La matematica dei modelli di riferimento

Introduzione: la nozione di MdR

L'intelligenza artificiale (IA) è basata sull'idea che quell'aggregato non ben definito di capacità che chiamiamo collettivamente "intelligenza" possa essere realizzata artificialmente – nello specifico, come un algoritmo. Non è sorprendente che, di conseguenza, l'IA fronteggi due difficoltà storiche: (1) come costruire la macchina destinata a implementare l'algoritmo ipotizzato e (2) come precisare quella nozione non ben definita: *intelligenza*. Gli odierni computer catturano in modo ancora rozzo alcuni aspetti di quella nozione – tipicamente, gli aspetti "non logici", analogici, associativi e contestuali. Ma i calcolatori di oggi a volte non se la cavano molto bene neppure quando giocano in casa, ossia nei campi della logica e della matematica (pensate alla fattorizzazione di grandi numeri e ai problemi non polinomiali).
Per affrontare le due difficoltà, il laboratorio di ricerca iLabs segue due idee fondamentali: (1) un isomorfismo teorico fra la realtà fisica e quella informazionale; e (2) un esteso sviluppo formale della nozione di *modello di riferimento* (MdR). Questo sviluppo formale è la *matematica dei modelli di riferimento*.

Quanto al primo punto: noi di iLabs crediamo che l'informazione sia radicata nella realtà fisica fin dalla sua base. Riteniamo che quel che accade nel mondo possa essere descritto in modo succinto, ma corretto, dicendo: ciò che l'universo *fa* in ogni istante di tempo è computare il suo stato complessivo all'istante successivo. Di conseguenza, l'intelligenza artificiale non è poi tanto artificiale: ci sono già ottimi processori là fuori! E ciò dipende dal fatto che la computazione universale è incorporata nell'essenza ultima della realtà.

Quanto al secondo punto: formalmente, un modello di riferimento (MdR) è una tripla ordinata <*percezione*, *pensiero*, *azione*>. Usando una nozione operazionale standard, f: A $\Rightarrow$ B è un MdR definito su un insieme A di percezioni, il cui pensiero proprio consiste nel mapparle su azioni od output in un insieme B. Possiamo allora scrivere:

$$f(x_1, x_2, \ldots, x_n) = y$$

con $x_1, x_2, \ldots, x_n \in$ A, e $y \in$ B. Al fondo del reale, un MdR funziona come una semplice modifica nei *pattern* informazionali delle più piccole unità fisiche che costituiscono il mondo (e su cui torneremo fra poco). Al livello umano, un MdR è ciò che spiega il nostro comportamento nei termini di come percepiamo ed elaboriamo informazioni.
Generalizzando: i modelli di riferimento sono entità di tipo "frattale", che esibiscono una struttura isomorfa a qualsiasi livello della realtà. E perciò, sono il nostro più promettente candidato al ruolo teorico di interfaccia funzionale fra materia e informazione. Potete leggere questa pagina perché i MdR rilevanti sono attivi nel vostro occhio, allo scopo di processare gli stimoli visivi e inviare segnali all'area del vostro cervello deputata alla visione.

Una rana può catturare una mosca attivando alcune centinaia di MdR. Reciprocamente, siamo fobici, o malati, o sviluppiamo il cancro, perché la nostra mente, il nostro corpo, le nostre cellule, attivano i modelli di riferimento "sbagliati". Ma anche le operazioni aritmetiche, un software gestionale, o la dimostrazione di un teorema in algebra universale, sono MdR. Quanto segue è una prospettiva sui risultati esposti con i dovuti dettagli formali nel libro iLabs [forthcoming], dove le opzioni teoriche qui esposte sono difese anche dal punto di vista filosofico.

1. Un universo discreto

Come concepire un universo in cui materia e informazione sono due facce della stessa medaglia? Ecco la proposta iLabs. L'universo (diciamo, U) dev'essere, per cominciare, *discreto* e *finito* – basato su *unità minime di spaziotempo* (quanto "minime"? 10^{-35} m per l'unità spaziale minima, e 10^{-44} secondi per quella temporale minima, potrebbero essere buone ipotesi di partenza; ma quali siano di preciso queste dimensioni ha un'importanza teorica limitata).

Chiamiamo *celle* questi atomi. Ci aspettiamo che lo spazio sia interamente occupato da celle morfologicamente identiche. Esiste dunque un numero finito *w* di celle, ossia di unità spaziotemporali minime. Dunque, anche il tempo è diviso in unità discrete minimali, gli istanti: t_0, t_1, ... (algebricamente: il tempo è un ordine lineare discreto). Il nostro spazio euclideo intuitivo è tridimensionale. Ma un universo discreto può anche avere la forma di uno spazio a una, due, tre, … *n* dimensioni dal punto di vista computazionale. Per modellare il nostro universo, abbiamo scelto una griglia esagonale bidimensionale, ma i risultati ottenuti dalla nostra ricerca sarebbero realizzabili anche implementando i nostri MdR in griglie quadratiche tradizionali, e anche in un ambiente a tre dimensioni suddiviso dall'analogo tridimensionale dell'esagono: il *dodecaedro rombico*.

Fig. 1: dodecaedro rombico

Esagono e dodecaedro rombico hanno svariati vantaggi topologici nella rappresentazione del movimento fisico – specificamente, la distanza fra celle può essere approssimata in termini di raggio:

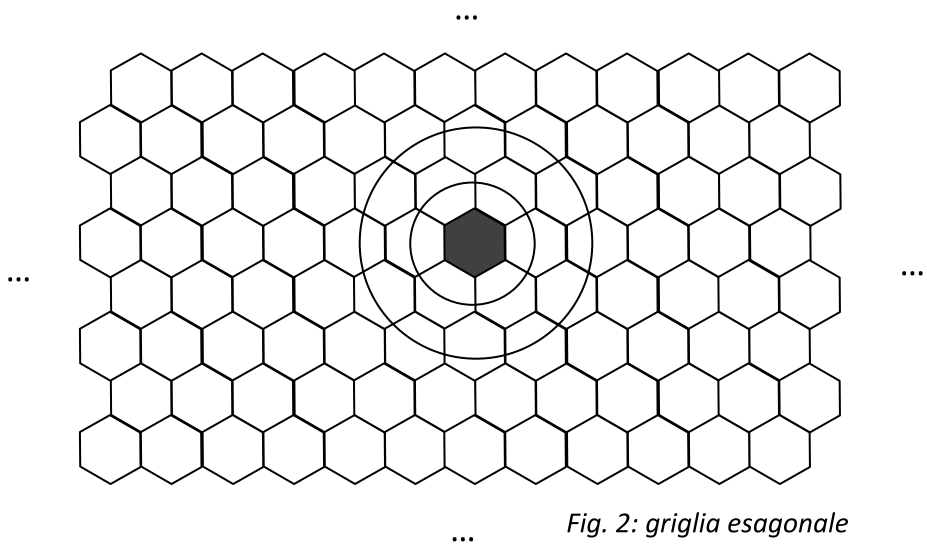

Fig. 2: griglia esagonale

Una volta fissata una base dello spazio, in un contesto bidimensionale ogni cella è univocamente individuata come pun punto in un reticolo da una coppia ordinata di numeri interi <*i*, *j*>. Ad ogni istante di tempo *t*, ogni cella <*i*, *j*> si trova in uno e un solo *stato* $\sigma \in \Sigma$, dove Σ è un insieme finito di stati discreti di cardinalità $|\Sigma| = k$. Indicheremo con "$\sigma_{i, j, t}$" lo stato della cella <*i*, *j*> al tempo *t*.

La nostra prospettiva è schiettamente convenzionalista: anzitutto, riteniamo che la sterminata varietà degli oggetti del mondo intorno a noi, con le loro proprietà, qualità e caratteristiche, emerga come una risultanza "di alto livello" di questi semplici ingredienti di base: celle atomiche, e un limitato numero di stati di base da esse istanziati. In secondo luogo, tutto è, da ultimo, un aggregato di celle. Chiamiamo *sistema* un qualsiasi aggregato. Allora, qualsiasi sistema è tanto "legittimo" quanto qualsiasi altro: i nostri oggetti ordinari sono semplicemente quegli aggregati che si accordano meglio col modo in cui suddividiamo il reale – e questo dipende dal nostro apparato cognitivo, dalle nostre capacità di apprendimento, e dai nostri interessi pratici.

L'universo non funziona a caso: regole precise determinano come ogni punto del reticolo *aggiorna* il proprio stato. Non sappiamo quali siano le regole di base, ma sappiamo per certo che devono essere *modelli di riferimento*: sequenze deterministiche di input, elaborazioni ed output. Ora, la nostra scommessa è che, al livello base della realtà, le regole debbano essere poche e semplici; complessità e varietà dovrebbero *emergere* a livelli superiori e dipendere dalla semplicità sottostante: *simplex sigillum veri*.

Non ci sono misteriose "azioni a distanza" nell'universo, ma solo interazioni locali: ogni punto <*i*, *j*> interagisce solo con le sei celle adiacenti, chiamate il suo *neighbourhood*:

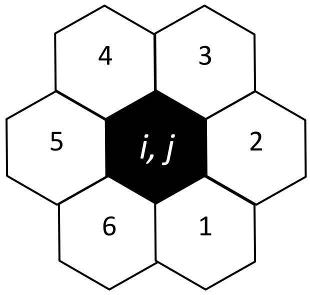

Fig. 3: una cella e il suo neighbourhood

Indichiamo con "[*i*, *j*]" il *neighbourhood* del punto <*i*, *j*>. Allora, una regola deterministica dinamica o MdR governa gli atomi – e dunque, il mondo: ad ogni istante *t*, ogni punto <*i*, *j*> aggiorna in sincrono il proprio stato rispetto all'istante *t*-1, seguendo un'unica regola o MdR ϕ: $\Sigma \Rightarrow \Sigma$, tale che per ogni σ, <*i, j*>, e *t*:

$$\sigma_{i,\, j,\, t+1} = \phi(\sigma_{[i,\, j],\, t})$$

Il nostro mondo ospita una *quantità globalmente finita di informazione*: data $|\Sigma| = k$ e un numero *w* di punti del reticolo, abbiamo al massimo k^w configurazioni globali per l'universo U. Di conseguenza, l'intera evoluzione di U è un grafo di transizione globale finito, G_Φ - il grafico della funzione di transizione globale Φ: $\Gamma \Rightarrow \Gamma$ (dove Γ è il *phase space* o l'insieme delle configurazioni globali di U) indotta dal MdR ϕ.

2. Il mondo di iLabs

2.1. Reversibilità perfetta per computazione *hi-tech*

Le leggi dinamiche della fisica sono reversibili a livello microscopico: distinti stati iniziali in un sistema microfisico portano sempre a distinti stati finali. È probabile che qualsiasi modello formale che miri a catturare le computazioni che hanno realmente luogo alle radici della realtà debba ospitare una dinamica *reversibile*.

D'altra parte, la *computazione* irreversibile comporta sempre uno spiacevole dispendio energetico. Un AND ci dà 0 come output a *t*+1. Qual era l'input a *t*? 0 e 1, o viceversa, o due zeri? Come aveva già mostrato von Neumann, questa "entropia informazionale" costa ~3 · 10^{-21} joules per ogni passo computazionale elementare a temperatura ambiente.
La perdita di informazione ha un costo termodinamico da pagare in termini di energia, per così dire, "non computazionale". La qual cosa non dipende da inefficacia nel *design* dei circuiti: segue direttamente dall'esistenza di computazioni irreversibili. Man mano che la tecnologia avanza, il problema dell'entropia informazionale è destinato a farsi vivo: potremmo aver bisogno di processori reversibili prima di quanto si creda.

Ai laboratori iLabs abbiamo sviluppato un modello matematico (un automa cellulare a stati finiti) dotato di una dinamica perfettamente *reversibile* e capace di conservare la totalità dell'informazione archiviata all'inizio dell'universo. Quindi, abbiamo fornito una dimostrazione effettiva del fatto che il nostro automa è capace di *computazione universale*: il nostro universo discreto computazionale può ospitare macchine di Turing universali, capaci di computare (data la Tesi di Turing) tutto ciò che è computabile. Ciò dipende dal fatto che ogni singola cella del nostro universo istanzia un singolo MdR come primitivo computazionale. Il nostro modello è quindi un buon candidato per la realizzazione di sistemi computazionali ad alte prestazioni: sistemi capaci di ospitare circuiti logici che computano con dissipazione di energia interna virtualmente nulla.

2.2. La regola iLabs

Ogni punto del reticolo nel modello iLabs istanzia stati che sono *sestuple* di bit dell'insieme {1, 0}. Per aiutare l'intuizione, si può pensarli come implementati nei lati di una cella esagonale:

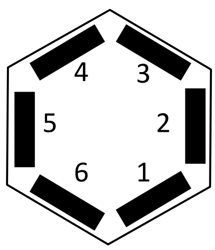

Fig. 4: i sei lati di una cella che istanzia una sestupla

Ogni stato ha il valore 1 ("attivo") o 0 ("inattivo") ad ogni istante di tempo: <*i*, *j*> a *t* ha stato $\sigma_{i, j, t} = \langle x_1, x_2, x_3, x_4, x_5, x_6 \rangle$, x_1 corrispondendo al lato n. 1, x_2 al lato 2, etc., dove ogni *x* ha valore $v \in V = \{1, 0\}$. Poiché $|V| = 2$, $|\Sigma| = 64$: ogni cella ha $2^6 = 64$ stati possibili.
Il *neighbourhood* [*i*, *j*] del punto <*i*, *j*> è determinato dal valore $v \in V$ dei sei lati adiacenti:

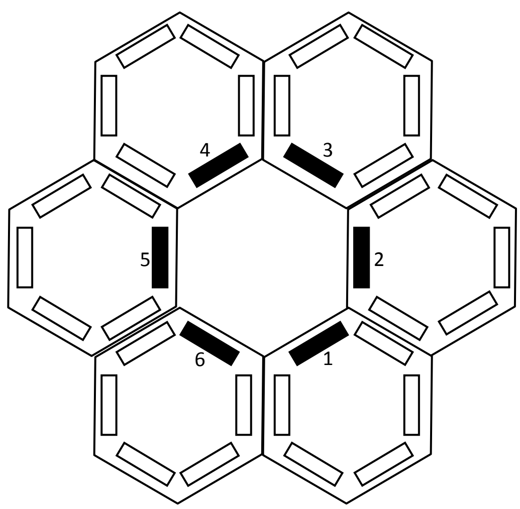

Fig. 5: i lati adiacenti del neighbourhood

Ancora intuitivamente: possiamo pensare ogni atomo o cella come in grado di *percepire* (input) l'output esibitole dal *neighbourhood*, e di *agire* (output) esponendo il risultato della propria computazione, determinato dal seguente MdR ϕ: $\{1, 0\}^6 \Rightarrow \{1, 0\}^6$, che non è altro che un *routing* condizionale di segnali:

$$\sigma_{i,\,j,\,t+1} = \begin{cases} Perm(\sigma_{[i,\,j],\,t}), \text{ se } \sum_{v[i,\,j],t} (\text{mod } 2) = 1 \\ \\ Id(\sigma_{[i,\,j],\,t}) \text{ altrimenti} \end{cases}$$

Dove "$\sum_{v[i,\,j],t}$ (mod 2)" indica la somma modulo 2 dei $v \in \{1, 0\}$ di ogni membro della sestupla che costituisce l'intput a t. *Id* è semplicemente la funzione identità, che mappa ogni sestupla $\sigma \in \Sigma$ in se stessa:

$Id(<x_1, x_2, x_3, x_4, x_5, x_6>) = <x_1, x_2, x_3, x_4, x_5, x_6>$.

Perm è una permutazione definita su Σ, ossia un operatore $\Sigma \Rightarrow \Sigma$ che scambia i primi tre membri della sestupla con gli ultimi tre:

$Perm(<x_1, x_2, x_3, x_4, x_5, x_6>) = <x_4, x_5, x_6, x_1, x_2, x_3>$.

Un numero dispari di 1 nella sestupla in input produce come output la loro permutazione: qualsiasi segnale in ingresso è trasmesso dal punto $<i, j>$ al suo lato opposto (daccapo, questa descrizione serve solo ad aiutare l'intuizione: i "lati" della cella sono codificati nella computazione). Un numero pari di 1 attiva *Id*, e l'effetto è un *kickback*: i segnali (parlando ancora intuitivamente) vengono rimbalzati da dove venivano.

ϕ è una (peculiare) regola *outer-totalistic* per automi cellulari, che garantisce una *conservazione totale* del numero di bit "attivi" e "inattivi" presenti all'inizio dell'universo: visto che *Perm* e (banalmente) *Id* sono permutazioni, l'output di ogni sestupla in input conserverà il numero di zero e di uno dell'input.

Inoltre, ϕ è *reversibile*: ogni input o percezione è mappato dalle due sottoregole di ϕ, *Id* e *Perm*, in un output o azione distinto. Esiste dunque un MdR inverso, ϕ^{-1}, che mappa gli output di ϕ ai rispettivi input.

Ma ϕ è anche *fortemente* reversibile, ossia, $\phi^{-1} = \phi$. Infatti, ϕ è una regola *time-reversal invariant*: è mappata a ϕ^{-1} dalla trasformazione $t \mapsto -t$ che inverte il corso del tempo. E questo vuol dire che in qualsiasi istante di tempo t possiamo recuperare qualsiasi stato dell'universo U a t-n (fino alla configurazione iniziale di U – l'inizio dell'universo) facendo scorrere il sistema all'indietro, sulla base di ϕ stessa.

2.3. Computazione universale

Il MdR ϕ è capace di computazione universale:il nostro universo discreto U può essere abitato da macchine di Turing universali (MTU) capaci di computare tutto il computabile. Abbiamo esposto una dimostrazione costruttiva di questo fatto nel volume iLabs [forthcoming], dove proviamo che ogni primitivo della computazione di una MTU convenzionale, ossia (a) memoria (b) trasmissione e (c) computazione di segnali attraverso un insieme funzionalmente completo di circuiti logici, può essere emulato da *pattern* prodotti mediante ϕ nell'automa cellulare in questione.
Nello specifico: implementando ϕ, ogni punto <*i*, *j*> del reticolo che costituisce il nostro universo discreto svolge la doppia funzione di archiviazione e trasmissione, cosicché ogni catena di celle adiacenti funziona come un circuito lungo il quale bit di informazione possono spostarsi.

Inoltre, ogni singola cella *è* essa stessa un *gate* logico universale, precisamente in quanto è un luogo per la trasmissione e la deviazione di segnali, implementando l'insieme funzionalmente completo {AND, OR}, e anche il FAN-OUT di segnali (daccapo: si veda il nostro libro [forthcoming] per i dettagli).

Se chiamiamo $\overline{\sigma}_0$ il vettore che esprime la configurazione iniziale di U (la *w*-pla ordinate di stati di ciascuno dei *w* atomi della griglia all'inizio dell'universo), l'evoluzione di U è prescritta da ϕ (con la corrispondente regola globale Φ). E una tale evoluzione può produrre macchine di Turing universali, e così implementare qualsiasi algoritmo e valutare qualsiasi affermazione computabile.

Visto che il nostro universo discreto può ospitare MTU, la sua evoluzione è *imprevedibile* in un senso preciso: dato il Teorema della Fermata, non c'è alcun algoritmo generale in grado di predire se una data MTU si arresterà dopo *n* passi, dato un certo input. Non c'è quindi alcuna scorciatoia computazionale per prevedere l'evoluzione di U: si può solo sedersi ed osservarla.

3. I MdR come concetti

Pensare ai nostril MdR come triple <percezione, pensiero, azione> può suonare, *prima facie*, come una sorta di "animismo semantico". Eppure, i MdR possono ricomprendere la nozione generale di *operatore*, a cui buona parte dei concetti matematici può essere ridotta. La matematica dei MdR è una "matematica del pensiero", in quanto fornisce un metodo formale, e matematicamente rispettabile, per precisare molte nozioni cognitive. La stessa nozione di *concetto* può esser trattata in modo soddisfacente attraverso la teoria dei MdR.

La natura dei concetti è stata al centro di discussioni filosofiche per oltre duemila anni: i concetti sono stati talvolta ridotti a entità fisiche e materiali, talvolta innalzati nel cielo delle "idee" platoniche, o dei *Sinne* freghiani, etc. Noi di iLabs crediamo che i concetti *siano* nient'altro che MdR: regole che mappano in modo effettivo un certo segnale in entrata o percezione o input ad un certo segnale determinato in uscita, o output, o azione, in seguito a un'elaborazione interna o pensiero. O, più modestamente: accettando di identificare convenzionalmente la nozione tradizionale di concetto con quella di MdR, e sviluppando formalmente la seconda, si può ottenere una chiarificazione della prima. I MdR sono "concetti al lavoro": svolgono i ruoli teoretici assegnati ai concetti, e forniscono precisi realizzatori fisico-informazionali per essi.

Trattando i MdR come operatori, iLabs ha potuto importare nella matematica dei MdR la teoria tradizionale delle funzioni computabili. Di seguito, forniamo solo qualche assaggio della teoria dispiegata.

3.1. MdR equivalenti

I MdR hanno *condizioni di equivalenza* chiaramente definite. Sia $\approx$ una relazione di equivalenza generica che determina una partizione dell'insieme dei MdR in classi di equivalenza. Possiamo introdurre criteri o condizioni di equivalenza fra MdR attraverso assiomi che istanziano lo schema:

(CE) $(\alpha[x/f] \leftrightarrow \alpha[x/g]) \rightarrow f \approx g$

dove α è la condizione in questione. Due MdR dati f e g possono essere considerati come equivalenti sotto la $\approx$ rilevante se soddisfano entrambi la condizione prescelta α.

Una condizione minimale per qualsiasi relazione di equivalenza fra MdR (ME) è la seguente. Affinché f e g siano minimalmente equivalenti ($f \approx_0 g$), è necessario che condividano lo stesso insieme di percezioni ed azioni. La condizione sufficiente per $f \approx_0 g$ è che mappino le stesse percezioni o input sulle stesse azioni:

Dati due MdR qualsiasi f: A $\Rightarrow$ B, e g: C $\Rightarrow$ D, se:

(a) A = C e B = D,
(b) Per ogni input $x_1, x_2, \ldots, x_n$, f e g danno lo stesso output y,

Allora, $f \approx_0 g$.

(ME) è ancora un criterio debole, non includendo il fattore tempo. Un'equivalenza temporalmente qualificata, $\approx_1$, aggiungerà ad (a) e (b) la clausola:

(c) … e l'output y è prodotto da f e da g nello stesso numero n di unità di tempo;

Allora, $f \approx_1 g$.

Si possono ottenere equivalenze più strette aggiungendo clausole – in particolare, tali che catturino un isomorfismo procedurale fra MdR. In questo caso, vogliamo che f e g siano equivalenti ($f \approx_2 g$) se e solo se non solo "fanno le stesse cose negli stessi tempi", ma anche, con "pensieri", ossia procedure di computazione passo-passo, diciamo, P_f e P_g, tali che si possa stabilire un isomorfismo *i* fra P_f e P_g. Ci serve allora una quarta condizione (d) della forma:

(d) … e l'output *y* è prodotto da *f* e da *g* attraverso due procedure di pensiero P_f e P_g, tali che $i(P_f, P_g)$,

Allora, $f \approx_2 g$.

Una buona caratterizzazione matematica di *i* è fornita nel libro iLabs [forthcoming] mettendo le computazioni effettuate da un insieme di MdR ricorsivi in forma canonica, e quindi aritmetizzandole attraverso procedure standard di codifica: l' *i* richiesto è poi definito sui codici numerici rilevanti.

4. MdR ricorsivi, meta-modelli

Intuitivamente, un MdR ricorsivo è un MdR che "si riferisce a se stesso" nella propria definizione. Un po' più precisamente: un MdR ricorsivo è un MdR tale che le sue azioni od output sulla base di certe percezioni od input possono essere determinate dall'output o azione di *quello stesso* MdR rispetto a input o percezioni più *semplici*: quello che un MdR ricorsivo fa per certe percezioni più complesse dipende da quello che *esso stesso* fa, o farebbe, per percezioni più semplici (e, come mostrato in dettaglio nel nostro libro, anche "più semplice" è caratterizzabile in modo matematicamente preciso). I MdR ricorsivi possono essere introdotti, quindi, con definizioni ricorsive standard: l'operatore nel *definiendum* ricorre nel *definiens*.

Catturando integralmente la teoria degli operatori ricorsivi, la matematica dei MdR ha consentito a iLabs di definire la nozione fondamentale di *meta-modello*. Informalmente, un meta-modello non è altro che *un modello di riferimento capace di percepire altri modelli di riferimento e di operare su di essi*. Tecnicamente, ciò è stato ottenuto codificando i MdR stessi come stati delle celle che abitano il nostro universo digitale U: un meta-modello può quindi prendere come input i codici dei MdR che percepisce.

Incorporando il Teorema di Ricorsione (Forte) di Kleene nella teoria dei MdR, è stato quindi dimostrato che ci può essere un MdR universale, capace di emulare i pensieri di qualsiasi MdR ricorsivo. Questo è un MdR (parziale) della seguente forma:

$$univ(e, \langle x_1, \ldots, x_n \rangle).$$

Dato un MdR ricorsivo *n*-ario $f(x_1, \ldots, x_n)$ con codice *e*, ossia, $[e]_n(x_1, \ldots, x_n)$, *univ* prende come input il codice di f e (il codice per) il suo input, e fornisce come output quello che sarebbe dato da *f* o $[e]_n$:

$$f(x_1, \ldots, x_n) \cong [e]_n(x_1, \ldots, x_n) \cong univ(e, \langle x_1, \ldots, x_n \rangle).$$

(daccapo: i dettagli sono nel libro iLabs). Ovviamente, *univ* corrispodne a una MTU e, dati i risultati presentati nella Sezione 2, può essere realizzato nel nostro universo digitale U.

5. Autoreferenza ricorsiva e oltre

L'autoreferenza ricorsiva è la situazione che ha luogo quando un MdR non soltanto richiama se stesso, ma *è "consapevole"* di farlo – cosicché un sistema che lo implementa si può "accorgere" di ciò che sta facendo mediante quello stesso MdR.

Daccapo, questo non è animismo semantico. Infatti, espressioni come "è consapevole" possono ottenere un significato matematico ben preciso. Molta della ricerca condotta da iLabs è guidata dalla persuasione che l'autoreferenza ricorsiva sia alla base di ciò che ordinariamente chiamiamo "coscienza": una differenza essenziale fra l'attività coscienziale e un qualsiasi altra procedura computazionale consiste nel fatto che la nostra mente, in quanto (auto)cosciente, ha la capacità di pensare, di avere un punto di vista su, se stessa – probabilmente con limiti operativi alla capacità di operare sul proprio "codice sorgente". Se l'Intelligenza Artificiale si deve realizzare, lo farà attraverso l'autoreferenza ricorsiva – o questa è la nostra scommessa.

Come mostrato nel nostro libro, i Teoremi di Ricorsione, applicati ai nostri MdR ricorsivi, garantiscono che possiamo definire MdR parziali che sono ricorsivamente autoreferenziali, perché *includono il proprio stesso codice nelle proprie definizioni ricorsive*. Queste sono, semplicemente, definizioni per punto fisso. Dato che i codici numerici sono percezioni prese in input da (meta-)modelli di riferimento, che possono anche emulare le procedure di pensiero effettuate dai MdR codificati, i MdR ricorsivamente autoreferenziali possono percepire se stessi in senso matematicamente preciso, e rappresentare in sé la stessa procedura computazionale in cui consistono.