

# LE AI SOGNANO PECORE ELETTRICHE?

di Nicola Villa, editor di Altreconomia

In “Blade runner”, titolo originale “Do androids dream of electric sheep?” (1968), Philip K. Dick si inventa un futuro post-apocalittico: sulla Terra, ormai ricoperta di palta, quasi tutte le specie sono estinte, gli umani sopravvissuti non sono più in grado di provare emozioni, neanche di empatizzare con le repliche robotiche di animali da compagnia, mentre gli androidi ingannano i test che dovrebbero smascherarli, iniziano ad avere coscienza di sé e della loro condizione. È probabile che vogliano ribellarsi. È fantascienza, ma ha la capacità di inquietare e di suggerire domande filosofiche: se una macchina pensa allora è? Può una macchina essere intelligente? Esiste un’intelligenza artificiale (Ai)? Le Ai sognano pecore elettriche?

Tutte queste domande – e allucinazioni – sono oggi molto attuali, dato che l’Ai sta squattando il dibattito pubblico e sta deviando tutti i soldi della Silicon Valley, dopo gli abbagli del Metaverso e di altre tecnomagie.

Per fortuna c’è la letteratura ad aiutarci a comprendere la realtà, lo sa bene un fisico come Stefano Borroni Barale, cavaliere Jedi della pedagogia hacker e profeta di Linux: per Altreconomia ha scritto, vent’anni fa, uno dei primi libri sul “pinguino”, “Come passare al software libero e vivere felici. Manuale di autolibrazione informatica”.

Borroni Barale ha avuto un’epifania affrontando l’intelligenza artificiale: ha riletto “i nostri antenati”, cioè “Il cavaliere inesistente” di Italo Calvino – proprio nell’anno del centenario dalla nascita del nostro scrittore più post-moderno che si era anche appassionato di cibernetica –, per arrivare a una felice

intuizione che dà il titolo a questo libro. Non è però una trovata di copertina: l’Ai attuale è una grande operazione ideologica e di marketing, confezionata per aumentare il controllo delle persone e restringere il margine di libertà digitale.

In questo agile testo, tra i più aggiornati su questi temi e figlio del dibattito internazionale e della comunità hacker, Borroni Barale ricostruisce la storia dell’Ai da Alan Turing fino a ChatGPT, scoprendo paradossi e obiezioni come quella del matematico Kurt Gödel sulla ricorsività e l’incompletezza delle teorie matematiche. “L’intelligenza inesistente” è un testo divulgativo alla portata di tutti, anche quando spiega concetti molto complessi. Nella prima parte analizza, inoltre, com’è fatta l’Ai, cosa sono i Large language model, gli assiomi e il deep learning, e soprattutto le conseguenze tecno-politiche di questa sulla sorveglianza, la privacy e i costi umani e ambientali.

Nella seconda parte, dopo un intermezzo che smonta la retorica dei technoentusiasti, Borroni Barale propone delle alternative autonome di Ai, da realizzare anche in classe con gli studenti e le studentesse. Ecco quindi l’altro nume tutelare di questo libro dopo Calvino: Ivan Illich, il filosofo che ha teorizzato la società conviviale. Non immaginiamoci tavole imbandite e gente che brinda: la società conviviale è quella in cui gli strumenti tecnologici sono a disposizione della collettività e sono compresi, condivisi e utilizzati da tutte le persone che ne fanno parte, anziché essere creati e controllati unicamente dagli specialisti dello sviluppo e della produttività industriale illimitata. Di conseguenza, la convivialità rappresenta la realizzazione concreta della libertà individuale all’interno di strutture organizzate in senso non necessariamente gerarchico, potenziale strumento per la costruzione di un futuro assai meno distopico di quello di “Blade Runner”.

Abbiamo bisogno di teste affilate, come quella di Borroni Barale, per provare a condurci, invece che a “dar la caccia agli androidi”, a costruire, insieme a essi, una nuova “convivialità elettrica”, dove ogni forma di comando e obbedienza sia dissolta per sempre.

# COS'È L'INTELLIGENZA ARTIFICIALE

## WarGames e Artificial Intelligence

Mi piacerebbe poter dire che non ci sia pericolo di un lancio nucleare accidentale oggi. Ma quando c'è in ballo lo spazio – quando giochiamo a fare dio – chi può sapere quale sarà la prossima sorpresa?

*Stanisláv Petrův*

Il 7 maggio 1983 a Cannes veniva presentato un film che sarebbe passato alla storia, tra gli altri motivi, per il modo in cui descriveva la figura degli hacker. Era “WarGames” di John Badham, con protagonista un giovanissimo, ma convincente, Matthew Broderick agli esordi. Broderick interpretava un adolescente un po' *nerd* e un po' ribelle con la passione per la tecnologia, maestro nel *phone phreaking*: l'arte hacker di sbloccare le chiamate a lunga distanza senza inserire monete nella cabina telefonica. Normalmente dedito a collegarsi illegalmente al computer scolastico per modificare i suoi voti,<sup>1</sup> il giovane finisce con l'intrufolarsi in uno dei computer del Norad (il Comando di difesa

aerospaziale del Nord-America), e comincia a giocare con questo una partita a un gioco che lui pensa innocente, nonostante il nome: “Guerra Termonucleare Globale”. Peccato che il computer in questione sia un vero simulatore di scenari di guerra atomica e, soprattutto, peccato che politici poco accorti – *per essere certi di eliminare dall’equazione l’elemento umano* – abbiano appena collegato questo computer, curiosamente battezzato con un nome proprio (Joshua), *direttamente* ai silos di lancio dei missili nucleari. La ragione di una scelta tanto scellerata è narrata nella potente scena di apertura del film, in cui uno dei tre militari rinchiusi in un silos missilistico si rifiuta di premere il bottone rosso allo scattare dell’allarme per un attacco che non sanno essere simulato.

Non racconto il finale, a vantaggio di chi non avesse mai visto il film, un *geek cult* assoluto; è più interessante invece ricordare che la realtà imitò da lì a poco la fantasia. Il 26 settembre di quello stesso anno Stanisláv Petrův, allora tenente colonnello dell’armata sovietica, aveva il turno di notte nel *bunker* Serpukhov 15, con il compito di controllare i dati dei satelliti spia Oko (occhio), puntati sui movimenti statunitensi. In caso di attacco il sistema di allerta, che Petrův aveva contribuito a progettare, avrebbe dato ai sovietici il tempo di reagire con un contrattacco. Era la dottrina Mad, Mutua assicurata distruzione, che garantiva *l’equilibrio del terrore* tra le due superpotenze atomiche.

Quella notte l’allarme scattò: gli schermi indicavano un missile in arrivo da una base americana nel Montana, diretto verso il territorio sovietico. Petrův fece spegnere l’allarme e analizzò la situazione. Se gli americani avessero voluto attaccare l’Urss – pensò – non l’avrebbero fatto con un solo missile: voleva dire rischiare di essere obliterati prima di avere il tempo di portare a termine il loro attacco. Ma pochi attimi dopo l’allarme suonò ancora: altri quattro missili in arrivo. Per fortuna la procedura burocratica non dettava con precisione i tempi per la comunicazione dell’allarme ai massimi livelli di comando. Il tenente colonnello decise che, anche in questo caso, si doveva trattare

di un falso allarme. Quando comunicò con i suoi superiori descrisse la situazione come “un errore nel sistema”. Aveva ragione: si trattava della luce abbagliante del sole riflesso dalle nuvole. Lui l’aveva intuito e, bloccando la rappresaglia, aveva salvato l’umanità.

Immagino la perplessità di fronte a questo racconto uscito dal pieno della Guerra fredda. Che cosa potrà mai avere a che vedere tutto ciò con il funzionamento dell’Ai? Nel prosieguo del nostro racconto, scopriremo diverse connessioni. La prima, storica: la disciplina che oggi chiamiamo Ai nasce nel 1956 a causa della guerra fredda e del maccartismo. La sua logica è militare, di intelligence. Il suo scopo: sottomettere il nemico e schiacciarlo con un perfetto dominio della nascente *information theory*; ponendo fine bruscamente agli approfondimenti e alle cautele che stavano al centro della sua disciplina-madre, la Cibernetica. Infatti, proprio a causa del maccartismo e della Guerra fredda, la Cibernetica era caduta in disgrazia per via delle accuse di comunismo rivolte al suo fondatore, Norbert Wiener. La seconda: un aspetto profondamente ideologico, una fascinazione per la superiorità della macchina sull’uomo di gusto quasi futurista, negli scritti degli entusiasti dell’Ai.

Ma esiste una connessione più diretta, ed è questa: all’epoca di Petróv le macchine a disposizione funzionavano in maniera perfettamente trasparente. Ciò rendeva il loro comportamento assolutamente prevedibile: “SE scatta l’allarme ALLORA lancia i missili”. Ovviamente nessuno sarebbe così folle da mettere in mano a un meccanismo simile il futuro della propria sopravvivenza sul pianeta.

Ma i sistemi di presa di decisione e gli automatismi odierni vengono dipinti come *intelligenti* e – cosa assai più importante – privi delle pericolose emozioni umane, spesso fonte di decisioni impulsive potenzialmente catastrofiche. Allora, come dichiarato di recente da un sorridente Elon Musk durante un’intervista, sembra assai più logico il suo parere secondo il quale “dovremmo lasciarci guidare dai numeri”: se i sistemi di apprendimento

automatico applicati alla guida causano statisticamente meno morti di un guidatore umano, *forse sarebbe il caso di lasciar fare a loro.*<sup>2</sup>

Peccato che, se al posto di controllo del *bunker* Serpukhov collocassimo oggi uno di questi sistemi, e si verificasse un errore simile a quello del 26 settembre 1983... nulla potrebbe salvarci dall'inverno nucleare. Questa previsione apparentemente *tranchant*, si basa in realtà sull'analisi del lavoro di alcuni tra i maggiori esperti al mondo di sistemi di presa di decisione automatizzati, come Yann LeCun, vincitore del premio Turing e leader della ricerca Ai di Meta (l'azienda precedentemente nota come Facebook).

Durante una recente conferenza a Parigi,<sup>3</sup> LeCun ha presentato la sua ultima fatica: l'architettura I-Jepa, un nuovo modello per costruire Ai, che è convinto rappresenterà un passo avanti sostanziale rispetto alla Matematica e alla tecnologia che costituiscono le fondamenta di ChatGPT.

---

### **Sistemi imperativi e trasparenti vs black box**

Con "trasparente" intendiamo che il funzionamento di tali sistemi era perfettamente comprensibile e verificabile usando nozioni di informatica elementari. Ad esempio, il codice Basic seguente:

```
10 INPUT "Inserisci un numero: "; numero
20 LET risultato = numero + 5
30 PRINT "Il nuovo numero è: "; risultato
```

è di facile comprensione: il programma stampa sempre il valore che gli abbiamo dato in input, incrementato del numero 5.

Lo stesso non si può dire dei moderni Llm. Non è solo questione della loro complessità: il modo in cui funzionano le reti di interruttori a soglia ("reti neurali" nel gergo Ai), con le loro soglie programmate statisticamente, è tale per cui comprendere cosa abbia generato un certo risultato è sempre operazione molto complicata e, se il designer non conserva tutti i dati utilizzati durante l'ottimizzazione del sistema ("addestramento" nel gergo Ai), può divenire addirittura impossibile.

Come vedremo, l'obiettivo del lavoro di LeCun è aumentare la capacità di questi sistemi, rendendoli capaci di *inferenze a termine multiplo*, ossia basate non solo sui dati passati, ma da più fonti differenti sia per il momento in cui vengono rilevati sia per la tipologia (per esempio: testuali e grafici). La ricerca di LeCun si ispira a recenti articoli pubblicati sulle maggiori riviste di neuroscienza. I neuroscienziati, infatti, sono convinti che il nostro cervello implementi meccanismi di funzionamento simili a quelli previsti dall'architettura I-Jepa, anche se suggeriscono che sia la *multimodalità* – ossia prendere in considerazione sia testo che immagini o audio – che la capacità di produrre inferenze a termine multiplo, potrebbero essere solo parte di un funzionamento ancora più complesso.

Torniamo dunque a Petróv: la sua decisione si basò su una *felice intuizione*. Non poteva avere la *certezza deterministica* di essere nel giusto, poteva al massimo sperarlo. Probabilmente ciò che salvò l'umanità fu il suo rifiuto di arrendersi all'avvento della morte assoluta che avrebbe avvolto il pianeta, colpendo almeno il 99% degli homo sapiens e la maggioranza delle specie viventi insieme a lui. Infatti, non c'è oggi come non c'era allora alcun dato storico su passate guerre atomiche o anche solo su guerre globali tra superpotenze nucleari; al massimo si potevano prendere in considerazione le strategie utilizzate da Stati Uniti e Urss nei

---

### **Inferenze a termine singolo vs multiplo**

Un sistema è detto a inferenza singola se basa il suo output su una sola fonte di dati, tipicamente dati rilevati nel passato. Al contrario il sistema si dirà a inferenza multipla se basa l'output su più fonti di dati, una delle quali sicuramente è rilevata nel presente. Lo sviluppo dell'AI si sta muovendo verso l'inferenza a termine multiplo tra dati storici e dati rilevati "a runtime" ossia durante l'esecuzione dell'algoritmo. Questa classificazione vale sia per i grandi modelli linguistici, che devono "predire" la parola successiva<sup>4</sup> nella serie concatenata che forma una frase, sia per i sistemi di presa di decisione che devono "scegliere" quale sia la migliore azione in base ai dati in loro possesso.

conflitti locali come il Vietnam o nella fase finale della Seconda guerra mondiale. Un'*inferenza singola* legata al passato – l'unica di cui sono capaci le Ai odierne – non fornirebbe quindi alcuna informazione utile alla decisione, per via della evidente mancanza di dati. A quel punto l'AI, a seconda delle scelte del suo programmatore, potrebbe andare in "blocco" o ricadere nella dottrina Mad: tu spari, io rispondo. Risultato: estinzione. Una mano di roulette russa, giocata contro il destino. Certamente non il più raccomandabile degli scenari.