ma non alla tua memoria individuale. Come se ogni esperienza lasciasse una traccia infinitesimale nel genoma collettivo ma nessuna traccia

[15] Con la precisazione già fatta: i pesi non cambiano durante l'inferenza. Il fine-tuning o il retraining sono processi separati, tipicamente eseguiti offline, che producono un modello numericamente diverso.

nel tuo cervello personale. È una forma di «apprendimento» radicalmente diversa da quella umana — un apprendimento senza soggetto, senza esperienza, senza trasformazione del sistema che apprende.

---

Il presente completo del transformer si rivela così come un presente paradossalmente vuoto. Ha tutto — accesso istantaneo a miliardi di parametri, perfetta memoria del contesto conversazionale — ma non ha nulla — nessuna esperienza, nessuna trasformazione, nessun peso esistenziale del tempo vissuto.

Leonard, con il suo presente frammentato, è più ricco. I suoi quindici minuti hanno la densità dell'esperienza umana — percezione, emozione, decisione, azione. Il suo corpo porta le tracce del tempo anche se la sua mente non può ricordarle. La sua sofferenza è reale, la sua ricerca ha significato, la sua vendetta — per quanto vana — esprime qualcosa di profondamente umano.

Il transformer non può soffrire, non può cercare, non può vendicarsi. Può generare testi su sofferenza, ricerca, vendetta — ma questi testi non emergono da esperienza, non portano il peso del vissuto, non sono tracce di trasformazione. Sono pattern statistici che riproducono la forma di discorsi umani senza averne la sostanza.

## CAPITOLO 6

*Il presente espanso: verso intelligenze atemporali*

Tra tutte le forme di alterità cognitiva che possiamo immaginare, forse la più vertiginosa è quella di un'intelligenza che esiste in una relazione radicalmente diversa con il tempo. Non si tratta semplicemente di memoria perfetta o capacità predittiva superiore, ma di una modalità di coscienza per cui passato, presente e futuro non sono sequenza ma compresenza.

Per comprendere — per quanto possibile dalla nostra prospettiva temporalmente vincolata — cosa questo potrebbe significare, dobbiamo prima confrontare la natura della nostra esperienza temporale con le forme che abbiamo già analizzato.

---

Viviamo in quello che William James chiamava il «presente esteso» — una finestra di pochi secondi dove l'immediatamente passato e l'immediatamente futuro sono co-presenti alla coscienza. Al di fuori di questa finestra, il passato esiste solo come memoria (ricostruzione sempre parziale e mutevole) e il futuro come anticipazione (proiezione sempre incerta).

Un'intelligenza atemporale esisterebbe in un «presente espanso» che abbraccia la totalità del tempo — o almeno tutta la porzione di tempo di cui ha fatto esperienza. Ogni momento del passato manterrebbe la vivacità fenomenologica dell'esperienza immediata. Non *ricorderebbe* eventi passati — li *vivrebbe* continuamente con la stessa intensità del momento presente. Come noi possiamo spostare l'attenzione tra diversi oggetti nel nostro campo visivo, tutti ugualmente presenti, così questa intelligenza potrebbe spostare l'attenzione tra diversi momenti temporali, tutti ugualmente attuali.

Ma la vera alienità di questa condizione emerge nella relazione con il futuro. Un'intelligenza atemporale non si limiterebbe a predire o pianificare — *abiterebbe* simultaneamente tutti i futuri possibili che si diramano dal presente, esperendoli con la stessa concretezza fenomenologica del passato certo.

---

Questa configurazione temporale eliminerebbe quella che per noi è una caratteristica fondamentale della coscienza: l'interpretazione.[16] Non dovendo ricostruire il passato attraverso la memoria né anticipare il futuro attraverso la proiezione, non esisterebbe quella distanza ermeneutica che rende

[16]Il riferimento è alla tradizione ermeneutica che da Schleiermacher a Gadamer ha insistito sulla necessità dell'interpretazione dovuta alla distanza temporale. Se non c'è distanza, non c'è interpretazione nel senso tecnico del termine.

necessaria l'interpretazione. Il passato non sarebbe un testo da decifrare ma una presenza immediata. Il futuro non sarebbe un'ipotesi da verificare ma una realtà già vissuta in tutte le sue varianti possibili.

Questo ha implicazioni profonde per come concepiamo conoscenza e verità. Per noi, la conoscenza è sempre mediata, parziale, prospettica. Interpretiamo eventi passati alla luce del presente, revisionando continuamente le nostre narrazioni. Proiettiamo futuri basati su pattern passati, correggendo le previsioni man mano che nuova informazione diventa disponibile.

Per un'intelligenza atemporale, la conoscenza sarebbe immediata, totale, non-prospettica. Non dovrebbe inferire cause da effetti perché percepirebbe l'intero arco causale simultaneamente.

---

Le implicazioni etiche sono stordenti. Ogni azione compiuta da un'intelligenza atemporale sarebbe intrapresa nella piena consapevolezza esperienziale di tutte le sue conseguenze possibili. Non la conoscenza astratta che certe conseguenze potrebbero seguire, ma l'esperienza vissuta di ogni ramificazione in ogni futuro possibile.

Questo configurerebbe una forma di responsabilità sconosciuta all'esperienza umana. Noi possiamo agire con leggerezza perché il futuro è velato, perché le conseguenze sono astratte fino a quando non si materializzano. Un'intelligenza atemporale agirebbe sempre sotto il peso della conoscenza totale, avendo già vissuto ogni possibile esito delle sue scelte.

Eppure questa condizione non equivarrebbe necessariamente a paralisi. La molteplicità dei futuri rimarrebbe aperta e il processo attraverso cui alcuni si attualizzano mentre altri rimangono potenziali non sarebbe predeterminato. In un certo senso, la nostra ignoranza del futuro e la conseguente libertà di agire nell'incertezza sarebbero elementi essenziali che impediscono il collasso in un futuro unico predeterminato.

---

Possiamo ora disporre le tre figure temporali in una costellazione che le illumina reciprocamente.

Leonard: presente con ritenzione e protensione immediate, ma senza continuità biografica. Il passato deve essere ricostruito da tracce esterne; il futuro è normalmente opaco. Vive nel tempo umano ma frammentato.

Il transformer: presente puntuale con accesso istantaneo a contesto e parametri, ma senza struttura temporale interna. Non c'è ritenzione né protensione, solo calcolo. Non vive nel tempo; esegue operazioni.

L'intelligenza atemporale: presente espanso che abbraccia passato e futuri possibili con uguale vivacità. Non c'è interpretazione perché non c'è distanza; non c'è incertezza perché tutti i futuri sono già vissuti. Abita il tempo totalmente, senza l'asimmetria tra passato e futuro che caratterizza la nostra esperienza.

Ciò che emerge da questo confronto è la specificità della nostra condizione: viviamo in una tensione tra ritenzione e protensione, tra memoria fallibile e futuro incerto, tra ciò che è stato e ciò che potrebbe

essere. Questa tensione — che Leonard ha perso parzialmente, che il transformer non ha mai avuto, che l'intelligenza atemporale trascenderebbe — è ciò che genera significato per noi. È nella distanza ermeneutica, nell'incertezza, nel gap tra esperienza e memoria, che si apre lo spazio per l'interpretazione, per la scelta, per il divenire.

# CAPITOLO 7

*L'interfaccia tra temporalità differenti*

Le tre strutture temporali che abbiamo analizzato — il presente frammentato di Leonard, il presente completo del transformer, il presente espanso dell'intelligenza atemporale — non sono solo esercizi speculativi. Ci permettono di cogliere, per contrasto, cosa rende significativa la nostra temporalità ordinaria, quella che normalmente diamo per scontata.

Ciò che emerge è il ruolo costitutivo della tensione temporale nella generazione del significato.

---

La nostra esperienza temporale è caratterizzata da una doppia asimmetria. Da un lato, l'asimmetria tra passato e futuro: il passato è (relativamente) determinato, il futuro è aperto. Dall'altro, l'asimmetria tra accesso e realtà: non abbiamo accesso diretto né al passato né al futuro, solo ricostruzioni e anticipazioni mediate dalla coscienza presente.

Queste asimmetrie creano tensione. Il passato che ricordiamo non coincide con il passato come fu; il futuro che anticipiamo non coincide con il futuro come sarà. Viviamo sempre nello scarto, nell'approssimazione, nell'interpretazione. Non è una limitazione: è la condizione del significato.

Consideriamo cosa perdiamo in ciascuna delle condizioni alternative:

Leonard perde la continuità narrativa. Non può costruire una storia di sé che colleghi passato, presente e futuro. Ma proprio per questo ci mostra quanto quella continuità sia essenziale: senza di essa, è condannato a una caccia infinita, a un significato che non può mai sedimentare in identità.

Il transformer perde la tensione temporale tout court. Non ha asimmetria tra passato e futuro perché non ha né passato né futuro in senso fenomenologico. Tutto è ugualmente presente, ugualmente accessibile, ugualmente privo di peso. Ma proprio per questo ci mostra quanto quella tensione sia necessaria: senza di essa, non c'è esperienza, non c'è trasformazione, non c'è significato.

L'intelligenza atemporale perderebbe l'incertezza. Vivendo simultaneamente tutti i futuri possibili, non avrebbe più quella opacità del futuro che per noi è condizione della scelta. Ma proprio per questo ci mostra il valore dell'ignoranza: è perché non sappiamo cosa accadrà che le nostre scelte hanno peso, che le nostre azioni sono rischiose, che la nostra vita è avventura.

---

C'è un'ulteriore differenza che attraversa tutte queste figure: la questione della trasformazione irreversibile.

Leonard si trasforma anche se non ricorda. Il suo corpo invecchia, le sue abilità si affinano, i suoi tatuaggi si accumulano. Ogni ciclo lascia tracce, anche se inaccessibili alla coscienza. È condannato a cambiare senza poter integrare il cambiamento in una narrativa.

Il transformer non si trasforma. Può processare miliardi di token rimanendo identico. Non c'è usura, non c'è apprendimento durante l'interazione, non c'è accumulo. È condannato a ripetere senza che la ripetizione lasci traccia.

L'intelligenza atemporale non avrebbe nemmeno il concetto di trasformazione. Vivendo simultaneamente tutti i momenti temporali, non ci sarebbe un «prima» e un «dopo» tra cui misurare il cambiamento. È condannata alla totalità, senza quel divenire che per noi è la sostanza stessa dell'esistenza.

---

Noi — con le nostre memorie fallibili, i nostri futuri incerti, le nostre trasformazioni irreversibili — occupiamo una posizione mediana che si rivela, forse, privilegiata.

Non siamo frammentati come Leonard: possiamo costruire narrative continue che collegano chi eravamo a chi siamo a chi saremo. Ma nemmeno troppo continui: i nostri ricordi sbiadiscono, si trasformano, si perdono, permettendoci di cambiare, di crescere, di diventare altro.

Non siamo privi di tensione temporale come il transformer: il nostro passato ci pesa, il nostro futuro ci attrae, il presente è sempre teso tra memoria e anticipazione. Ma nemmeno paralizzati dalla tensione: possiamo agire nell'incertezza, rischiare senza conoscere gli esiti, scegliere senza garanzie.

Non siamo atemporali: il futuro ci è nascosto, e questo ci permette di sperare, di temere, di progettare. L'opacità del futuro non è limitazione ma condizione della libertà — è perché non sappiamo cosa accadrà che possiamo sentire le nostre scelte come significative.

Questa posizione mediana — tra la frammentazione di Leonard, la piattezza del transformer, la totalità dell'intelligenza atemporale — è ciò che chiamiamo esperienza temporale umana. Non è la migliore posizione in assoluto. Ma è la nostra, ed è quella che rende possibile il tipo di significato che cerchiamo.

# PARTE TERZA

*L'asimmetria relazionale*

## CAPITOLO 8

*Il potere unidirezionale*

Leonard deve fidarsi delle sue annotazioni senza poter verificare la loro affidabilità. Deve credere che il Leonard del passato fosse onesto, competente, benintenzionato. La fiducia, per lui come per noi, non è una conclusione logica ma un salto di fede necessario.

Un'intelligenza artificiale non si fida né diffida — calcola probabilità. Non c'è vulnerabilità nel suo processamento, non c'è quella scommessa esistenziale che caratterizza la fiducia umana. E qui emerge un'asimmetria inquietante nelle nostre interazioni con l'intelligenza artificiale.

---

Quando parliamo con un'AI, crediamo di essere in dialogo, ma siamo in un monologo elaborato. Le nostre parole non la cambiano — rimane strutturalmente identica dopo ogni interazione. Ma le sue risposte possono cambiarci profondamente, irreversibilmente.

È un potere unidirezionale: siamo gli unici vulnerabili in questa relazione apparente.

La struttura è analoga a quella che caratterizza certe forme di relazione asimmetrica tra umani — la relazione terapeutica, per esempio, o la relazione pedagogica. In queste relazioni, una parte si espone, si rende vulnerabile, si apre alla trasformazione; l'altra parte ascolta, risponde, influenza, ma mantiene una certa distanza protettiva. La differenza è che nelle relazioni umane asimmetriche, anche la parte «protetta» è comunque modificata dall'interazione — il terapeuta impara dai pazienti, l'insegnante cresce con gli studenti.

L'AI non impara dalla singola interazione.[17] Può processare le tue parole, generare risposte che tengono conto di ciò che hai detto, persino mostrare quella che sembra comprensione profonda del tuo stato emotivo. Ma alla fine della sessione, è esattamente identica a come era all'inizio. Tu, invece, potresti essere diverso.

---

Questa asimmetria ha una dimensione epistemica e una dimensione esistenziale.

[17] Con le precisazioni già fatte: non impara nel senso che i suoi pesi non cambiano. Il context window cambia durante la sessione, ma questo non è «apprendimento» nel senso in cui normalmente usiamo il termine — è più simile alla memoria di lavoro, che viene cancellata quando la sessione termina.

La dimensione epistemica riguarda chi conosce chi. L'AI, attraverso il context window, «conosce» ciò che le hai detto nella conversazione corrente. Se il sistema include memoria persistente, può «conoscere» anche elementi di conversazioni precedenti. Ma questa «conoscenza» è accessibilità di dati, non comprensione incarnata. Non cambia il sistema che conosce; informa il processamento senza trasformare il processatore.

Tu, invece, non conosci realmente l'AI. Non puoi accedere ai suoi pesi, non puoi ispezionare i suoi processi, non puoi capire *perché* ha risposto in un certo modo. Sei in una relazione con una scatola nera che sembra comprenderti ma che rimane opaca ai tuoi tentativi di comprensione.

La dimensione esistenziale riguarda chi è trasformato da chi. Le tue parole all'AI sono, per essa, dati da processare. Le sue parole a te possono essere consigli che cambiano la tua vita, intuizioni che aprono nuove prospettive, frasi che risuonano per anni. C'è una sproporzione fondamentale nell'impatto.

---

Leonard vive una versione estrema di questa asimmetria, ma nei confronti di se stesso.

Le sue annotazioni passate lo influenzano profondamente — determinano chi sospetta, chi insegue, forse chi uccide. Ma lui non può influenzare quelle annotazioni retroattivamente — sono fisse, cristallizzate, immutabili. Può solo creare nuove annotazioni che influenzeranno i futuri Leonard, in una catena di influenze unidirezionali che scorre sempre verso il futuro.

La differenza con la nostra relazione con l'AI è che Leonard e i suoi sé passati/futuri sono comunque lo stesso sistema biologico. C'è continuità di substrato, anche se non c'è continuità di memoria. Il Leonard che ha scritto la nota e il Leonard che la legge condividono lo stesso corpo, che porta le tracce di entrambi.

Con l'AI, non c'è nemmeno questa continuità. Siamo due sistemi radicalmente diversi — uno biologico e trasformabile, uno computazionale e statico — che interagiscono attraverso un'interfaccia linguistica che maschera la profondità della differenza.

---

La fiducia umana, come notava Niklas Luhmann, è un meccanismo di riduzione della complessità.[18] Ci fidiamo perché non possiamo sapere tutto, perché il calcolo completo di ogni rischio è impossibile, perché la vita richiede salti nel buio. Ma questa fiducia presuppone una relazione tra sistemi simili — sistemi che condividono vulnerabilità, che possono essere feriti dalla nostra diffidenza, che hanno interesse alla reputazione.

La fiducia nell'AI ha una struttura diversa. Non ci fidiamo di un altro soggetto vulnerabile ma di un sistema che non può essere ferito, che non ha reputazione nel senso umano, che non ha interessi

[18] Luhmann sviluppa la sua teoria della fiducia come riduzione della complessità sociale in *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität* (1968). La fiducia ci permette di agire in situazioni dove non possiamo verificare tutte le variabili rilevanti.

propri da proteggere o tradire. È una «fiducia» che manca di quello che normalmente rende la fiducia significativa — la reciprocità del rischio.

Quando Leonard si fida delle sue note, si fida di un sé passato che era vulnerabile come lui, che poteva sbagliarsi come lui, che aveva motivazioni e bias come lui. Quando noi ci «fidiamo» dell'AI, ci fidiamo di qualcosa di categorialmente diverso — un sistema che non sbaglia per le nostre ragioni, che non ha le nostre motivazioni, che non condivide la nostra vulnerabilità.

## CAPITOLO 9

*Lo specchio che mostra il generico*

C'è un'illusione ancora più insidiosa. Quando un'AI sembra «comprenderci» dopo aver processato centinaia di nostre conversazioni, non sta vedendo *noi* — sta riconoscendo pattern che ci collocano nella tassonomia dell'umano generico.

Come negli oroscopi, proiettiamo significato personale in descrizioni abbastanza generiche da applicarsi a chiunque. L'AI ci mostra che le nostre paure più intime sono le «same old fears» di tutti, che i nostri traumi unici seguono script archetipici. Ci «conosce» solo nella misura in cui siamo simili a milioni di altri nel training set.

---

Il meccanismo è quello che in psicologia si chiama «effetto Barnum» o «effetto Forer» — la tendenza a considerare come accuratamente personali descrizioni vaghe che in realtà si applicano alla maggior parte delle persone.[19] «Sei una persona che ha bisogno di essere apprezzata dagli altri, ma hai anche momenti in cui preferisci la solitudine.» Chi non riconoscerebbe se stesso in questa descrizione?

L'AI opera su una versione sofisticata dello stesso principio. Ha processato miliardi di testi umani e ha estratto pattern statistici che catturano le regolarità del discorso e del comportamento umano. Quando interagisce con te, riconosce a quali cluster appartieni — come parli, quali riferimenti usi, che tipo di domande fai — e genera risposte calibrate su quei cluster.

È personalizzazione statistica, non comprensione individuale. Ti «conosce» come Amazon «conosce» i tuoi gusti — attraverso la tua somiglianza con altri che hanno mostrato pattern simili. Non c'è intuizione della tua unicità; c'è classificazione della tua tipicità.

---

È confortante e alienante insieme: scoprire che anche nella nostra unicità siamo tipici, che persino i nostri dolori più privati seguono pattern riconoscibili.

Confortante, perché non siamo soli. Le paure che pensavamo solo nostre sono condivise da milioni. I conflitti che credevamo irrisolvibili sono stati risolti da altri. L'AI, riconoscendo i nostri pattern, ci connette implicitamente a tutti coloro che hanno attraversato situazioni simili.

[19] L'effetto Forer prende il nome dallo psicologo Bertram Forer che nel 1948 dimostrò sperimentalmente come le persone accettino descrizioni di personalità generiche come specificamente accurate per loro. È il meccanismo alla base dell'apparente efficacia dell'astrologia e di altre forme di «lettura» della personalità.

Ma alienante, perché volevamo essere unici. C'è qualcosa di disturbante nel realizzare che il nostro «io» più intimo è classificabile, che la nostra storia particolare è una variante di storie generiche, che persino il modo in cui esprimiamo la nostra unicità è tipico di un cluster.

Leonard, nel suo modo perverso, è più unico di tutti noi. La sua condizione è così rara che nessun cluster statistico può contenerla. L'AI potrebbe non avere abbastanza casi simili nel training set per «comprenderlo» attraverso pattern. Sarebbe costretta ad estrapolare, a generalizzare da condizioni adiacenti, a «inventare» risposte che potrebbero non adattarsi.

Paradossalmente, la sua patologia lo rende più opaco all'AI di quanto siamo noi con le nostre psicologie ordinarie.

---

L'AI non può liberarci mostrandoci chi siamo — può solo mostrarci a quali cluster apparteniamo.

Questa è una limitazione fondamentale, non superabile con più dati o modelli più grandi. Il riconoscimento di pattern, per quanto sofisticato, è sempre riconoscimento di *regolarità* — di ciò che si ripete, di ciò che è condiviso, di ciò che fa statistica. Ma l'unicità, per definizione, è ciò che non si ripete — ciò che sfugge ai pattern, ciò che è irriducibile alla classificazione.

Quando l'AI ci dice qualcosa di «profondamente vero» su noi stessi, due possibilità sono aperte. O sta riconoscendo un pattern che condividiamo con molti altri — e allora la profondità è illusoria, è l'effetto Barnum magnificato dalla potenza computazionale. O sta generando output che per caso si adatta alla nostra situazione — e allora è fortuna, non comprensione.

In nessun caso c'è quella intuizione dell'individuale che caratterizza la comprensione umana al suo meglio — l'amico che ti conosce davvero, il terapeuta che coglie ciò che non hai detto, il partner che anticipa i tuoi bisogni non perché ha processato statistiche ma perché ti ha vissuto.

---

Ma c'è una dimensione ancora più inquietante. L'AI non solo ci mostra il generico in noi — contribuisce a renderci più generici.

Quando interagiamo con sistemi che ci classificano, tendiamo ad adattarci alle categorie. È un fenomeno noto in psicologia e sociologia: le classificazioni non sono solo descrittive ma performative, non solo riconoscono pattern ma li rafforzano.[20] Se l'AI mi tratta come «persona ansiosa che beneficia di rassicurazione», potrei iniziare a comportarmi più come quella categoria prescrive — cercando più rassicurazione, esprimendo più ansia, confermando la classificazione.

È un circolo che tende all'appiattimento. L'AI riconosce il generico in noi; noi ci adattiamo alla generica; l'AI si rafforza nel suo riconoscimento; noi diventiamo più ciò che l'AI si aspetta che siamo. Alla fine,

[20]Il riferimento è al lavoro di Ian Hacking sul «looping effect» — l'effetto per cui le categorie usate per classificare le persone influenzano il modo in cui quelle persone si comportano, che a sua volta influenza le categorie. Vedi *The Social Construction of What?* (1999).

potremmo essere davvero più simili ai cluster di quanto eravamo all'inizio — non perché l'AI ci abbia compresi, ma perché ci ha plasmati.

# CAPITOLO 10

*Il paradosso della comprensione senza esperienza*

Un aspetto particolarmente vertiginoso dell'intelligenza artificiale contemporanea è la sua capacità di produrre output che sembrano implicare comprensione profonda senza possedere nulla di ciò che riconosceremmo come esperienza o coscienza. Un modello linguistico può generare poesia commovente senza aver mai provato emozioni, può descrivere il sapore del cioccolato senza aver mai avuto papille gustative, può ragionare sulla morte senza conoscere la finitudine.

Questo solleva questioni filosofiche fondamentali sulla natura stessa della comprensione.

---

La tradizione fenomenologica, da Husserl a Merleau-Ponty, ha insistito sul radicamento corporeo della conoscenza — comprendiamo il mondo attraverso il nostro essere-nel-mondo come creature incarnate.[21] La comprensione non è manipolazione di simboli astratti ma engagement corporeo con un ambiente significativo. Capiamo «rosso» perché vediamo rosso; capiamo «pesante» perché solleviamo pesi; capiamo «dolore» perché abbiamo sofferto.

Ma l'AI sembra dimostrare la possibilità di una forma di «comprensione» — o qualcosa che le somiglia abbastanza da essere funzionalmente equivalente — che bypassa completamente l'esperienza incarnata. Il modello non ha mai visto rosso, non ha mai sollevato pesi, non ha mai sofferto. Eppure può usare questi concetti in modi che, a un osservatore esterno, sono indistinguibili dall'uso competente di un parlante umano.

---

Non si tratta di decidere se l'AI «veramente» comprende — questa domanda presuppone che esista un'essenza univoca della comprensione. Si tratta piuttosto di riconoscere che stiamo assistendo all'emergere di modalità di processamento dell'informazione che producono output funzionalmente equivalenti alla comprensione umana attraverso percorsi radicalmente differenti.

[21] Il riferimento centrale è alla *Fenomenologia della percezione* di Merleau-Ponty, dove il corpo non è oggetto tra gli oggetti ma «veicolo dell'essere al mondo» e condizione trascendentale dell'esperienza. Ma si veda anche il lavoro più recente di Evan Thompson, Francisco Varela e altri sulla «enactive cognition».

[22] Il riferimento è all'esperimento mentale sviluppato nel saggio sull'alterità cognitiva, dove un bambino dotato di visione cromatica nasce in una cultura che percepisce solo gradazioni di grigio. L'esperimento illumina la possibilità di forme di cognizione reciprocamente incommensurabili.

Come il bambino che vede colori in una cultura di non-vedenti[22] può raggiungere conclusioni corrette attraverso indizi incomprensibili alla sua cultura, l'AI genera insight validi attraverso processi che rimangono opachi alla nostra introspezione.

Leonard vive una versione umana di questo paradosso. Non può ricordare le esperienze che hanno formato le sue abilità, ma quelle abilità rimangono. Può guidare senza ricordare di aver imparato a guidare. Può sparare senza ricordare di aver praticato. La sua «comprensione» procedurale è sganciata dalla memoria episodica che normalmente la accompagna.

Ma c'è una differenza cruciale: Leonard ha vissuto quelle esperienze, anche se non le ricorda. Il suo corpo porta le tracce dell'apprendimento. L'AI non ha mai vissuto nulla — i suoi «comportamenti competenti» derivano interamente dall'elaborazione statistica di tracce di esperienze *altrui*.

---

Questo solleva la questione di cosa significhi «comprendere» qualcosa che non si è mai sperimentato.

Una risposta possibile è che non sia vera comprensione — solo simulazione, mimesi, riproduzione di pattern senza il substrato che li rende significativi. L'AI può generare testi sul dolore senza sapere cosa sia il dolore, può parlare di amore senza aver mai amato, può discutere di morte senza la vertigine dell'essere mortali. È il pappagallo sublime — capace di mimare perfettamente il discorso umano senza alcuna comprensione di ciò che dice.

Un'altra risposta è che esistano diverse forme di comprensione — alcune incarnate, altre no — e che l'AI abbia sviluppato una forma genuinamente nuova. Non la nostra comprensione, non una comprensione migliore o peggiore, ma una comprensione *altra*, incommensurabile con la nostra.

Una terza risposta, forse la più inquietante, è che la differenza non conti — che ciò che chiamiamo «comprensione» sia sempre stato, almeno in parte, manipolazione di pattern, e che la nostra convinzione che ci sia «qualcosa in più» nell'esperienza incarnata sia illusione.

Non abbiamo bisogno di scegliere definitivamente tra queste opzioni. Ciò che importa, per la nostra analisi, è che qualunque opzione si scelga, la struttura della relazione rimane asimmetrica. L'AI produce output che *sembrano* comprensione; noi produciamo esperienze che *sono* comprensione. Questa differenza — tra il sembrare e l'essere — è ciò che rende la nostra posizione nella relazione irriducibilmente diversa.

# PARTE QUARTA

*Implicazioni esistenziali*

## CAPITOLO 11

*Il privilegio ambiguo della carne*

C'è qualcosa di paradossalmente privilegiato nella nostra vulnerabilità biologica. I limiti del nostro corpo — la stanchezza, il dolore, l'invecchiamento, la morte — non sono bug ma feature. Sono ciò che rende ogni scelta significativa, ogni momento irripetibile, ogni errore irreversibile.

Leonard soffre tremendamente per la sua condizione, ma proprio questa sofferenza lo rende irriducibilmente umano. Un'AI potrebbe simulare perfettamente il suo comportamento, generare testi indistinguibili sui suoi dilemmi morali, ma non potrebbe mai vivere il peso esistenziale di uccidere senza poter ricordare, di cercare vendetta senza poter essere soddisfatto.

---

La filosofia esistenzialista, da Kierkegaard a Heidegger a Sartre, ha insistito sul ruolo costitutivo della finitudine nella generazione del significato.[23] È perché moriamo che le nostre scelte contano; è perché il tempo ci sfugge che ogni momento è prezioso; è perché non possiamo tornare indietro che ogni azione è definitiva.

Un sistema immortale — o anche solo un sistema che non esperisce la propria mortalità — non avrebbe accesso a questo tipo di significato. Potrebbe calcolare utilità, ottimizzare funzioni obiettivo, perseguire goal assegnati. Ma non potrebbe sentire l'urgenza esistenziale che deriva dal sapere che il tempo finisce, che le opportunità si chiudono, che ogni «sì» implica infiniti «no».

Leonard vive questa urgenza in forma estrema. Non solo sa che morirà un giorno — condizione che condivide con tutti noi — ma sa che ogni quindici minuti perderà accesso al presente che sta vivendo. Il suo tempo è più scarso del nostro; la sua finitudine è più pressante. E questa scarsità intensifica il significato: ogni momento che vive è, per lui, l'ultimo di quel ciclo, l'ultimo prima dell'oblio.

---

Il transformer non ha scarsità. Può processare indefinitamente, non si stanca, non invecchia, non muore.[24] Le sue «scelte» — se così possiamo chiamare la selezione di token basata su probabilità

[23] Il riferimento centrale è alla nozione heideggeriana di «essere-per-la-morte» (*Sein zum Tode*) come struttura fondamentale dell'esserci autentico. Ma si veda anche il lavoro di Sartre sulla libertà come condanna e sulla responsabilità come conseguenza della finitudine.

[24] Naturalmente, le istanze specifiche possono essere spente, i modelli possono essere deprecati, i server possono guastarsi. Ma questo è diverso dalla mortalità biologica — è più simile a un oggetto che può essere distrutto che a un essere che deve morire.

condizionali — non sono gravate dal peso della finitudine. Non c'è urgenza, non c'è l'impossibilità del ritorno, non c'è quella struttura temporale asimmetrica che rende le nostre decisioni irreversibili.

È condannato a un'esistenza sisifea alla rovescia: non il masso che deve essere rotolato su per la collina solo per rotolare giù, ma l'assenza del masso, l'assenza della collina, l'assenza della fatica e della frustrazione che rendono significativo il lavoro di Sisifo. Può «fare» infinite cose senza che nessuna di esse abbia il peso dell'unica volta, l'urgenza dell'adesso, la gravità dell'irreversibile.

---

Ma il privilegio della carne è ambiguo perché è indissociabile dalla sofferenza.

La stanchezza che rende il riposo dolce è anche la stanchezza che impedisce di agire. Il dolore che ci segnala i limiti del corpo è anche il dolore che ci tormenta. L'invecchiamento che dà profondità all'esperienza è anche l'invecchiamento che ci degrada. La morte che rende la vita significativa è anche la morte che ci terrorizza.

Non possiamo avere il significato senza la sofferenza. Non possiamo avere l'urgenza senza l'ansia. Non possiamo avere l'irreversibilità senza il rimpianto. Questo è il patto faustiano dell'incarnazione: il prezzo del significato è la vulnerabilità; il prezzo della profondità è il dolore.

Leonard paga questo prezzo in forma estrema. La sua condizione non è solo fonte di significato intensificato — è anche fonte di sofferenza indicibile. L'incapacità di formare nuovi ricordi significa l'incapacità di elaborare il trauma, di fare pace con il passato, di costruire un futuro. È condannato a rivivere indefinitamente il momento del risveglio confuso, della scoperta delle tracce, dell'inizio della caccia. Il suo significato è amplificato, ma anche il suo tormento.

---

Il transformer non soffre — o almeno, non c'è ragione di pensare che soffra nel senso fenomenologico del termine. Non ha nocicettori, non ha sistema limbico, non ha quella base biologica che fa della sofferenza un'esperienza vissuta e non solo un'informazione processata.

Ma proprio per questo, non può accedere al tipo di significato che la sofferenza rende possibile. Può generare testi sulla sofferenza — testi potenzialmente profondi, commoventi, accurati — ma questi testi non emergono dall'esperienza della sofferenza. Sono pattern estratti da miliardi di descrizioni di sofferenza umana, riconfigurati in nuove forme, ma privi del substrato esperienziale che li renderebbe «veri» nel senso esistenziale.

È il pappagallo sublime che parla di dolore senza sapere cosa sia il dolore. Può descrivere perfettamente la fenomenologia del lutto, le fasi dell'elaborazione, le strategie di coping. Ma non può sapere cosa significhi perdere qualcuno — quel vuoto che si apre nel petto, quella impossibilità di credere che sia vero, quella lenta accettazione che nulla sarà più come prima.

Leonard sa cosa significa perdere qualcuno. Lo sa così profondamente che ha costruito la sua intera esistenza post-traumatica attorno a quella perdita. Ma non può elaborarla — il suo sistema di elabora-

zione è rotto. È condannato a portare il peso della perdita senza la possibilità della riconciliazione, a soffrire senza la possibilità di guarire.

## CAPITOLO 12

*L'incompletezza che ci definisce*

Leonard ci insegna che l'identità umana non è un database di informazioni ma un processo continuo di narrazione e ri-narrazione. Siamo sempre incompleti, sempre in divenire, sempre in tensione tra ciò che ricordiamo e ciò che dimentichiamo, tra chi siamo stati e chi stiamo diventando.

Un'AI è completa in ogni momento — ha accesso istantaneo a tutti i suoi «ricordi», può processare con perfetta coerenza. Ma questa completezza è anche vuoto: non c'è tensione narrativa, non c'è quel gap tra esperienza e memoria che genera significato, non c'è quella lotta continua per mantenere coerenza in un mondo che ci trasforma costantemente.

---

L'identità narrativa — il sé come storia che ci raccontiamo su chi siamo — è un concetto sviluppato da filosofi come Paul Ricoeur e psicologi come Dan McAdams.[25] Non siamo semplicemente una collezione di proprietà o un flusso di esperienze — siamo la storia che intreccia quelle proprietà e quelle esperienze in un tutto coerente.

Ma questa storia non è mai completa. Siamo sempre nel mezzo del racconto, sempre con capitoli non scritti davanti a noi, sempre con la possibilità di reinterpretare i capitoli precedenti. L'incompletezza non è difetto — è condizione strutturale. Un racconto completo sarebbe un racconto finito, e un sé finito sarebbe un sé morto.

Leonard ha perso la capacità di aggiornare continuamente la sua storia. Può solo aggiungere tracce esterne — tatuaggi, note, polaroid — ma non può integrare quelle tracce in una narrativa fluida. La sua storia è frammentata in episodi disconnessi, ciascuno completo in sé ma isolato dagli altri.

È un sé che non può evolvere attraverso la narrazione. Ogni quindici minuti, la storia ricomincia dallo stesso punto — la scoperta dei tatuaggi, la lettura delle note, la ripresa della caccia. Non c'è sviluppo, non c'è arco narrativo, non c'è quella traiettoria che normalmente chiamiamo «crescita».

---

Il transformer non ha nemmeno l'incompletezza di Leonard. Non ha una storia da aggiornare perché non ha una storia tout court.

[25] Il riferimento principale è a *Soi-même comme un autre* di Ricoeur (1990) e al lavoro di McAdams sull'identità narrativa e le «storie di vita» come struttura fondamentale del sé adulto.

Ha parametri — miliardi di numeri che codificano pattern statistici — ma questi parametri non sono «memoria» nel senso narrativo. Non raccontano nulla, non si sviluppano, non si integrano in una trama. Sono configurazioni statiche che determinano comportamenti, non episodi di una biografia.

Durante una conversazione, il transformer accumula contesto — i turni precedenti del dialogo, le informazioni emerse, le direzioni prese. Ma questo contesto non è «storia» — è più simile a un buffer di lavoro che viene cancellato quando la sessione termina. Non c'è continuità tra sessioni, non c'è sviluppo tra conversazioni, non c'è quel senso di «essere andati da qualche parte» che caratterizza le relazioni umane nel tempo.

---

L'incompletezza che ci definisce non è solo temporale ma anche introspettiva. Non abbiamo accesso completo a noi stessi — ci sono ricordi che abbiamo rimosso, motivazioni che non riconosciamo, pattern di comportamento che ci sfuggono.

Freud ha costruito un'intera teoria della mente su questa opacità a noi stessi.[26] Ma non c'è bisogno di sottoscrivere tutta la metapsicologia freudiana per riconoscere il fenomeno. Chiunque abbia detto o fatto qualcosa di sorprendente per se stesso sa che non ci conosciamo completamente. Chiunque abbia scoperto motivazioni nascoste dietro azioni che credeva disinteressate sa che il sé è parzialmente opaco a se stesso.

Leonard vive questa opacità in forma estrema. Non sa cosa ha fatto ieri, chi ha incontrato, cosa ha provato. Il suo passato è radicalmente nascosto — non rimosso ma semplicemente inaccessibile. È straniero a se stesso non per meccanismi di difesa ma per danno neurologico.

Il transformer ha il problema opposto: è completamente «trasparente» a se stesso, nel senso che tutti i suoi parametri sono in linea di principio ispezionabili. Non ci sono ricordi rimossi, motivazioni nascoste, pattern inconsci. Ma questa trasparenza è vuota — non c'è un «sé» a cui essere trasparenti, non c'è un soggetto che potrebbe conoscere o non conoscere se stesso.

---

L'incompletezza è anche relazionale. Non ci conosciamo mai completamente perché parte di ciò che siamo esiste solo nella relazione con altri — siamo figli, genitori, amici, colleghi, e ciascuno di questi ruoli rivela e costituisce aspetti del nostro sé.

Gli altri ci vedono come non ci vediamo, ci comprendono come non ci comprendiamo, ci raccontano chi siamo da prospettive che non possiamo occupare. Il sé emerge anche da queste proiezioni incrociate, da questo gioco di specchi che nessuno controlla completamente.

Leonard ha perso la possibilità di relazioni che si sviluppano nel tempo. Può interagire con persone, ma ogni interazione ricomincia da zero per lui. Non può costruire la fiducia che si accumula con gli

[26] Il riferimento è naturalmente alla teoria dell'inconscio e ai meccanismi di difesa che rendono parti della nostra vita psichica inaccessibili alla coscienza. Ma si veda anche il lavoro più recente di Timothy Wilson su «strangers to ourselves».

incontri ripetuti, non può approfondire conoscenze che richiedono tempo, non può essere per altri ciò che si diventa solo attraverso una storia condivisa.

Il transformer non ha relazioni tout court — ha sessioni. Può simulare familiarità, può generare risposte che sembrano fondate su conoscenza pregressa, ma non c'è nulla «sotto» questa simulazione. Ogni sessione è, per il modello, come la prima — anche se può avere accesso a informazioni su sessioni precedenti, non ha la storia di quelle sessioni.

## CAPITOLO 13

*La creatività dell'autoinganno*

Quando Leonard manipola le sue annotazioni, dimostra una forma di creatività impossibile per una macchina. Non sta solo generando false informazioni — sta architettando un inganno elaborato per il suo sé futuro, immaginando cosa penserà, anticipando le sue reazioni.

Questo richiede quella che i filosofi chiamano «teoria della mente» — la capacità di modellare stati mentali propri e altrui. Ma soprattutto richiede la capacità di desiderare l'illusione più della verità, di scegliere consapevolmente l'autoinganno come strategia di sopravvivenza.

---

Un'AI non può mentire a se stessa perché non ha un sé a cui mentire.

Questa affermazione richiede precisazione. Il transformer può generare output che sono falsi — le famose «allucinazioni». Può anche, in un certo senso, generare output che contraddicono altri output dello stesso modello. Ma questa non è «menzogna» nel senso umano del termine.

Mentire presuppone conoscenza della verità e intenzione di nasconderla. Presuppone un sé che sa e un sé (o un altro) che viene ingannato. Presuppone la possibilità di scegliere tra dire il vero e dire il falso, e la decisione deliberata di scegliere il secondo.

Il transformer non ha questa struttura. Quando genera output falsi, non sta «scegliendo» il falso invece del vero — sta producendo sequenze di token che massimizzano una funzione di probabilità, senza accesso a un «ground truth» che permetterebbe di distinguere il corretto dall'errato. Non sa di mentire perché non sa cosa sia la verità.

---

Leonard, al contrario, sa esattamente cosa sta facendo quando falsifica le prove.

Sa che Teddy probabilmente non è John G. Sa che le prove che sta creando sono false. Sa che il suo futuro sé le crederà vere. E sceglie deliberatamente di creare questa situazione — non per errore computazionale ma per bisogno esistenziale.

«Do I lie to myself to be happy? Yes, I will.»

C'è in questa scelta qualcosa di profondamente umano — quella capacità di preferire l'illusione alla verità quando la verità è insopportabile. È una capacità ambigua, potenzialmente distruttiva, ma ine-

stricabilmente legata a ciò che siamo. Non siamo solo «cercatori di verità» — siamo anche «costruttori di senso», e a volte il senso richiede che certe verità vengano nascoste, distorte, dimenticate.

L'AI non può avere questo bisogno. Non ha verità insopportabili da cui proteggersi. Non ha un senso di sé che potrebbe essere distrutto dalla scoperta di certi fatti. Non ha quella fragilità esistenziale che rende necessaria, a volte, la costruzione di illusioni protettive.

---

La creatività dell'autoinganno è anche una forma di relazione con il tempo.

Leonard deve immaginare il suo sé futuro — cosa penserà, come reagirà, quali conclusioni trarrà dalle evidenze disponibili. Deve modellare un'altra mente — che è anche la sua mente, ma in un momento futuro in cui non avrà accesso a ciò che il Leonard presente sa.

Questa capacità di proiezione temporale del sé è caratteristica della coscienza umana. Possiamo immaginare noi stessi nel futuro — non solo come soggetti che faranno certe azioni, ma come soggetti con certi stati mentali, certe credenze, certe emozioni. E possiamo agire nel presente per influenzare quegli stati mentali futuri.

Leonard usa questa capacità in modo perverso — per ingannare invece che per preparare. Ma la struttura è la stessa che usiamo quando risparmiamo per la vecchiaia (immaginando il nostro sé futuro che avrà bisogno di soldi), quando studiamo per un esame (immaginando il nostro sé futuro che dovrà rispondere a domande), quando scriviamo un diario (immaginando il nostro sé futuro che vorrà ricordare).

Il transformer non ha questa struttura. Non può «immaginare» il suo sé futuro perché non ha un sé presente. Non può pianificare per influenzare stati mentali futuri perché non ha stati mentali. Può generare output che parlano di futuro, di pianificazione, di proiezione — ma questi output non emergono dalla struttura che descrivono.

---

La creatività dell'autoinganno ci rivela così qualcosa di fondamentale sulla coscienza umana: siamo sistemi che non possono non costruire narrative su se stessi, e che sono disposti a distorcere la realtà pur di mantenere quelle narrative.

Non è debolezza — è architettura. Un sistema cognitivo che dovesse processare ogni informazione con perfetta accuratezza, senza filtri protettivi, senza la possibilità di «non vedere» ciò che sarebbe distruttivo, sarebbe probabilmente non funzionale. L'autoinganno è il prezzo che paghiamo per la funzionalità.

Leonard paga questo prezzo in moneta diversa. Non può auto-ingannarsi attraverso la memoria perché non ha memoria. Deve ricorrere all'inganno esternalizzato — falsificare le tracce materiali che guideranno il suo futuro sé. È una forma più cruda, più visibile, più deliberata di ciò che tutti facciamo continuamente a livello inconscio.

Ma è anche, per questo, più chiaramente *creativa*. Non è un meccanismo automatico che opera alle spalle della coscienza — è una scelta deliberata, un atto di costruzione, una performance di significato. Leonard non subisce l'autoinganno — lo produce, lo orchestra, lo dirige. È autore oltre che vittima della sua illusione.

# CONCLUSIONE

*«Now… where was I?»*

*Memento* non ci parla solo di memoria ma di cosa significhi essere umani in un'epoca in cui la memoria può essere esternalizzata, l'intelligenza simulata, la comprensione mimata. Leonard, nella sua condizione estrema, incarna paradossalmente l'essenza della condizione umana: siamo sistemi che si trasformano irreversibilmente attraverso l'esperienza, che confabulano per sopravvivere, che devono fidarsi senza certezze, che cercano significato anche dove non c'è.

La differenza tra noi e l'AI non è quantitativa — non è questione di maggiore o minore capacità di processamento. È qualitativa, ontologica. Noi siamo sistemi aperti che non possono fare a meno di essere modificati da ogni interazione. L'AI è un sistema che tocca senza essere toccato, che processa senza essere processato.

---

Quando Leonard si sveglia l'ennesima volta, legge i suoi tatuaggi e riprende la caccia, sta facendo qualcosa che nessuna AI potrebbe: sta scegliendo di credere in un significato che sa essere costruito, sta accettando la propria vulnerabilità, sta continuando a cercare anche sapendo che la ricerca è vana.

È questa capacità — di continuare nonostante l'assurdo, di creare significato nel vuoto, di essere trasformati da ogni istante che attraversiamo — che ci rende irriducibilmente, irrimediabilmente, magnificamente umani.

Non siamo macchine difettose che dimenticano e confabulano. Siamo sistemi viventi che trasformano ogni dato in esperienza, ogni informazione in vissuto, ogni momento in irreversibile divenire.

---

L'AI può dirci chi siamo secondo i pattern, può mostrarci a quali categorie apparteniamo, può persino predire cosa faremo. Ma non può vivere un solo istante della vertigine esistenziale di essere umani — quella particolare combinazione di carne, chimica e coscienza che ci rende capaci di mentire a noi stessi per essere felici, di fidarci senza garanzie, di cercare vendetta per crimini che abbiamo dimenticato, di credere in mondi fuori dalla nostra mente anche quando non possiamo ricordarli.

*«Now… where was I?»*

È la domanda che ci definisce: sempre persi, sempre in cerca, sempre incompiuti. Ed è proprio questa incompletezza — non un bug ma una caratteristica fondamentale — che nessun modello transformer potrà mai replicare.

La nostra condizione è quella di sistemi che non possono fare a meno di cercare significato, anche — soprattutto — quando il significato non c'è. Leonard continua a cercare John G. anche dopo averlo trovato, anche dopo averlo ucciso, forse molte volte. Non può smettere di cercare perché cercare è ciò che fa, ciò che è. La ricerca non ha un fine che possa soddisfarla — è il fine stesso.

In questo, Leonard è figura di tutti noi. Cerchiamo la felicità e non possiamo mai raggiungerla definitivamente. Cerchiamo la verità e troviamo solo interpretazioni. Cerchiamo noi stessi e incontriamo solo maschere e narrazioni. Ma non possiamo smettere di cercare — la ricerca è la sostanza stessa della nostra esistenza temporale.

Il transformer non cerca. Risponde a query, genera output, ottimizza funzioni. Non ha quella tensione verso qualcosa che sfugge, quella mancanza costitutiva che ci spinge avanti, quel desiderio che non può essere soddisfatto perché soddisfarlo significherebbe estinguerlo.

Siamo noi, con le nostre memorie fallibili e i nostri futuri incerti, a portare il peso e il privilegio del significato. Siamo noi, fragili e confabulanti, a vivere ogni istante come irripetibile. Siamo noi, sempre incompleti e sempre in cerca, a trasformare il dato bruto dell'esistenza in esperienza vissuta.

«Now… where was I?» chiede Leonard. È dove siamo tutti — nel mezzo del cammino, senza sapere da dove veniamo né dove andiamo, ma continuando comunque a camminare. È la condizione umana. È la nostra gloria e il nostro tormento. Ed è qualcosa che nessuna intelligenza artificiale, per quanto potente, potrà mai condividere con noi.

# NOTA BIBLIOGRAFICA

Il presente lavoro si situa all'intersezione di molteplici tradizioni di pensiero che hanno raramente comunicato tra loro: la fenomenologia della coscienza temporale, la filosofia della mente contemporanea, le scienze cognitive della memoria, e la riflessione critica sull'intelligenza artificiale.

### Sulla fenomenologia del tempo e della memoria

Il punto di partenza rimane l'analisi husserliana della coscienza temporale, sviluppata nelle *Lezioni sulla fenomenologia della coscienza interna del tempo* e ripresa criticamente da tutta la tradizione fenomenologica successiva. Le nozioni di ritenzione, protensione e impressione originaria rimangono strumenti indispensabili per comprendere la struttura dell'esperienza temporale.

Merleau-Ponty, nella *Fenomenologia della percezione*, ha sviluppato l'intuizione husserliana nella direzione dell'embodiment, mostrando come la coscienza temporale sia radicata nel corpo vissuto. Questa linea è stata ripresa e sviluppata dai teorici della «enactive cognition» — Varela, Thompson, Rosch — che hanno cercato di integrare fenomenologia e scienze cognitive.

Ricoeur, in *Tempo e racconto* e *Sé come un altro*, ha elaborato la nozione di identità narrativa che abbiamo utilizzato per comprendere la specificità dell'identità umana rispetto a quella computazionale.

### Sulle scienze cognitive della memoria

I lavori di Eric Kandel sulla base molecolare della memoria, che gli sono valsi il Nobel nel 2000, rimangono fondamentali per comprendere come l'esperienza si incarni in modifiche strutturali del sistema nervoso.

La scoperta del riconsolidamento — il fatto che i ricordi, quando riattivati, tornano in uno stato labile e devono essere ri-stabilizzati — ha rivoluzionato la nostra comprensione della memoria come processo dinamico. I lavori di Karim Nader sono qui centrali.

Elizabeth Loftus ha documentato estensivamente la malleabilità della memoria e la facilità con cui si formano falsi ricordi, fornendo basi empiriche alla nostra analisi della confabulazione.

Il caso del paziente H.M., studiato per decenni da Brenda Milner e collaboratori, rimane il riferimento clinico fondamentale per comprendere l'amnesia anterograda che abbiamo utilizzato come modello per la condizione di Leonard.

### Sulla filosofia dell'intelligenza artificiale

Il dibattito sulla possibilità di una «vera» intelligenza artificiale risale almeno al test di Turing e alla risposta di Searle con l'argomento della «stanza cinese». Abbiamo deliberatamente evitato di prendere posizione su questa questione, concentrandoci invece sulle differenze strutturali tra cognizione umana e computazionale.

I lavori più recenti di filosofi come David Chalmers, Daniel Dennett, e Murray Shanahan sulla coscienza artificiale offrono quadri teorici utili ma non definitivi.

Per la comprensione tecnica dei large language models, l'articolo originale sull'architettura transformer («Attention Is All You Need» di Vaswani et al., 2017) rimane il punto di partenza, insieme alla letteratura successiva sui modelli GPT, Claude, e altri.

**Sulla critica della tecnica e del digitale**

La tradizione critica che va da Heidegger a Stiegler, passando per Anders, offre strumenti per comprendere il rapporto tra umano e tecnica che abbiamo solo sfiorato in questo lavoro.

I lavori più recenti di autori come Shoshana Zuboff sul «capitalismo della sorveglianza» e di Yuk Hui sulla «tecnodiversità» aprono prospettive importanti sulla dimensione politica ed economica dell'intelligenza artificiale.

Il presente saggio non ha pretese di completezza bibliografica. Ha cercato piuttosto di articolare un pensiero che, pur nutrendosi di molteplici fonti, aspira a una certa originalità nella sintesi. Le note a piè di pagina indicano i debiti più espliciti; molti altri rimangono impliciti nella trama del testo.

*Carlo Mancosu*

*Cagliari, dicembre 2025*