

Gino Roncaglia

L'architetto e l'oracolo

Forme digitali del sapere da Wikipedia a ChatGPT

Editori  Laterza



i Robinson / Letture

Gino Roncaglia

L'architetto e l'oracolo

Forme digitali del sapere da Wikipedia a ChatGPT



Editori Laterza

© 2023, Gius. Laterza & Figli

Edizione digitale: ottobre 2023

www.laterza.it

Proprietà letteraria riservata
Gius. Laterza & Figli Spa, Bari-Roma

Realizzato da Graphiservice s.r.l. - Bari (Italy)
per conto della
Gius. Laterza & Figli Spa

ISBN 9788858153659

È vietata la riproduzione, anche parziale, con qualsiasi mezzo effettuata

Indice

Introduzione

Parte I

1.
Frammentazione e complessità (una sintesi)
2.
Dall'enciclopedia mondiale alle enciclopedie digitali
3.
Pionieri dimenticati e primi servizi online: alle origini dell'enciclopedismo digitale
4.
Il multimedia all'assalto della carta
5.
Le enciclopedie online e la nascita di Wikipedia
6.
Wikipedia funziona?
7.
Il sogno del web semantico

Parte II

8.
Architetti, oracoli, pappagalli
9.
Il contesto: IA e reti neurali

10.

Le IA generative

11.

Allucinazioni e pregiudizi

12.

Noi e loro: le IA generative ci assomigliano?

13.

Nove tesi (e uno sguardo al mondo del libro)

Tesi generali

Tesi specifiche

L'oracolo può collaborare con l'architetto?

Parte III

14.

Memorie personali e mente estesa

15.

Trasferimenti di memoria

16.

Il sogno della memoria totale

17.

Nuove memorie personali e big data

Parte IV

18.

Fantascienza?

19.

La biblioteca e l'impero

20.

Bibliotecari nel metaverso

21.

Virtual Librarian

Conclusioni

a Grazia

Il terzo imperatore della dinastia Ming, conosciuto come Yung Lo (1403-1405), istituì una commissione per la realizzazione di un'opera su una scala che era colossale persino per la Cina. La sua idea era di raccogliere tutto ciò che era stato scritto nei quattro ambiti rappresentati da (1) il canone Confuciano, (2) la storia, (3) la filosofia e (4) la letteratura generale, inclusi astronomia, geografia, cosmogonia, medicina, divinazione, Buddismo, Taoismo, arti e mestieri. Nel 1408 questa enciclopedia fu presentata al Trono, ricevette l'approvazione imperiale e fu intitolata *Wung Lo Ta Tien*, o *Il Grande Standard di Yung Lo*.

Per realizzarla, 3 commissari, con 5 direttori, 20 sottodirettori, e uno staff di 2.141 redattori, avevano lavorato per cinque anni. Il suo contenuto era suddiviso in non meno di 22.877 sezioni separate, a cui si doveva aggiungere un indice che riempiva 60 sezioni.

Ogni sezione conteneva circa 20 fogli, portando il totale a 917.480 pagine per tutto il lavoro. Ogni pagina consisteva di sedici colonne di caratteri con una media di venticinque caratteri per ogni colonna, per un totale di 366.992.000 caratteri. [...] Questo straordinario lavoro non fu mai stampato, poiché la spesa sarebbe stata eccessiva, sebbene fosse stato effettivamente trascritto a tal fine; successivamente ne furono realizzate altre due copie, una delle quali fu infine conservata a Pechino, e l'altra, con l'originale, a Nanchino. Entrambe le copie di Nanchino andarono distrutte con la caduta della dinastia Ming, e una sorte analoga toccò alla copia di Pechino, con l'eccezione di pochi volumi sparsi, durante l'assedio delle legazioni nel 1900. Quest'ultima era divisa in 11.000 volumi, rilegati in seta gialla; ciascun volume misurava circa 51 cm di lunghezza per 30,5 cm di larghezza, e aveva in media oltre 3,8 cm di spessore¹.

Encyclopaedia Britannica, ed. 1911 (voce 'China', sezione 'Encyclopaedias')

La sola versione inglese di Wikipedia comprende a metà 2023 oltre 6.600.000 articoli, con un totale di oltre 4,3 miliardi di parole. Calcolando una media di 5 caratteri a parola, si tratta di oltre 21 miliardi di caratteri (senza tener conto degli spazi e della punteggiatura)².

Wikipedia, voce 'Size of Wikipedia'

Il contenuto di Wikipedia costituiva circa il 3% del corpus di addestramento di GPT-3 (non abbiamo dati esatti relativi a GPT-4)³.

Tom B. Brown et al., *Language Models are Few-Shot Learners*

¹ *Encyclopaedia Britannica*, 11th ed. (1911), vol. 6, p. 230 (voce 'China', sezione 'Encyclopaedias'). La digitalizzazione è disponibile sull'Internet Archive: https://archive.org/details/Encyclopaediabrit06chisrich_201303/page/n257/mode/2up. Se la descrizione evoca Borges, la cosa non è casuale: cfr. Paolo Cherchi, *La "enciclopedia cinese" di Jorge Luis Borges*, in «Storie e linguaggi. Rivista di studi

umanistici – A Journal of the Humanities» 4, 2018, 2, pp. 103-115, in rete alla pagina https://www.libreriauniversitaria.it/download_free.php?code=2217341802073. Tutti i link nelle note di questo libro sono stati verificati l'ultima volta nel giugno 2023.

² Wikipedia, voce 'Size of Wikipedia': https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia, 3 luglio 2023.

³ Il dato è fornito nell'articolo che presenta il sistema scritto dai ricercatori di Open AI che l'hanno creato: Tom B. Brown et al., *Language Models are Few-Shot Learners*, in «arXiv.org», 22 luglio 2020, <https://arxiv.org/pdf/2005.14165.pdf>.

Introduzione

Nel 1960 – l’anno in cui sono nato – i computer erano macchinari enormi che potevano riempire una stanza delle dimensioni di una palestra, non erano collegati fra loro, e avevano capacità di calcolo e di elaborazione dati molto minori di quelle di uno smartphone. In quello stesso anno Ted Nelson – all’epoca studente dell’università di Harvard, in seguito uno dei pionieri della riflessione su come gli strumenti informatici e di rete possano modificare la gestione e la natura stessa delle nostre conoscenze – formulava la prima idea di un progetto che avrebbe pervicacemente (e senza troppo successo) cercato di realizzare nei sessant’anni successivi: Xanadu.

Il progetto di Nelson prendeva il nome dalla capitale estiva dell’impero mongolo di Kublai Khan, nota alla cultura occidentale attraverso una serie di mediazioni che comprendono *Il Milione* di Marco Polo e, soprattutto, la famosa poesia *Kubla Khan* di Coleridge. In questa successione di testi, Xanadu diventa progressivamente simbolo di complessità e perfezione architettonica; ma la Xanadu a cui pensa Nelson non è più un’architettura fisica, è un’architettura del sapere. La sua descrizione più suggestiva – inizialmente anche nella forma, a metà fra la stampata di un computer anni ’80 e un volantino ciclostilato: ve ne fate un’idea dalla scannerizzazione raggiungibile attraverso il QR-Code qui di seguito – è in *Literary Machines*, il libro ‘in progress’ di Nelson, pubblicato in nove edizioni diverse fra il 1981 e il 1994¹. Il secondo capitolo di quel libro si intitola «Proposta per un sistema di pubblicazione e archiviazione elettronica universale» e contiene alcune delle idee-chiave del progetto. Innanzitutto, l’idea che la letteratura sia «un sistema in continua evoluzione di documenti interconnessi»², dove il termine ‘letteratura’ copre, in senso ampio, qualunque forma di testualità: narrativa, poetica, saggistica o di altro tipo. Poi, l’idea che l’esplicitazione di quelle

connessioni attraverso una struttura ipertestuale (dobbiamo a Nelson l'introduzione del termine 'ipertesto') aiuti a farne comprendere il rilievo, e nel contempo a comprendere la natura relazionale e collaborativa del processo di costruzione del sapere, catturandone la complessità.



QR-Code 1

Il secondo capitolo di *Literary Machines*

<http://bit.ly/3CKEcRQ>

Il sistema ipertestuale a cui pensava Nelson è spesso visto come un'anticipazione del world wide web; un accostamento che lo stesso Nelson ha rifiutato, considerando invece il web come una sorta di 'brutta copia', piena di difetti, della sua idea originaria: il web prevede solo link unidirezionali, non permette di differenziarli per tipologia, non tiene traccia dei cambiamenti dell'informazione nel tempo, non garantisce la permanenza né dei link né dei dati archiviati, non dichiara esplicitamente autore e proprietario dei contenuti, si basa su una strutturazione dei contenuti su server in cartelle e sottocartelle che Nelson considera obsoleta... Il web, per Nelson, non è insomma altro che «un'altra imitazione della carta», che «tradisce la nostra idea originale di ipertesto»³.

Ted Nelson non realizzerà mai Xanadu se non in implementazioni parziali e problematiche, che ne consolideranno la fama di primo e più famoso 'vaporware': un software presentato con grandi promesse e aspettative, alle quali non segue però nulla di concreto. Ma Xanadu resta una delle più ambiziose manifestazioni di un'idea alla base di molta parte del lavoro della prima generazione di evangelisti dell'informatica personale e della rete: costruire le nuove infrastrutture della conoscenza. Infrastrutture capaci di ospitare e rendere accessibile l'insieme del sapere umano del passato, evidenziandone la natura relazionale e interconnessa,

e nel contempo di favorire la costruzione del sapere futuro, sfruttando al meglio le possibilità inedite offerte dall'ecosistema digitale.

Nel mio libro precedente, *L'età della frammentazione*⁴, ho contrapposto modelli architettonici di organizzazione delle informazioni come Xanadu (mai realizzato) o Wikipedia (realizzata e popolarissima) – molto ambiziosi e nati con l'aspirazione di costruire in digitale strumenti di gestione delle conoscenze migliori e capaci di maggiore complessità strutturale rispetto a quelli tradizionali – con la realtà rappresentata dalla grande diffusione in rete di contenuti brevi, granulari, frammentati, legati in particolare al mondo dei social network. In quella sede mi sono interrogato sulle ragioni e sulle caratteristiche di questa prevalenza della frammentazione sulla complessità, e sui problemi che essa pone soprattutto a livello di formazione scolastica e nella realizzazione di contenuti di apprendimento.

Questo libro intende proseguire quella ricerca da un punto di vista diverso, ponendo al centro dell'attenzione non più l'età della frammentazione e i suoi effetti sul sistema formativo, ma alcuni fra i principali modelli per la creazione di contenuti digitali complessi e le loro conseguenze rispetto alla costruzione – e alla natura stessa – dell'edificio del sapere. La scelta è legata alla convinzione, che nel libro precedente avevo cercato di argomentare, della necessità di riconquistare complessità anche nell'ecosistema digitale: un'evoluzione che credo corrisponda a linee di tendenza abbastanza chiaramente individuabili negli strumenti che sono stati e vengono sviluppati in questi ultimi anni.

In particolare, vorrei soffermarmi su due modelli – apparentemente alternativi, ma forse integrabili – di organizzazione complessa delle conoscenze: quello architettonico, basato sull'idea di un sapere fortemente strutturato, organizzato in base a ontologie e gerarchie solide e ben fondate, e quello organico, che vede l'insieme delle nostre conoscenze come una sorta di organismo vivente capace di modificarsi e crescere in forme non sempre ben definite e comprensibili, non sempre rigorosamente analizzabili, in cui i livelli superiori possono essere caratterizzati da proprietà emergenti difficilmente riducibili alla pura interazione meccanica delle componenti di livello inferiore (anche perché

le componenti da considerare sarebbero troppe, e le loro interazioni troppo numerose e in parte sconosciute).

Il primo modello è sostanzialmente quello della tradizione enciclopedica: l'enciclopedia rappresenta l'esempio paradigmatico della concezione architettonica del sapere, e l'enciclopedismo digitale ha cercato di aggiornare questa tradizione sfruttando le nuove armi offerte dalla multicodicalità, dall'ipertestualità, dall'interattività, nonché – sul piano della descrizione e categorizzazione dei contenuti – dalle ontologie e dai metadati, al centro di progetti come quelli del web semantico e dei linked data (su cui tornerò). Wikipedia ne è la manifestazione più nota, e la sua evoluzione nel tempo ci dice molto anche sull'evoluzione e sulle potenzialità future di questo modello. Come vedremo, al modello enciclopedico si collega spesso anche quello rappresentato da un'altra delle tradizionali cattedrali del sapere, la biblioteca, che però – soprattutto nelle sue manifestazioni più recenti – assorbe anche alcuni elementi del modello organico, come ben testimoniato dalla famosa 'quinta legge' della biblioteconomia proposta da Ranganathan: «La biblioteca è un organismo che cresce»⁵.

Il secondo modello ha invece il suo esempio paradigmatico nelle reti neurali, che hanno oggi portato all'impetuoso sviluppo delle intelligenze artificiali generative come ChatGPT. Le intelligenze artificiali generative sono state spesso paragonate a oracoli, basati su associazioni statistico-probabilistiche più che su algoritmi logico-deterministici, e questa contrapposizione fra modello architettonico-deterministico e modello oracolare-probabilistico è alla radice anche del titolo del libro che avete in mano. Titolo che – per le lettrici e i lettori più nerd – gioca anche sulla contrapposizione fra l'architetto e l'oracolo nei film della serie *Matrix*, in cui i due termini sono usati a proposito dei due software di intelligenza artificiale, per molti versi contrapposti ma in ultima analisi costretti alla collaborazione, che Neo, il protagonista, scopre essere alla base della costruzione della 'matrice'.

Questa distinzione è anche alla base della struttura che ho dato al libro. Le prime due parti – dopo un capitolo introduttivo che sintetizza rapidamente alcune delle conclusioni de *L'età della frammentazione*, in modo da rendere il testo pienamente autonomo e comprensibile anche a

chi non avesse sottomano quel lavoro – esplorano infatti rispettivamente la storia e l'evoluzione, fino ai risultati attuali, dell'enciclopedismo digitale, considerato appunto come esempio per eccellenza del paradigma architettonico, e delle reti neurali, alla base dei nuovi risultati nel campo dell'intelligenza artificiale generativa.

La parte terza del libro mette alla prova le due impostazioni attraverso lo studio di un caso che mi sembra particolarmente interessante, da molti punti di vista: quello delle memorie personali e della loro progressiva 'esternalizzazione' digitale. Come vedremo, si tratta di un campo che offre non solo un'occasione di confronto fra prospettive diverse ma anche un banco di prova delle loro possibili convergenze, oltre a uno sguardo suggestivo e in parte inquietante su possibili sviluppi futuri.

La parte quarta guarda al futuro, e lo fa in una prospettiva particolare e forse poco abituale rispetto al mio ambito specifico di competenze e di studio: quella della fantascienza. Oltre a permettermi di esplorare un campo per il quale, cresciuto negli anni delle missioni Apollo e della corsa allo spazio, ho sempre avuto una passione particolare, la prospettiva offerta dalla letteratura di fantascienza mi è sembrata adatta a offrire un punto di vista meno tradizionalmente 'accademico' sui temi affrontati nelle altre parti del libro. Inoltre, la possibilità di unire la dimensione della 'narrativa di anticipazione', che propone, esamina e mette alla prova possibili sviluppi futuri non solo in ambito tecnologico ma anche sociale e culturale, e quella di espressione dell'immaginario attuale, che inevitabilmente si riflette e si proietta sul futuro costruito e immaginato dalla narrazione, ben corrisponde a un momento nel quale l'analisi dei cambiamenti in corso sembra richiedere categorie e strumenti interpretativi in parte nuovi rispetto al passato.

Spero che l'insieme di queste scelte tematiche permetta al libro che avete in mano di risultare accessibile anche a lettrici e lettori non specialisti: ho sempre cercato di unire, nei miei lavori, la riflessione e l'esigenza di rigore proprie della ricerca universitaria con il tentativo di raggiungere un pubblico più vasto di quello strettamente accademico, nella convinzione che le tematiche legate al rapporto fra cultura e digitale siano di interesse generale e legate a cambiamenti culturali, sociali ed economici di grande portata, che toccano direttamente tutte e tutti noi. Una considerazione particolarmente rilevante nel prendere in esame gli strumenti usati per

raccogliere, organizzare e condividere il sapere: si tratta infatti di strumenti che costituiscono la struttura portante, potremmo dire l'ossatura, del nostro orizzonte culturale.

Questa esigenza mi ha spinto anche a quattro scelte che è bene esplicitare subito. In primo luogo, quella di organizzare il testo in capitoli piuttosto brevi, affidando invece all'articolazione in parti il compito di fornirne una suddivisione di più alto livello: è una scelta che avevo adottato anche per il libro precedente e mi pare possa facilitare la lettura, anche per la migliore identificazione e fruizione dei contenuti, più specificamente individuati. In secondo luogo, quella di tradurre in italiano tutte le citazioni (i testi originali sono sempre indicati in nota, e la maggior parte di essi sono ormai facilmente reperibili anche via web, per chi volesse verificare la citazione nella lingua originale); dove il riferimento in nota non riporta l'origine della traduzione, le traduzioni sono mie. In terzo luogo, quella di differenziare – in maniera ancor più netta di quanto già abituale in lavori di questo tipo – il corpo del testo, pensato per una lettura più accessibile, e l'apparato di note, decisamente nutrito. La comprensione delle tesi e della linea argomentativa del libro non richiede la consultazione delle note, pensate invece, oltre che per offrire riferimenti puntuali alle fonti, come occasione per approfondimenti e rimandi più specifici e dettagliati. Dato che il corpo delle note è più piccolo, questo rende anche più agile (e meno caro) il libro. Infine, quella di applicare una distinzione analoga ad alcune delle risorse di rete citate, che ho diviso in due grandi categorie: quelle semplicemente richiamate o utili come possibili approfondimenti, citate in nota attraverso il loro indirizzo web, e quelle che invece andrebbero considerate come integrazioni di particolare interesse o funzionali alla migliore comprensione del testo, quasi come si trattasse di illustrazioni. Per queste ultime ho usato – oltre all'indirizzo web, comunque presente – dei codici QR inseriti nel corpo del testo, che, se inquadrati attraverso la fotocamera di uno smartphone, permettono l'accesso diretto al contenuto. Ne avete già trovato uno in questa introduzione e un altro – che rimanda a una pagina i cui contenuti saranno aggiornati nel tempo anche per dar conto di novità successive alla pubblicazione del libro – in quarta di copertina.

Spetterà al lettore giudicare se e in quale misura il tentativo potrà dirsi

riuscito. Dal canto mio – ferma restando la mia esclusiva responsabilità per le scelte fatte (non sempre allineate e a volte decisamente divergenti rispetto alle opinioni e ai consigli ricevuti) e per gli errori che inevitabilmente saranno rimasti – devo ringraziare almeno alcune fra le molte persone che attraverso osservazioni, suggerimenti, commenti sui temi discussi e in alcuni casi su alcune parti del testo o sui loro materiali preparatori, hanno permesso a questo lavoro di svilupparsi e di assumere la sua forma attuale: Simone Arcagni, Giulio Blasi, Emmanuela Carbé, Roberto Casati, Fabio Ciotti, Fabio Ciracì, Lauro Colasanti, Cesare Cozzo, Mario De Caro, Derrick De Kerckhove, Chiara Faggiolani, Francesco Farina, Grazia Farina, Riccardo Fedriga, Mauro Guerrini, Maurizio Lana, Francesco Leonetti, Federico Meschini, Rossana Morriello, Teresa Numerico, Giorgio Parisi, Giovanni Rizzoni, Alessandro Roncaglia, Matteo Sanfilippo, Giovanni Solimine, Pete Stevens, Maurizio Vivarelli, i reviewer anonimi degli articoli in cui ho affrontato temi ripresi in questa sede, e tutte le partecipanti e i partecipanti della lista ‘Il grillo parlante’. Scusandomi con le moltissime altre persone che avrei pure dovuto menzionare: ricordare tutte e tutti supera sia le mie capacità di memoria, non ancora adeguatamente supportate dagli strumenti di cui parleremo nella parte terza del libro, sia i limiti di lunghezza di questa Introduzione.

La parte seconda del libro, e occasionalmente anche le altre, hanno comportato lunghe interazioni con ChatGPT (nelle versioni 3.5 e 4) e con altri Large Language Models (LLM); la questione del possibile coinvolgimento ‘autorale’ delle intelligenze artificiali generative è oggi molto discussa, ma nel caso specifico rappresentato da questo libro non si tratta in alcun modo di una co-authorship, neanche parziale: il testo del libro è interamente mio, i passi in cui ho fatto riferimento al testo prodotto da ChatGPT sono esplicitamente segnalati e le citazioni sono sempre virgolettate o in corpo rientrato. A una intelligenza artificiale generativa – Midjourney – si deve invece l’immagine di copertina del libro: il relativo prompt è riportato nei credits, e il suo progressivo affinamento (fino a ottenere un’immagine soddisfacente sia per me sia per la responsabile grafica della casa editrice, che ringrazio per la lunga interazione al riguardo e il parere sulle tante immagini alternative prodotte in questo processo) è stato fatto da me.

Infine, buona parte del libro è stata scritta ai tavolini del chiosco Mediterranea di Capocotta, i cui gestori, dovendo scegliere se cacciarmi per occupazione abusiva e prolungata di spazi pubblici o adottarmi come caso disperato di compulsione lavorativa, hanno optato per la seconda ipotesi.

Alcuni capitoli di questo libro riprendono, con modifiche, integrazioni e aggiornamenti, testi già comparsi in articoli e interventi scritti negli ultimi anni. In questi casi ho sempre ricordato in nota, all'inizio delle relative parti, i rispettivi riferimenti bibliografici.

¹ Theodor H. Nelson, *Literary Machines*, pubblicato dall'autore attraverso la sua casa editrice personale Mindful Press, Sausalito 1981-1994. L'edizione che utilizzo in questa sede è quella del 1987, disponibile su Internet Archive all'indirizzo <https://archive.org/details/literarymachines00nels/>. Per chi voglia farsi un'idea della costruzione grafica delle prime edizioni del testo, nel momento in cui scrivo un PDF del secondo capitolo, tratto dalla terza edizione, è disponibile all'indirizzo <https://cs.brown.edu/people/nmeyrowi/LiteraryMachinesChapter2.pdf>.

² Theodor H. Nelson, *Literary Machines* cit., p. 2/9.

³ Così la dichiarazione che ha aperto per anni la pagina 'Project Xanadu': cfr. ad es. <https://web.archive.org/web/20090215000911/http://xanadu.com/>. Nel 2012 la dichiarazione esplicitamente polemica contro il web è stata sostituita da una formulazione più neutra, ma la posizione di Nelson non è cambiata, e resta quella da lui espressa in forma ancor più radicale nella brevissima nota «I Don't Buy in»: <http://ted.hyperland.com/buyin.txt>. Si veda anche Theodor H. Nelson, *Geeks Bearing Gifts v. 1.1*, Mindful Press, Sausalito 2009. Per un'analisi delle idee di Ted Nelson si veda Douglas R. Dechow e Daniele C. Struppa, *Intertwined. The Work and Influence of Ted Nelson*, SpringerOpen, Berlin 2015, disponibile in accesso aperto all'indirizzo <https://link.springer.com/book/10.1007/978-3-319-16925-5>.

⁴ Gino Roncaglia, *L'età della frammentazione. Cultura del libro e scuola digitale*, seconda edizione accresciuta, Laterza, Bari-Roma 2020.

⁵ Shiyali Ramamrita Ranganathan, *The Five Laws of Library Science*, Madras Library Association, Chennai 1931.

Parte I

1.

Frammentazione e complessità (una sintesi)

Molta parte del mondo in cui viviamo è stata trasformata, spesso radicalmente, dagli effetti della rivoluzione digitale. La lista delle abitudini, delle pratiche lavorative e comunicative, degli strumenti oggi di uso quotidiano che sarebbero stati totalmente incomprensibili anche solo trenta o quarant'anni fa, sarebbe lunghissima. E la velocità dei cambiamenti in corso suggerisce che lo stesso potrebbe accadere nel prossimo futuro: fra qualche decennio uno smartphone potrebbe suscitare nelle nuove generazioni la stessa curiosa meraviglia con cui le bambine o i bambini di oggi guardano un telefono a disco. Eppure, nonostante la velocità dell'innovazione, la rivoluzione digitale è un fenomeno articolato, che ha conosciuto fasi diverse e il cui sviluppo non è stato e non è necessariamente lineare. Le prospettive da cui possiamo analizzarlo sono molteplici, e non tutte compatibili fra loro.

Alcune di queste prospettive sembrano effettivamente delineare una direzione di sviluppo ben definita: è il caso, ad esempio, della tendenza alla progressiva miniaturizzazione dei dispositivi e di quelle, parallele, all'aumento della loro potenza di calcolo e alla loro progressiva interconnessione. Se adottiamo questa prospettiva possiamo forse proporre una periodizzazione che (1) parte dai grandi 'mainframe' degli anni successivi alla Seconda guerra mondiale, le cui singole unità di lavoro erano grandi come un armadio: computer basati prima su valvole e poi su transistor, utilizzati soprattutto come macchine per il calcolo numerico; (2) prosegue con gli home e personal computer degli anni '80 – risultato dell'invenzione del microprocessore – che hanno trasformato il computer in un oggetto familiare sia a casa sia nel luogo di lavoro e dotato di limitate capacità di connessione, via accoppiatore acustico o via

modem, ai primi servizi di rete; (3) vede poi la fase dell'esplosione di Internet, legata all'introduzione del world wide web intorno al 1993 e alla sua rapida diffusione nel decennio successivo; e (4) arriva all'era dell'Internet mobile in banda larga e degli smartphone sempre connessi. Il futuro di questa linea di sviluppo – non facile da prevedere – potrebbe portare all'ulteriore diffusione di interfacce di 'human augmentation', di ambienti intelligenti, di sistemi di realtà aumentata e mista, in un contesto segnato dal crescente rilievo dei sistemi di intelligenza artificiale, sempre più spesso capaci di guidare anche il comportamento di dispositivi fisici.

Ma questa prospettiva non è l'unica possibile. Anziché partire dall'evoluzione dell'hardware e delle infrastrutture potremmo guardare all'evoluzione sul versante del software, con il passaggio dai sistemi operativi a caratteri a quelli a grafici¹, dai videogiochi che hanno contribuito alla diffusione degli home computer ai primi pacchetti di applicazioni per il mondo del lavoro, per arrivare ai software di navigazione online, che hanno trasformato la rete in un ambiente comunicativo, e poi ai social network, fino agli strumenti di videoconferenza che hanno accompagnato, sia per lo studio sia per il lavoro, il periodo della pandemia e che sono chiaramente destinati ad avere un ruolo rilevante anche in futuro. O, ancora, potremmo guardare all'evoluzione della cornice ideologica ed economica che ha accompagnato la rivoluzione digitale, partendo dalla cosiddetta 'ideologia della Silicon Valley', curioso mix tutto statunitense di anarco-individualismo anti-statalista con sfumature di destra e di cultura hippy, di Ayn Rand e Robert Heinlein, di Dungeons & Dragons e radioelettronica, che ha creato negli anni '70 l'ambiente culturale propizio alla nascita del personal computer come strumento di 'personal empowerment', per arrivare al suo capovolgimento nel turbocapitalismo neoliberista dei colossi Big Tech, in cui il controllo e lo sfruttamento economico dei dati – inclusi i dati personali – ha completamente sostituito la precedente rivendicazione dell'intangibilità della sfera personale da ogni intromissione esterna.

Ognuna di queste dimensioni – e delle molte altre che potrebbero essere individuate – offre spunti e motivi di riflessione capaci di rivelare qualcosa

del quadro d'insieme, ma non di esaurirlo o di renderlo perfettamente definito e coerente.

Nel libro *L'età della frammentazione* avevo provato ad avanzare una proposta interpretativa ulteriore e in parte diversa, basata sul paragone – sicuramente non privo di qualche rischio di artificialità – fra l'evoluzione della rete e l'evoluzione delle società umane. Un paragone che si concentrava in sostanza su due polarità: quella fra scarsità e abbondanza di risorse informative, e quella fra frammentazione e complessità dei relativi contenuti. La riassumo qui rapidamente, perché – in particolare per chi non avesse letto o non avesse presente quel lavoro – può aiutare a capire le ragioni e il contesto dell'attenzione alle diverse forme di complessità che costituisce la prospettiva specifica del libro che state leggendo. Chi preferisse può anche far riferimento alla sintesi di questa ricostruzione proposta in una serie di quattro lezioni disponibili su YouTube, e in particolare alla quarta di esse, visualizzabile attraverso il QR-Code qui di seguito.



QR-Code 2

Videolezione: alcune tappe della rivoluzione digitale

<http://bit.ly/44scK7l>

Come punto di partenza del paragone, possiamo pensare alla fase iniziale di diffusione della rete, fra gli anni '80 e i primissimi anni '90, con l'intreccio fra servizi amatoriali privati e prime esplorazioni delle possibilità offerte da Internet, come a qualcosa di simile all'età dei cacciatori-raccoglitori. Piccole tribù di nerd, armate dei loro home computer, si affacciavano nel nuovo spazio virtuale per brevi scorribande – quelle consentite dalla lentezza e precarietà della connessione via modem – alla ricerca, abbastanza casuale, dei contenuti disponibili:

contenuti raccolti e consumati poi nella ‘caverna’ del proprio computer, una volta staccata la connessione.

La nascita e la diffusione del web, così come il diffondersi di connessioni più stabili alla rete, può poi essere paragonata al passaggio dall’epoca dei cacciatori-raccoglitori a quella dei primi insediamenti urbani e dell’agricoltura. Un sito web (e il ricorso al termine ‘sito’ non è casuale) è così di fatto un insediamento informativo, luogo di coltivazione organizzata di contenuti che progressivamente crescono di numero, e insieme punto di attrazione e convergenza per gli utenti. La scelta del nome ‘Geocities’ – e l’uso della metafora dei diversi centri urbani corrispondenti a tipologie diverse di contenuti –, fatta da uno dei primi servizi per la creazione di siti e pagine personali, ben coglie le caratteristiche di questo periodo, che corrisponde a grandi linee al decennio che va dal 1993 fino ai primissimi anni 2000.

Con il cosiddetto ‘web 2.0’ lo scenario cambia: si moltiplicano i contenuti generati dagli utenti (il paradigma dello ‘user generated content’), che non sono più necessariamente ancorati a un singolo sito web ma possono circolare, rimbalzare da un sito all’altro, venire scambiati, diventare virali. I social network rappresentano un nuovo modello di produzione e diffusione capillare di informazioni granulari e a bassa complessità: si tratta di quella che, proprio pensando alla natura in fondo artigianale (per quanto mediata da strumenti di produzione e di editing sempre più sofisticati) del contenuto generato dagli utenti sui social, possiamo immaginare come l’età dell’artigianato e del commercio informativo.

Il passaggio dalla relativa scarsità di risorse che caratterizzava l’età dei cacciatori-raccoglitori alla loro abbondanza o sovrabbondanza nell’età dell’artigianato e del commercio non ha dunque modificato, almeno finora, il loro carattere prevalentemente frammentato e granulare. Le forme di testualità di rete – dalle mail ai messaggi WhatsApp, dai post di un blog ai messaggi di stato sui social network, dai tweet alle pagine di un sito web – sono quasi tutte brevi, sia per la velocità che ne caratterizza la scrittura, la trasmissione, la ricezione, l’interpretazione, sia per il fatto di essere, appunto, contenuti ‘artigianali’ e autoprodotti. Ovviamente, forme di testualità breve abbondano anche nella tradizione pre-digitale, ma nel mondo digitale la loro presenza sembra preponderante; del resto,

osserviamo la stessa situazione anche se dai testi scritti ci spostiamo, ad esempio, ai contenuti video (basti pensare alle storie Instagram o TikTok o ai video su YouTube) o audio (in musica, l'epoca dei 'concept album' sembra tramontata a favore dei brani singoli).

Certo, la complessità non implica necessariamente qualità: una teoria del complotto può essere anche molto complessa, pur essendo irrazionale o infondata. E d'altro canto la frammentazione non implica necessariamente scarsa qualità o scarsa rilevanza: alcuni contenuti granulari possono essere preziosi. Ma la progressiva tendenza all'aumento di contenuti brevi e alla diminuzione di contenuti complessi – e la conseguente perdita di competenze legate alla complessità – è comunque un fattore che dovrebbe preoccuparci: il nostro processo di costruzione di conoscenze e di interpretazione del mondo richiede infatti necessariamente la capacità di produrre e lavorare con contenuti complessi, di raccogliere e organizzare dati, di formulare teorie. E questa esigenza cresce con l'aumentare dei fattori e delle dimensioni in gioco, come accade in società sempre più interconnesse. Se l'esplosione di contenuti granulari pregiudica le competenze e l'abitudine alla complessità, la nostra capacità di capire il mondo intorno a noi diminuisce.

Cosa spiega questa prevalenza della granularità e della frammentazione sulla complessità? La tentazione più immediata (una tentazione in cui sono caduti in molti) è quella di considerarla come caratteristica in qualche misura 'essenziale' dell'ecosistema digitale. A ben vedere, però, questa ipotesi non funziona: il digitale di per sé è solo un formato di codifica dell'informazione, ormai abbastanza indifferente rispetto alla lunghezza dei contenuti. Per i dispositivi di cui disponiamo oggi, un testo di mille pagine non è molto più complicato da convertire in bit di un tweet di pochi caratteri, e un video di tre ore, pur se certo più pesante di uno di pochi secondi, è comunque perfettamente gestibile. La tendenza alla granularità mi sembra semmai legata alla situazione attuale e contingente dell'ecosistema digitale, e in particolare all'età dell'artigianato e del commercio informativo, più che a una sua caratteristica necessaria e permanente. E d'altro canto, come abbiamo già visto parlando di Ted Nelson e come vedremo ancor meglio nei prossimi capitoli, l'attenzione

verso contenuti complessi – pur se minoritaria – non è mai sparita dall'orizzonte delle culture digitali.

Se queste ipotesi interpretative sono corrette, l'evoluzione che ci aspetta potrebbe essere quella di una progressiva espansione di questi nuclei ancora embrionali di contenuti digitali complessi, producendo una tendenza alla riconquista della complessità. Anche in digitale, l'età dell'artigianato e del commercio potrebbe cedere il passo all'età delle cattedrali, e dunque alla costruzione collaborativa di architetture del sapere più sofisticate e meglio strutturate. Sulla scia delle prime cattedrali informative che abbiamo già cominciato a costruire: l'esempio più chiaro, e su cui mi soffermerò a lungo più avanti, è quello di Wikipedia².

Ne *L'età della frammentazione* avevo discusso le conseguenze di questa dialettica fra frammentazione e complessità con particolare riferimento al sistema formativo e alle risorse di apprendimento. La tesi di fondo che ho cercato di sostenere in quella sede è che se l'ipotesi di un progressivo recupero della complessità in ambiente digitale è corretta, ciò configura uno specifico bisogno formativo, legato alle competenze di complessità, trasversale rispetto a ordini e gradi scolastici e rispetto alle diverse discipline. In altri termini: il nostro sistema formativo deve orientare le sue politiche scolastiche e universitarie verso la necessità prioritaria di fornire a studentesse e studenti la capacità di produrre, comprendere, valutare, selezionare, conservare, diffondere contenuti strutturati e complessi, e non solo contenuti granulari e frammentati, e di farlo anche e soprattutto all'interno del nuovo ecosistema digitale, guardando alle forme di complessità che ci aspettano nel futuro e non solo a quelle ereditate dal passato. Mi sono poi soffermato, in particolare, sulle conseguenze di questo sviluppo per la forma-libro, che nella nostra tradizione culturale costituisce il principale paradigma di riferimento per la costruzione di informazione strutturata e complessa.

Le osservazioni critiche che mi sono state fatte più spesso rispetto a queste tesi sono due: l'assenza di indizi sul fatto che in rete si vada effettivamente verso un recupero di contenuti complessi, e la (conseguente) impressione che la mia impostazione possa peccare di eccessivo ottimismo.

In un certo senso, questo libro cerca di rispondere a tutte e due queste

critiche. Alla prima, esaminando alcuni fra i campi in cui il lavoro di costruzione di contenuti complessi sta avendo successo, favorendo il progressivo sviluppo di strumenti più potenti e sofisticati: l'ambito enciclopedico, che rappresenta un po' l'idea paradigmatica di architettura della conoscenza; quello legato alla nascita delle intelligenze artificiali generative, che – come ho già accennato nell'Introduzione – sembra orientato a un diverso tipo di complessità, ma si muove comunque in quella direzione; infine, quello della conservazione delle memorie (nel caso che prenderò in esame, delle memorie personali), che trasforma in complessità, in forme sempre più sofisticate, una pluralità di contenuti granulari (testi, immagini, video, dati personali...).

È bene osservare, peraltro, che la tendenza a un recupero della complessità si nota – credo – anche da indizi più minuti. Abbiamo osservato poc'anzi come YouTube, Instagram, TikTok siano esempi abbastanza evidenti della granularizzazione anche di contenuti non testuali. Ebbene, proprio queste piattaforme mostrano oggi segni di una inversione di direzione. Ne è un primo esempio la crescita, osservata ormai da diversi anni, della durata media dei video su YouTube³. Certo, questo dato potrebbe essere una naturale reazione proprio alla nascita di Instagram e di TikTok, programmaticamente orientati alle storie video brevi o brevissime... se non fosse che anche la durata media delle storie Instagram e TikTok sta aumentando: Instagram ha portato la durata delle sue storie da 15 a 60 secondi, e TikTok ha alzato l'iniziale limite di 15 secondi prima a 60 secondi, poi a 3 minuti, e nel 2022 a 10 minuti, tanto da far parlare di 'YouTube-ification' della piattaforma⁴. La crescita nel consumo di audiolibri e di podcast, a loro volta spesso seriali, mostra come anche nel caso dell'audio i contenuti più lunghi sembrano funzionare man mano meglio (mentre la tendenza alla diminuzione della durata media sembra proseguire sui brani musicali, che però per un verso hanno meccanismi di ritorno economico che favoriscono i brani più brevi, per altro verso vengono aggregati in playlist man mano più lunghe). Per quanto riguarda poi la scrittura, sullo sviluppo dei contenuti 'long form' in ambito giornalistico e di rete – spesso associati a forme sempre più sofisticate di visualizzazione dei dati – esiste una copiosa letteratura (che anche in questo caso ha creato nuovi neologismi, come

‘scrollytelling’⁵). È possibile che un’ulteriore manifestazione di questa stessa tendenza di fondo sia alla base di alcuni dati almeno in parte incoraggianti emersi dopo la pandemia, e in particolare negli ultimi mesi, anche su uno degli indicatori più importanti della nostra frequentazione di contenuti strutturati e complessi: quello relativo alla propensione alla lettura di libri (indipendentemente dalla scelta di leggere su carta o in digitale)⁶.

Non intendo affermare, si badi, che la rete si stia trasformando da un giorno all’altro da regno dei social veloci in spazio dominato da contenuti ponderosi e meditati: velocità e brevità restano comunque caratteristiche distintive dell’ecosistema comunicativo attuale. Vorrei sostenere piuttosto che la tendenza alla frammentazione, esasperata dai meccanismi di funzionamento dei social network negli anni ’10, non è necessariamente irreversibile e potrebbe anzi non corrispondere alle prospettive di sviluppo che ci aspettano negli anni ’20, e che in parte sono già in atto.

D’altro canto, ed è qui la mia risposta alla seconda fra le critiche che ho sopra ricordato, credo occorra guardarsi anche da un rischio per qualche verso speculare: quello di considerare come necessariamente ‘ottimista’ la prospettiva di un aumento dei contenuti complessi e strutturati e – in generale – di un aumento della complessità informativa. Per un verso, come ho accennato, la complessità non implica necessariamente qualità: una crescita nella produzione di contenuti complessi non accompagnata da adeguate competenze di comprensione e valutazione può risultare addirittura pericolosa. Per altro verso, se all’aumento della complessità non corrisponde una crescita il più possibile diffusa delle competenze necessarie a gestirla, la complessità rischia di diventare un fattore di esclusione anziché un fattore di sviluppo. E si tratterebbe di una direzione che non possiamo davvero permetterci di seguire.

Spero che queste considerazioni possano essere tenute presenti nella lettura dei capitoli che seguono: per capovolgere un detto abbastanza diffuso, non vorrei che, nei già pochi casi in cui la ragione trova qualche motivo di ottimismo, fosse poi la volontà – e dunque le scelte fatte in materia di politica culturale e formativa – a mettersi di traverso.

¹ Da non perdere, su questo tema, il bel saggio di Neal Stephenson *In the Beginning... Was the Command Line*, William Morrow, New York 1999. Su Stephenson – questa volta in veste di autore di uno dei più famosi libri di fantascienza dedicati al mondo del digitale e delle reti, *Snow Crash* – torneremo nell’ultima parte di questo libro.

² La metafora della cattedrale è qui usata in senso abbastanza diverso da quello che le attribuisce Eric Steven Raymond in *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O’Reilly Media, Sebastopol (CA) 1999, trad. it. *La cattedrale e il bazaar. Riflessioni di un rivoluzionario per caso su Linux e l’open source*, Apogeo, Milano 2022, anche se molte delle considerazioni contenute nel libro sono interessanti anche dal mio punto di vista.

³ Si veda ad es. Julia Alexander, *YouTube videos keep getting longer*, in «The Verge», 26 luglio 2019, <https://www.theverge.com/2019/7/26/8888003/youtube-video-length-contrapoints-lindsay-ellis-shelby-church-ad-revenue>.

⁴ Sarah Roach, *The YouTube-ification of TikTok is almost complete*, in «protocol», 1° marzo 2022, <https://www.protocol.com/entertainment/tiktok-longer-videos-ads>.

⁵ Cfr. Doris Seyser e Michael Zeiller, *Scrollytelling – An Analysis of Visual Storytelling in Online Journalism*, in *22nd International Conference Information Visualisation (IV)*, IEEE, Fisciano 2018, pp. 401-406, <https://doi.org/10.1109/iV.2018.00075>.

⁶ Mi riferisco ad esempio, per quel che riguarda l’Italia, ai dati presentati nell’incontro su ‘La lettura in digitale’ organizzato il 20 giugno 2023 da MediaLibrary On Line in collaborazione con AIE e CEPELL: ne è disponibile la registrazione video all’indirizzo <https://www.youtube.com/watch?v=WaV1oxNZt8g>. Si tratta di un tema che non affronto in questa sede, ma su cui spero di poter tornare altrove.

2.

Dall'enciclopedia mondiale alle enciclopedie digitali¹

Herbert George Wells è oggi ricordato soprattutto come romanziere, ma a lui dobbiamo anche decine di saggi, su moltissime tematiche diverse. Uno dei suoi argomenti preferiti, tanto nei romanzi quanto nei saggi, è il futuro. E se il Wells romanziere guarda al futuro intrecciando abilmente anticipazione scientifica, utopia e distopia, il Wells saggista lo considera soprattutto nelle sue dimensioni politiche, culturali e sociali. Nonostante le sue posizioni filosofiche deterministiche, una delle sue preoccupazioni principali è quella di individuare strategie per la costruzione di un futuro politicamente, socialmente, culturalmente migliore: un futuro senza guerre, guidato da un governo mondiale, capace di riconoscere il valore unificante della scienza, di garantire istruzione e formazione di qualità (con l'adozione di modelli pedagogici che oggi definiremmo costruttivisti), di promuovere il progresso del sapere.

Tra le opere dedicate a questi temi, ne troviamo una dal titolo curioso: *World Brain*². Il 'cervello mondiale', per Wells, corrisponde a una sorta di raccolta organizzata e sistematizzata del nostro sapere, che – resa universalmente accessibile attraverso le nuove tecnologie (Wells pensa ai microfilm, ma è chiaro che la rivoluzione digitale ha fornito a questo sogno strumenti ben più potenti) – possa funzionare anche come fondamento per la risoluzione di conflitti e la collaborazione universale.

Wells non considera il cervello mondiale solo come un'idea astratta: propone una strada ben definita che ne dovrebbe garantire la realizzazione pratica. E questa strada è rappresentata dalla costruzione di un'enciclopedia mondiale, strumento insieme conoscitivo e politico. A suo avviso il progetto di enciclopedia, nucleo fondante del cervello

mondiale, dovrebbe essere il contraltare ‘conoscitivo’ della Società delle Nazioni, che da sola non può funzionare proprio per l’assenza di un adeguato fondamento epistemico: la politica ha bisogno di basare le proprie scelte su conoscenze e competenze ben organizzate.

Wells si fa così interprete di quella che è stata per molta parte degli intellettuali europei, nella breve ma intensa parentesi fra le due guerre mondiali, una ricerca quasi disperata di soluzioni capaci di offrire stabilità politica e di evitare lo spettro di nuovi conflitti (una frase del libro si rivela al riguardo quasi profetica: «Viviamo in quelli che gli storici del futuro chiameranno ‘gli spauriti anni Trenta’»³). Alla base della sua proposta è l’idea di estendere all’intero edificio del sapere le considerazioni fatte da John Maynard Keynes in *Le conseguenze economiche della pace*: la politica non funziona se non è supportata da conoscenze e competenze solide. I politici, nella sintesi offerta da Wells delle tesi di Keynes, «sono così poco avvezzi all’utilizzo del pensiero, così ignari dell’esistenza del sapere e di cosa il sapere sia, che non ne capiscono l’importanza»⁴.

L’Enciclopedia Mondiale immaginata da Wells dovrebbe contrastare questa insipienza, aiutando a costruire complessità e struttura a partire dall’enorme ma frammentata massa di conoscenze disponibili al ‘nuovo mondo’ che esce dalla catastrofe del primo conflitto mondiale:

La mia idea di fondo è che per comporre il mosaico eterogeneo cui accennavo sopra, per rimediare alla dispersione e all’inefficacia delle nostre forze cognitive e coinvolgerle in un comune orizzonte di senso, [...] occorre un qualcosa – un nuovo organo sociale, una nuova istituzione – che per il momento chiamerò Enciclopedia Mondiale [...], il necessario atto di congiunzione fra le nostre risorse mentali e le nostre forze politiche. [...] [Nei volumi dell’Enciclopedia Mondiale, chiunque potrebbe trovare] senza alcuna fatica, redatti in un linguaggio chiaro e immediato e tenuti costantemente aggiornati, i principi fondamentali del nostro ordine sociale, i concetti portanti e i dettagli più significativi di ogni campo del sapere, un quadro esatto e sufficientemente particolareggiato dell’universo, un compendio della storia mondiale, e qualora il nostro comune cittadino volesse studiare a fondo un determinato tema, una bibliografia esaustiva e affidabile lo rimanderebbe alle fonti primarie. Se in una certa disciplina dovessero sommarsi più metodi e teorie diverse, il lettore non si troverebbe a sfogliare raccolte di opinioni ammucciate alla rinfusa, ma summe di idee e di spiegazioni attentamente scelte e ordinate. [...] Una tale Enciclopedia svolgerebbe nella nostra cultura laica il ruolo di una Bibbia antidogmatica. [...] Farebbe del sapere il collante che tiene insieme il mondo⁵.

Quella proposta da Wells è una costruzione certo utopica; ma per certi versi si tratta di un'utopia molto concreta: un'enciclopedia concepita come cattedrale del sapere, continuamente aggiornata, a disposizione di ogni persona. Non sorprende il fatto che l'idea, probabilmente poco nota al grande pubblico, abbia invece avuto un'enorme fortuna fra chi ha lavorato a immaginare prima, e costruire poi, il nuovo ecosistema digitale e di rete. Arthur C. Clarke, che prima di essere uno scrittore di fantascienza è stato uno scienziato di prim'ordine, già nel 1962 proiettava l'idea di Wells su due dimensioni fra loro collegate: una «Global library» a suo avviso realizzabile intorno all'anno 2000, e il cervello mondiale vero e proprio, un'intelligenza artificiale di livello superiore rispetto a quella umana, a suo avviso realizzabile entro il 2100⁶. Dove meritano di essere sottolineati sia il collegamento stabilito da Clarke fra enciclopedia mondiale e biblioteca globale (uno dei molti esempi di collegamento fra il concetto di enciclopedia e quello di biblioteca), sia le date proposte: è presto per verificare la previsione relativa al World Brain, ma Wikipedia, come vedremo, è nata nel 2001, e anche se non esiste una singola biblioteca digitale globale, si potrebbe ragionevolmente affermare che il decennio 1995-2005 corrisponde più o meno al periodo in cui i progetti di biblioteca digitale (il primo dei quali, il progetto Gutenberg, è stato avviato da Michael Hart nel 1971) hanno raggiunto una loro prima maturità, grazie anche alla possibilità di accedervi attraverso il web.

Oltre a Clarke, moltissimi altri autori⁷ hanno sottolineato le affinità fra le idee di Wells e da un lato – sul versante enciclopedia globale – il world wide web e Wikipedia, dall'altro – sul versante cervello mondiale – il lavoro nel campo dell'intelligenza artificiale. Tornerò più avanti su uno di questi paragoni, mentre qui mi limito a segnalare che la più recente edizione del libro di Wells, non a caso pubblicata dalla MIT Press⁸, ha affidato a Bruce Sterling (uno dei più noti autori di fantascienza cyberpunk) e a Joseph Reagle (che si è occupato a lungo di Wikipedia e ha curato un libro sul tema) due prefazioni dedicate proprio all'attualità delle idee di Wells nel nuovo ecosistema digitale.

Le proposte di Wells meritano però un'ulteriore considerazione: l'enciclopedia globale è concepita come un progetto non solo culturale ma anche e soprattutto politico. Per un verso, questo corrisponde all'idea

– a mio avviso corretta – che la lotta alla dispersione e alla frammentazione e la spinta verso un’architettura complessa e strutturata delle conoscenze siano un’esigenza (e per certi versi una necessità) *politica*, se vogliamo sfuggire al rischio di collasso culturale e sociale che l’assenza di comprensione della complessità può comportare: basti pensare a temi come il riscaldamento globale, la fragilità degli equilibri geopolitici, il controllo degli armamenti, e come vedremo anche le nuove prospettive aperte dalla ricerca in intelligenza artificiale. Per altro verso, rovesciando la prospettiva, suggerisce che anche il lavoro politico, e in particolare il lavoro parlamentare, sia oggi in qualche misura un lavoro di tipo enciclopedico, proprio per la complessità dei temi da affrontare e per l’esigenza di solidità, coerenza e rigore che il lavoro legislativo porta con sé: un’idea esplorata in un interessantissimo volume di Giovanni Rizzoni sul rapporto fra parlamentarismo ed enciclopedia⁹.

I prossimi capitoli intendono esplorare il percorso seguito dall’idea di enciclopedia e di enciclopedismo nel nuovo ecosistema digitale, proprio per comprendere se, come e in quale misura gli strumenti offerti dal digitale permettano di organizzare e strutturare un panorama conoscitivo che è oggi per molti versi ancor più ricco, ma anche più disperso e frammentato, di quello con il quale si confrontava Wells nel 1938. Nell’avviare questa indagine, ho ovviamente l’obbligo di ricordare che la storia delle enciclopedie e dell’enciclopedismo è ben più antica: non proverò qui a riassumerla o a discutere le differenze fra i diversi modelli enciclopedici (a partire dalla distinzione fra enciclopedie sistematiche e alfabetiche), compito che sarebbe superiore alle mie forze e poco funzionale rispetto agli obiettivi del libro che state leggendo, ma segnalo in nota almeno alcuni fra i moltissimi testi dedicati al tema, scelti fra quelli di riferimento e più strettamente funzionali rispetto al seguito della trattazione¹⁰.

Un’ultima premessa da fare riguarda le scelte di periodizzazione. Nel primo capitolo, ho discusso alcuni fra i molti modi possibili per scandire cronologicamente le varie fasi della rivoluzione digitale. Un discorso analogo va fatto anche per il tema specifico delle enciclopedie e dell’enciclopedismo. Come vedremo, già prima della nascita di Wikipedia quasi tutti i principali editori operanti nel settore avevano abbandonato

l'idea della centralità della forma stampata tradizionale, sostituendola o integrandola con versioni online. Da questo punto di vista, le enciclopedie si differenziano da molte altre tipologie di prodotti editoriali e in particolare dai libri di narrativa e di saggistica generalista, che nonostante lo sviluppo dei libri elettronici rimangono per lo più legati – almeno per il momento – al formato cartaceo.

Le ragioni di questo sviluppo sono molteplici: da un lato, le opere di reference, come appunto un'enciclopedia, sono in effetti assai più vicine di un libro tradizionale al modello 'informatico' della base di dati e sono destinate alla consultazione occasionale e puntuale piuttosto che alla lettura lineare. Sono quindi molto meno dipendenti dalle caratteristiche suggerite o richieste da situazioni di lettura lineare e prolungata, rispetto alle quali la carta sembra mantenere ancora alcuni vantaggi sui dispositivi digitali di lettura. Inoltre, i contenuti online possono essere facilmente modificati e quindi aggiornati; le enciclopedie digitali permettono di aggiungere risorse multimediali, e in particolare audio e video, alle informazioni testuali e alle immagini; permettono di utilizzare link per dare la possibilità di esplorare facilmente i riferimenti incrociati da una voce all'altra e di utilizzare strumenti interattivi (la visualizzazione delle funzioni matematiche in Wolfram-Alpha, un'enciclopedia online orientata alla scienza¹¹, è un buon esempio di questa caratteristica). Infine, ma non meno importante, va considerato il fattore prezzo: in generale, le enciclopedie stampate hanno un formato ingombrante e sono molto costose, e il costo marginale di una nuova copia è elevato. Al contrario, nel caso di un'enciclopedia digitale una nuova copia non ha praticamente alcun costo marginale, e lo spazio richiesto al lettore per la sua archiviazione è molto ridotto (se si utilizza un supporto fisico) o inesistente (se l'enciclopedia è online).

Il processo di sviluppo delle enciclopedie digitali, tuttavia, ha richiesto tempi piuttosto lunghi e ha attraversato fasi diverse, strettamente connesse all'evoluzione generale dell'ecosistema digitale. Propongo qui di individuare cinque fasi principali di questo processo, le prime quattro strettamente legate al modello 'architettonico' di organizzazione delle conoscenze da cui siamo partiti, e la quinta, proiettata sul futuro, legata alla sua ibridazione con il modello rappresentato dalle intelligenze

artificiali generative e che abbiamo qui battezzato ‘oracolare’. Per seguire meglio la divisione delle tematiche nei prossimi capitoli, può essere utile anticiparle qui sinteticamente:

1. L’epoca delle prime enciclopedie testuali online, spesso offerte attraverso servizi come Minitel, Prestel, CompuServe: come vedremo, alcuni di essi hanno iniziato a fornire l’accesso a enciclopedie testuali online all’inizio degli anni ’80 del secolo scorso.

2. L’era dei CD-ROM multimediali, basata principalmente sull’idea di enciclopedia multimediale offline: corrisponde in linea di massima al decennio che intercorre tra la fine degli anni ’80 e la fine degli anni ’90, e raggiunge il suo apice nella ‘guerra’ tra l’Encyclopaedia Britannica ed Encarta, l’enciclopedia prodotta da Microsoft, che è anche l’esempio più rappresentativo del passaggio dal modello delle enciclopedie cartacee alle nuove enciclopedie multimediali.

3. La prima generazione di enciclopedie su web, in cui la versione web ha dapprima integrato e poi sostituito completamente i CD-ROM: è il periodo che corrisponde più o meno al primo decennio degli anni 2000, e il suo prodotto chiave è senza dubbio Wikipedia, che sostituisce a sua volta Encarta come enciclopedia più usata.

4. L’era delle enciclopedie semantiche orientate ai dati: è l’epoca che stiamo vivendo ed è legata agli sviluppi del web semantico; la sua caratteristica principale è la progressiva formalizzazione dell’idea di basi di conoscenza enciclopediche altamente strutturate, fondate su ontologie (ovvero sistemi di classificazione: il concetto di ontologia sarà discusso nel cap. 7) organizzate in forma di linked data, consultabili non solo da utenti umani ma anche e soprattutto da una pluralità di agenti software ‘intelligenti’ utilizzati per il recupero dei dati. Alcuni di questi agenti possono rispondere in modo interattivo a domande formulate in linguaggio naturale: è la strada che porta – tra l’altro – ai cosiddetti ‘assistenti vocali’, come Google Assistant, Amazon Alexa, Siri, Bixby, Cortana.

5. L’era dell’integrazione fra enciclopedie semantiche e intelligenze artificiali generative: è l’evoluzione che potrebbe caratterizzare il prossimo futuro, e che è già prefigurata dall’inclusione di Wikipedia all’interno del corpus di addestramento di molti sistemi di intelligenza artificiale

generativa. Vedremo che questo sviluppo, di cui mi occuperò più direttamente nella parte seconda del libro, non è privo di aspetti problematici, legati proprio all'ibridazione fra due paradigmi di base – quello 'architettonico' e quello 'oracolare' – fra loro in realtà assai diversi.

Va detto subito, comunque, che anche l'evoluzione sopra delineata, proprio come quella della rivoluzione digitale nel suo insieme, non è stata sempre lineare: i diversi modelli di enciclopedia e di enciclopedismo si sono intrecciati in forme di volta in volta diverse con gli sviluppi tecnologici. Un esempio di casa nostra è rappresentato dall'Istituto dell'Enciclopedia Italiana, uno dei pochi editori a livello internazionale che continua a vendere enciclopedie cartacee, pur consentendo l'accesso alla maggior parte dei contenuti attraverso un popolare portale web. Il marchio Treccani conserva un'autorità innegabile anche nell'era di Wikipedia, mentre l'Encyclopaedia Britannica non viene più pubblicata in versione cartacea da più di un decennio e la sua versione online ha perso da tempo – se mai l'ha avuta – la sua centralità.

Nonostante queste differenze, ritengo che la distinzione delle cinque fasi sopra individuate sia utile e fundamentalmente corretta; la loro analisi e discussione sarà dunque al centro dei prossimi capitoli di questo libro, anche se ciò imporrà a volte di procedere seguendo criteri diversi da quelli strettamente cronologici.

¹ I capitoli dal 2 al 7 del libro riprendono in parte, in forma aggiornata e notevolmente allargata, alcune considerazioni svolte in Gino Roncaglia, *Encyclopedias and encyclopedism in the era of the Web*, in «JLIS.it» 12, 3 (settembre 2021), pp. 69-90, in rete all'indirizzo <https://www.jlis.it/index.php/jlis/article/view/375/374>, doi: 10.4403/jlis.it-12757, a sua volta sviluppato a partire dall'intervento tenuto in occasione della conferenza internazionale *Enciclopedismo antico e moderno* organizzata da Luciano Canfora a Bari il 17-18 aprile 2017; l'intervento, nella sua forma originaria, era stato pubblicato negli atti del convegno: Gino Roncaglia, *Enciclopedie ed enciclopedismo ai tempi della rete*, in Vincenzo D'Alba, Francesco Maggiore, Vanna Maraglino (a cura di), *Enciclopedismo antico e moderno*, Fondazione Gianfranco Dioguardi («I Quaderni di Varia Cultura», 12), Bari 2020, pp. 193-214.

² H.G. Wells, *World Brain*, Methuen, London 1938. L'edizione originale è

disponibile sull'Internet Archive all'indirizzo <https://archive.org/details/worldbrain00wells/page/n1/mode/2up>. Trad. it. *Per una enciclopedia mondiale*, Elliot, Roma 2019.

³ H.G. Wells, *Per una enciclopedia mondiale* cit., p. 43.

⁴ Ivi, p 17.

⁵ Ivi, pp. 24-27.

⁶ Arthur C. Clarke, *Chart of the future*, in *Profiles of the Future: An Inquiry into the Limits of the Possible*, Harper & Row, New York 1962; la carta è a p. 235 nella revised edition del libro, Harper & Row, New York 1973.

⁷ Alcuni sono citati nella sezione 'Influence' della voce 'World Brain' nell'edizione inglese di Wikipedia: https://en.wikipedia.org/wiki/World_Brain.

⁸ H.G. Wells, *World Brain*, MIT Press, Cambridge (MA) 2021.

⁹ Giovanni Rizzoni, *Parliamentarism and Encyclopaedism: Parliamentary Democracy in an Age of Fragmentation*, Hart Publishing, Oxford 2024 (in corso di pubblicazione). Sono grato a Giovanni Rizzoni per avermi fornito una prima stesura del suo libro.

¹⁰ Per evitare che questa nota sia a sua volta enciclopedica mi limito a citare, in ordine cronologico di pubblicazione, Robert Collison, *Encyclopaedias: their History throughout the Ages*, Hafner, New York-London 1966; Walter Tega (a cura di), *L'unità del sapere e l'ideale enciclopedico nel pensiero moderno*, il Mulino, Bologna 1983; Peter Binkley (a cura di), *Pre-Modern Encyclopedic Texts*, Brill, Leiden 1997; Peter Burke, *A Social History of Knowledge: From Gutenberg to Diderot*, Polity Press, Cambridge 2000, trad. it. *Storia sociale della conoscenza. Da Gutenberg a Diderot*, il Mulino, Bologna 2002; Andrew Brown, *A Brief History of Encyclopedias: From Pliny to Wikipedia*, Hesperus Press, London 2011; Peter Burke, *A Social History of Knowledge: From the Encyclopédie to Wikipedia*, Polity Press, Cambridge 2012, trad. it. *Dall'Encyclopédie a Wikipedia*, il Mulino, Bologna 2013; Jason König e Greg Woolf (a cura di), *Encyclopaedism from Antiquity to the Renaissance*, Cambridge University Press, Cambridge 2013; Jeff Loveland, *The European Encyclopaedia: From 1650 to the Twenty-First Century*, Cambridge University Press, Cambridge 2019; Vincenzo D'Alba, Francesco Maggiore, Vanna Maraglino (a cura di), *Enciclopedia antico e moderno* cit.; Jean-Jacques Chardin, Sorana Corneanu, Richard Somerset (a cura di), *Ordering Knowledge: Disciplinarity and the Shaping of European Modernity*, Presses universitaires de Strasbourg, Strasbourg 2023.

¹¹ Il sito web è <https://www.wolframalpha.com/>.

3.

Pionieri dimenticati e primi servizi online: alle origini dell'enciclopedismo digitale

Il nome di Albert Ducrocq è oggi quasi sconosciuto, e nel momento in cui scrivo non è registrato né dall'edizione in italiano né da quella in inglese di Wikipedia (vedremo se la menzione in questo libro contribuirà a cambiare la situazione...)¹. Eppure, a questa affascinante figura di giornalista, saggista e pioniere della cibernetica francese – all'epoca assai popolare nel suo paese anche per le frequenti apparizioni televisive, e capace di spaziare su tutti i terreni scientifici di frontiera, dalla fisica atomica all'astronautica, dalla chimica alla biologia, dalla matematica alla cibernetica e alla robotica – dobbiamo non solo testi divulgativi in deciso anticipo sui tempi, alcuni dei quali tradotti anche in italiano², ma anche sorprendenti sperimentazioni. Fra le altre cose, Ducrocq lavorò già alla fine degli anni '40 del secolo scorso a piccoli animali-robot programmabili, capaci di raccogliere e interpretare attraverso sensori alcuni dati relativi al loro ambiente esterno, inventò nel 1952 una macchina elettronica per la generazione di testi e poesie chiamata Calliope, costruì nel 1956 la macchina da scrivere elettronica Electro Styl, e produsse quelli che sono probabilmente i primi esempi di arte digitale, generati in codice binario da una rudimentale stampante³. Non stupirà dunque il fatto di trovarlo autore di quello che, pur se lontano dal modello enciclopedico in senso stretto, sembra essere il primo strumento elettronico di reference: l'«informateur électronique», un elaboratore basato su transistor e presentato in occasione della fiera internazionale di Parigi del 1954. Nelle parole del suo inventore, l'informateur électronique aveva l'obiettivo di «informare i giovani sulle carriere che si offrono loro

in base ai loro desideri, alle loro inclinazioni, ai loro risultati scolastici, alle loro preferenze per un lavoro stanziale o in movimento, da soli o in équipe»⁴: un dispositivo informatico interattivo di orientamento al lavoro, dunque, che anticipa di decenni l'attenzione odierna verso applicazioni di questo tipo.



QR-Code 3

Albert Ducrocq presenta alcune delle sue invenzioni cibernetiche

<http://bit.ly/3NmBITV>

L'informateur électronique precede di quattro anni quella che possiamo probabilmente considerare la prima enciclopedia digitale in senso stretto, anch'essa assai poco nota e dimenticata da molte storie dell'enciclopedismo digitale: il 'Larousse elettronico', che fu presentato all'Esposizione universale di Bruxelles del 1958, e che funzionava su un computer Control Data 3600 con memoria a nucleo magnetico⁵. La guida ufficiale della fiera lo descrive come segue: «Larousse mette l'universo delle conoscenze al servizio dell'uomo. [...] Un dizionario elettronico 'risponde' alle domande che gli vengono poste sui temi più vari»⁶.

Un'enciclopedia di questo tipo, basata su mainframe, era tuttavia di scarsa utilità senza la possibilità di accedervi a distanza. Per questo passo dobbiamo aspettare il 1980, quando l'OCLC – sigla che all'epoca si riferiva all'Ohio College Library Center, e che diventerà in seguito famosa in ambito bibliotecario⁷ – annunciò il lancio di un nuovo progetto, presentato come rivoluzionario:

Un servizio di informazione domestica basato su computer che trasforma un normale televisore e un telefono in un servizio di banca domestica, in un'enciclopedia, in un catalogo di biblioteca e in una fonte di informazioni per la comunità, sarà testato in 200

case nell'area di Columbus a partire da ottobre. Chiamato Channel 2000, il servizio di informazione domestica viene sviluppato e testato sul mercato da OCLC, Inc. e Bane One Corporation. Il sistema utilizza un normale televisore, un telefono e uno speciale adattatore progettato da OCLC, Inc.⁸.

Tra i contenuti, resi disponibili attraverso un sistema di accesso remoto alle banche dati in parte simile al Minitel francese (introdotto tra il 1980 e il 1982 e che entro il 1986 avrebbe anch'esso consentito l'accesso ad alcune enciclopedie tematiche⁹), c'era l'Academic American Encyclopedia. E c'era il catalogo integrato delle biblioteche dei comuni di Columbus e Franklin, nell'Ohio. Va notata subito la scelta dei servizi offerti, e in particolare l'associazione stabilita già in questa primissima sperimentazione fra accesso all'enciclopedia e accesso al catalogo online di una biblioteca. Troviamo qui nuovamente associate le due risorse che, integrate, costituiscono – evidentemente anche nella percezione dei progettisti di un sistema come quello proposto – un po' l'architrave di quella cattedrale del sapere che attraverso le nuove tecnologie informatiche e di rete ci si proponeva di rendere direttamente raggiungibile da ogni casa. La lentezza dei collegamenti di rete allora disponibili e le limitate capacità grafiche dei terminali imponevano l'uso di un'interfaccia testuale: ce se ne può fare una buona idea attraverso il suo manuale utente, oggi disponibile sul sito dell'archivio e museo OCLC¹⁰. Si capisce dunque come mai l'enciclopedia stessa fosse in versione solo testuale, senza integrazioni multimediali; caratteristica che, come vedremo, sarà in seguito mantenuta anche nella versione laserdisc, e che troviamo del resto in tutte le prime enciclopedie digitali, proprio a causa delle limitate capacità di gestione di contenuti multimediali nei dispositivi informatici usati all'epoca.



QR-Code 4

Il manuale del sistema Channel 2000 (1980)

<http://bit.ly/3CIJ0XY>

Quanto al contenuto, l'Academic American Encyclopedia era nata proprio nell'agosto 1980 per iniziativa di Aretê Publishing, sussidiaria statunitense del gruppo editoriale olandese VNU, con l'obiettivo di coprire una fetta di mercato a metà strada fra la World Book Encyclopedia (una delle enciclopedie 'storiche' in lingua inglese, pubblicata per la prima volta nel 1917 e basata su successive edizioni annuali, con una formula ibrida fra enciclopedia e annuario e il proposito di seguire man mano tanto lo sviluppo degli avvenimenti storici quanto il progresso in particolare in campo scientifico-tecnologico, rivolgendosi a un pubblico non specialistico e in prevalenza giovanile) e l'Encyclopaedia Britannica, opera di riferimento ma percepita come più complessa e costosa. L'Academic American Encyclopedia era venduta a 400 dollari, un prezzo piuttosto basso considerati i 21 volumi che la componevano, attraverso librerie e grande distribuzione anziché con il tradizionale metodo porta a porta ad opera di venditori specializzati. Le intenzioni – testimoniate anche dalla partecipazione alla sperimentazione fatta con OCLC – erano quelle di svecchiare tanto i meccanismi produttivi quanto i canali e i modelli distributivi tradizionalmente propri dell'editoria enciclopedica. Occorre tener presente che nel 1980 la strada del digitale era, per un'opera editoriale destinata al grande pubblico, davvero pionieristica: basti pensare che l'anno successivo un articolo del «New York Times» dedicato ai primi risultati di vendita (abbastanza deludenti) della nuova enciclopedia, nel discutere la differenza fra i vari canali distributivi, non accenna affatto alla sua inclusione all'interno di Channel 2000¹¹.

Il servizio OCLC sembra essere stato il primo di questo tipo rivolto a un pubblico generico, che accede a un'enciclopedia da casa¹². Se dal campo delle enciclopedie allarghiamo però lo sguardo ad altre tipologie di opere di reference, va ricordato che già all'inizio degli anni '70 proprio il «New York Times» aveva creato un sistema informativo informatizzato, fra i primissimi nel settore del reference, denominato New York Times Information Bank. La sua ideazione si deve a un'altra figura oggi quasi dimenticata (in questo caso, e nel momento in cui scrivo, anche da Wikipedia in lingua inglese) e che merita invece di essere ricordata: John Rothman.

Rothman aveva una professione non comune ma affascinante: quella di indicizzatore. Lavorava, fin dal 1946, a *The New York Times Index*, lo strumento che dal 1863 in poi – e con maggior regolarità e diffusione dal 1913 – ha offerto indici a stampa, mensili, quadrimestrali e annuali, degli articoli del giornale¹³. E il lavoro di chi crea indici¹⁴, di chi lavora a schedature, di chi produce classificazioni, liste e cataloghi, ha inevitabilmente molti punti di contatto tanto con il lavoro enciclopedico di sistematizzazione 'astratta' del sapere quanto con quello bibliotecario, archivistico e museale di sistematizzazione e di accesso alle sue manifestazioni concrete: lavori a loro volta basati su indicizzazioni, schedature, ontologie e classificazioni, organizzazione di contenuti, voci e concetti¹⁵. Lo stesso Rothman ha ben sintetizzato in un articolo del 1965 la funzione di mediazione informativa affidata agli indici ai quali lavorava:

L'indicizzazione è un gigantesco gioco di predizione. Gli indicizzatori devono saper valutare in anticipo quali informazioni gli utenti potrebbero prevedibilmente cercare, dove potrebbero cercarla, e quanto dettaglio l'abstract o la voce debbano includere per evitarli un viaggio fino agli archivi del giornale¹⁶.

L'*Index* era un'opera sterminata: una sola annualità poteva comprendere migliaia di pagine e i volumi che la componevano, ricorda Rothman, avevano spesso una forma quasi cubica¹⁷. Un indice così ampio e complesso rappresentava il contenuto ideale per una gestione informatizzata, e già nel 1965 Rothman propone di provare a farlo; l'anno successivo avvia una collaborazione con l'IBM, e nella prefazione all'edizione annuale del 1968 può affermare con orgoglio che l'*Index* «ha appena completato una transizione durata due anni a un processo di

produzione assistito da computer, che ci permette di predisporre abstract e indici di una quantità maggiore di materiale, in maniera più esaustiva, accurata ed efficiente»¹⁸. Nel 1972 l'accesso alla banca dati così realizzata diventa possibile anche attraverso i terminali interni a disposizione dei giornalisti, e lo stesso anno il primo accesso esterno è attivato presso la biblioteca Hillman dell'università di Pittsburgh. Sei mesi dopo gli accessi esterni erano diventati quattordici, fra cui la NBC, la Associated Press, il Dipartimento di Stato, la CIA, la Library of Congress, la Exxon e la Chase Manhattan Bank¹⁹. Una volta identificati online gli articoli di loro interesse, gli utenti delle postazioni esterne – collegati al mainframe attraverso linea telefonica e un apparato proprietario di modulazione e demodulazione del segnale – potevano richiederne l'invio dal giornale in forma di microfiche.

La presenza delle biblioteche fra le prime utenze di un sistema di questo tipo²⁰ testimonia ancora una volta una vicinanza determinata dal comune radicamento nella funzione di mediazione e offerta informativa: la biblioteca può così girare alla propria utenza l'accesso a un sistema che archivia e organizza i contenuti ai quali offre l'accesso, trasformandoli in tal modo in componenti di un edificio conoscitivo architettonicamente più ampio, solido e strutturato. Se si considera che la rivoluzione rappresentata dall'informatica personale era ancora lontana diversi anni, ci si renderà conto del carattere assolutamente pionieristico – sia dal punto di vista tecnologico, sia da quello dell'organizzazione e dell'accesso alle informazioni – del lavoro realizzato da Rothman e dal suo giornale.

A partire dall'inizio degli anni '80 – quando di informatica personale si comincia invece a parlare – entrano direttamente in gioco, come abbiamo visto, anche le enciclopedie, e i diversi sistemi di gestione e mediazione informativa si moltiplicano e si intrecciano fra loro, in una dinamica che diventa ben presto commerciale e concorrenziale. Abbiamo visto come l'esperimento dell'OCLC fosse destinato a un gruppo di utenti certo ristretto ma composto da cittadini 'comuni', legati dall'appartenenza alla stessa realtà locale; poteva dunque permettersi, anche e proprio per la sua natura sperimentale, un carattere abbastanza inclusivo e socializzante: ricordiamo la presenza fra i suoi servizi di una bacheca destinata allo scambio di informazioni all'interno della comunità. Le offerte che

spuntano come funghi negli anni successivi seguono invece il modello che era stato adottato dall'Information Bank, e propongono servizi a pagamento, inizialmente abbastanza esclusivi. Nel 1981, l'anno successivo al lancio di Channel 2000, ancora il «New York Times» sceglierà di includere l'Academic American Encyclopedia all'interno dell'Information Bank (che nel frattempo si era allargata dagli archivi alla versione corrente del quotidiano), integrando per la prima volta informazione enciclopedica e informazione giornalistica, mentre LexisNexis, «una rete chiusa, basata su abbonamento, che permetteva ai professionisti del settore legale, tra gli altri, di ricercare documenti in formato elettronico»²¹, offrirà a sua volta accesso sia all'Information Bank del «New York Times» sia a una enciclopedia: in questo caso la versione digitale dell'Encyclopaedia Britannica, che arriva seconda ma comunque ben piazzata – come vedremo, i problemi della Britannica inizieranno dopo – in questa primissima corsa al digitale. Nel 1982, è il CompuServe Information Service, uno dei primi servizi online dial-up a costi accessibili (quelli cioè che offrono accesso ai propri server – e in seguito a Internet – attraverso collegamenti via modem, usando normali linee telefoniche e rivolgendosi a un'utenza generalista), a includere, accanto a quello a diversi quotidiani, fra cui il «New York Times», l'accesso online alla World Book Encyclopedia; nello stesso anno, l'accesso all'Academic American Encyclopedia è offerto da Dow Jones News/Retrieval, un servizio a pagamento di informazione finanziaria, e nel 1983 faranno lo stesso il Bibliographic Research Service e il servizio britannico di videotext Prestel²². Sempre nel 1983, Atari pubblicizza uno dei suoi home computer, il modello 1400XL (che disponeva di un modem incorporato), facendo riferimento alla sua capacità di collegarsi proprio a CompuServe e al Dow Jones News/Retrieval, e di accedere in tal modo a «un ampio ventaglio di basi dati di reference»²³.

Meno di tre anni dopo il lancio del progetto sperimentale dell'OCLC, dunque, enciclopedie online sono incluse nell'offerta di molti fra i primi servizi che permettono – da casa o dall'ufficio – un accesso a pagamento a risorse di rete. E i prezzi di questi servizi sono, se non proprio economici, tutto sommato abbastanza accessibili: nel caso di CompuServe, 5-6 dollari l'ora; con l'aggiunta però spesso di costi aggiuntivi legati ai singoli

servizi, e con tempi di accesso alle informazioni che, per la lentezza dei collegamenti, erano spesso assai lunghi: se qualcuno avesse voluto scaricare l'intera edizione del «New York Times», anziché un singolo articolo, si sarebbe dovuto rassegnare a tempi d'attesa di diverse ore e a pagare una ventina di dollari, per un contenuto che in edicola costava venti centesimi.

Senza seguire in dettaglio uno sviluppo che a questo punto, con il moltiplicarsi dei fornitori di servizi dial-up, diventa abbastanza frenetico, possiamo dire che in pochi anni l'Academic American Encyclopedia vince la prima battaglia tra le enciclopedie online, venendo offerta da «quasi tutti i servizi commerciali di informazione dial-up, tra cui CompuServe, Delphi e Prodigy»²⁴.

Le vicende che ho provato a ricostruire in questo capitolo mostrano che la storia delle enciclopedie digitali inizia ben prima della data più spesso citata in letteratura, quella della prima enciclopedia su laserdisc. Traguardo che – come a questo punto non stupirà – sarà raggiunto ancora una volta dall'Academic American Encyclopedia. Siamo nel 1984, ma la commercializzazione di questo prodotto, denominato Grolier KnowledgeDisc, inizia l'anno dopo, quando Aretê è acquistata da Grolier, un editore statunitense specializzato in edizioni raffinate di opere classiche e che fin dal 1910 si era affacciato anche nel mercato delle opere di reference²⁵. Per l'enciclopedismo digitale inizia una nuova epoca²⁶.

¹ Esiste invece una voce a lui dedicata nell'edizione francese di Wikipedia, all'indirizzo https://fr.wikipedia.org/wiki/Albert_Ducrocq.

² Per fare solo un esempio, il suo *La science à la conquête du passé*, Plon, Paris 1955, tradotto in italiano nel 1959 nella collana Universale Cappelli con il titolo *La scienza alla scoperta del passato*, è stato probabilmente il primo libro divulgativo a discutere il metodo della datazione con carbonio 14.

³ Su alcune di queste realizzazioni si veda Ryan Walach, *Afterthought to "A Robo-Poet Does Not Scare Us"*, in «Hypocrite Reader» 28, maggio 2013, in rete all'indirizzo <https://hypocritereader.com/28/afterthought-to-robo-poet>.

⁴ Albert Ducrocq, *Le changement global*, J.-C. Lattès, Paris 1993, p. 29. Un'immagine del dispositivo è nella collezione dell'agenzia fotografica Roger-

Viollet: <https://www.roger-viollet.fr/image-photo/informateur-electronique-paris-le-1er-mars-1954-roger-viollet-roger-viollet-433767>.

⁵ Si veda al riguardo Jean-Yves Mollier e Bruno Dubot, *Histoire de la librairie Larousse (1852-2010)*, Fayard, Paris 2012, pp. 537-538, cit. in Jeff Loveland, *The European Encyclopedia* cit., pp. 358-359.

⁶ *Guide Officiel Exposition Universelle de Bruxelles 1958*, Desclée & Cie, Bruxelles 1958, cit. in https://www.worldfairs.info/expopavillondetails.php?expo_id=14&pavillon_id=168.

⁷ Il significato dell'acronimo cambierà successivamente in Online Computer Library Center e l'organizzazione diventerà uno dei principali servizi di ambito bibliotecario al mondo, responsabile della gestione di WorldCat, il più grande catalogo internazionale online, e del sistema di classificazione decimale Dewey.

⁸ «LITA Newsletter» 2, estate 1980, p. 4, disponibile in rete all'indirizzo https://alair.ala.org/bitstream/handle/11213/2104/LITAnewsletter_1980_02.pdf.

⁹ Cfr. Michel Mennig, *L'indispensable pour Minitel*, Marabout, Paris 1986; edizione digitale, FeniXX, Paris 2016. Per un parallelo fra il Minitel francese e i servizi dial-up statunitensi dell'epoca si veda Julien Mailland, *Minitel, the Open Network Before the Internet*, in «The Atlantic», 16 giugno 2017, disponibile in rete all'indirizzo <https://www.theatlantic.com/technology/archive/2017/06/minitel/530646/>.

¹⁰ In rete all'indirizzo <https://library.oclc.org/digital/collection/p267701coll27/id/4438>.

¹¹ N.R. Kleinfeld, *Arête Encyclopedia Misses Sales Target*, in «The New York Times», 11 marzo 1981, consultabile in rete all'indirizzo <https://timesmachine.nytimes.com/timesmachine/1981/03/11/issue.html>.

¹² L'altrimenti prezioso volume di Jeff Loveland, *The European Encyclopedia* cit., non menziona questa esperienza e afferma erroneamente che «La prima azienda a mettere un'enciclopedia online è stata l'Encyclopaedia Britannica» (p. 367).

¹³ Si veda al riguardo Jo A. Cates, *The Record of Record: The New York Times Index*, in James Rettig (a cura di), *Distinguished Classics of Reference Publishing*, Oryx Press, Phoenix 1992, pp. 174-179.

¹⁴ Non posso qui non ricordare, su questi temi, il recente e bellissimo libro di Dennis Duncan *Index, A History of the. A Bookish Adventure from Medieval Manuscripts to the Digital Age*, Penguin, London 2021 (trad. it. *Indice, storia dell': Dai manoscritti a Google, l'avventurosa storia di come abbiamo imparato a orientarci nel sapere*, UTET, Torino 2022), e – con un taglio diverso, ma non meno poliedrico – il testo ormai classico di Umberto Eco *Vertigine della lista*, Bompiani, Milano 2009. La rivista «The Indexer. The International Journal of Indexing» (<https://www.theindexer.org/>), pubblicata a partire dal 1958 dalla Society of

Indexers, offre nei suoi fascicoli, che al momento in cui scrivo coprono ormai sessantacinque annate, un panorama sterminato dell'affascinante mondo dell'indicizzazione.

¹⁵ Per un panorama costruito attraverso uno studio di casi tratti dai tre mondi interconnessi delle enciclopedie, delle biblioteche e dell'indicizzazione si vedano i saggi raccolti in James Rettig (a cura di), *Distinguished Classics* cit.

¹⁶ Cit. in Richard Sandomir, *John Rothman, Who Made The Times's Archives Accessible, Dies at 95*, in «The New York Times», 1° ottobre 2019, <https://www.nytimes.com/2019/10/01/business/media/john-rothman-dead.html>.

¹⁷ Jo A. Cates, *The Record of Record* cit., p. 175.

¹⁸ Cit. *ivi*, pp. 177-178. Una descrizione più dettagliata di queste fasi del progetto è in John Rothman, *The New York Times Information Bank*, in «Special Libraries» 63, 3, marzo 1972, pp. 111-115.

¹⁹ Richard Sandomir, *John Rothman* cit.

²⁰ Si veda al riguardo Rhoda Garoogian, *Library Use of the New York Times Information Bank: A Preliminary Survey*, in «RQ» 16, 1, autunno 1976, pp. 59-64, <https://www.jstor.org/stable/41354525>.

²¹ Jeff Loveland, *The European Encyclopedia* cit., p. 369.

²² Per queste informazioni si vedano Julien Mailland, *Minitel* cit.; Jeff Loveland, *The European Encyclopedia* cit., p. 396; James Rice, *Videotext, Teletext, and the Impact of Television Information Systems on Library Services*, in Allen Kent (a cura di), *Encyclopedia of Library and Information Science*, vol. 39, suppl. 4, Marcel Dekker Inc., New York 1985, pp. 418-430.

²³ Pubblicità Atari del giugno 1983; una digitalizzazione è disponibile all'interno dell'Internet Archive, all'indirizzo <https://archive.org/details/AtariHomeComputersUpdateJune1983/mode/2up?view=theater>.

²⁴ Jeff Loveland, *The European Encyclopedia* cit., p. 369, con riferimento a Gordon Flagg, *Online Encyclopedias: Are They Ready for Libraries? And Are Libraries Ready for Them?*, in «American Libraries» 14, 1983, pp. 134-136.

²⁵ Grolier fu poi acquistata dalla francese Hachette nel 1988 e rivenduta a Scholastic nel 2000.

²⁶ Un'utile risorsa di riferimento sulle enciclopedie prodotte in questo periodo è Kenneth Kister, *Kister's Best Encyclopedias: A Comparative Guide to General and Specialized Encyclopedias*, Oryx Press, Phoenix 1994.

4.

Il multimedia all'assalto della carta

La prima edizione su laserdisc dell'Academic American Encyclopedia aveva reso subito chiaro uno dei vantaggi principali delle enciclopedie digitali rispetto a quelle cartacee: il prezzo. Nel 1985 la versione a stampa in 21 volumi costava infatti 650 dollari (il prezzo iniziale di 400 dollari si era rivelato economicamente non sostenibile), mentre quella su laserdisc ne costava 89,95. Anche considerando il costo del lettore laserdisc (circa 450 dollari), che era ovviamente utilizzabile anche per altri contenuti, il bilancio economico era dunque positivo fin dall'inizio¹. Negli anni successivi, Grolier pubblicò una serie di nuove edizioni dell'enciclopedia con nomi diversi: The Electronic Encyclopedia (1986), The Grolier Electronic Encyclopedia (1987), The New Grolier Electronic Encyclopedia (1988-91), The New Grolier Multimedia Encyclopedia (1992)². L'edizione del 1987 passò dal laserdisc a una prima versione di CD-ROM, il cosiddetto CD-32, disponibile per i computer Atari; nelle versioni successive il supporto scelto fu un CD-ROM standard³.

Mentre indici sempre più sofisticati miglioravano la velocità di accesso – nella versione del 1987 dell'enciclopedia il testo occupava 60 Mbyte e gli indici 48 Mbyte, permettendo così di recuperare informazioni su oltre sette milioni di termini indicizzati in meno di tre secondi⁴ – il contenuto rimaneva inizialmente solo testuale. Per avere una vera enciclopedia multimediale, con l'inclusione di contenuti audiovisivi, dobbiamo aspettare il 1989⁵, quando viene pubblicata la Compton's Multimedia Encyclopedia.

La Compton Encyclopedia era nata nel 1922 come Compton's Pictured Encyclopedia e si concentrava dunque fin dalle origini sull'elemento visivo, anche in relazione al fatto che si rivolgeva principalmente alle

famiglie e alle scuole. Nel 1989, Compton era di proprietà di Encyclopaedia Britannica, Inc. (che ha detenuto la maggioranza delle azioni dal 1961 al 1993, e poi di nuovo dal 2002 in poi), per cui il CD-ROM del 1989 può anche essere considerato il primo passo compiuto dalla Britannica nella direzione delle enciclopedie basate su CD-ROM⁶. E tuttavia in quegli anni la Britannica sembra ancora incapace di capire la portata della rivoluzione in corso: una sua curiosa pubblicità televisiva del 1987 presenta l'enciclopedia come complemento perfetto del computer, ma lo fa con riferimento all'enciclopedia cartacea, di cui propone l'acquisto telefonico (il prezzo dell'enciclopedia cartacea era all'epoca di circa 1.400 dollari). Il risultato è un mix rivelatosi presto abbastanza anacronistico, che fa ben capire tanto la portata della transizione, quanto la difficoltà di gestirla da parte di un soggetto editoriale tradizionale.



QR-Code 5

Publicità dell'Encyclopaedia Britannica, 1987

<http://bit.ly/44kOK66>

Nel corso degli anni '90 le enciclopedie su CD-ROM divennero sempre più diffuse, con l'ingresso nel nuovo campo di tutti i maggiori editori del settore, e la concorrenza cominciò a riguardare anche l'inclusione di immagini e brevi spezzoni audio o video. Fra il 1991 e il 1992 una nuova versione della Compton's Encyclopedia fu resa disponibile, con il nome di Compton's Interactive Encyclopedia, sia per la prima versione a larga diffusione di Windows (la 3.1) sia per Mac, e l'IBM iniziò a distribuirla in associazione con il suo personal computer di punta, l'IBM/PS1. Sempre nel 1992 anche Grolier produsse una propria enciclopedia multimediale, la già ricordata New Grolier Multimedia Encyclopedia, che grazie a un accordo con Apple veniva inclusa gratuitamente nell'acquisto di un

computer Mac. L'associazione fra computer ed enciclopedia, che trasforma il computer in strumento di accesso e gestione organizzata delle conoscenze, viene evidentemente rafforzata dalla diffusione di offerte di questo tipo.

Nel 1995, sia l'Encyclopedia Americana (sempre di Grolier, in cui il testo e i contenuti multimediali erano memorizzati in due diversi CD-ROM) sia la Hachette Multimedia Encyclopedia (EHM – all'epoca Grolier era di proprietà di Hachette; i due prodotti potevano quindi usufruire di un'infrastruttura tecnica comune) furono pubblicate in versione CD-ROM. Negli stessi anni, molti altri editori entrarono nel nuovo mercato digitale. In Italia, l'enciclopedia multimediale Omnia della De Agostini fu uno dei maggiori successi: pubblicata per la prima volta nel 1997, nel giro di pochi anni fu adattata ad altri paesi (Polonia, Germania e Grecia) e prodotta in una versione per bambini (Omnia Junior) e in una versione settoriale (Omnia Art). Inoltre, sempre nel 1997, si affacciò per la prima volta nel settore anche l'Istituto dell'Enciclopedia Italiana, con un CD-ROM che conteneva la seconda edizione del Vocabolario Treccani.



QR-Code 6

Videorecensione della Compton's Interactive Encyclopedia, dal canale YouTube
'The Doppler Effect'

<http://bit.ly/3r0wSPx>

Le enciclopedie prese in considerazione finora erano principalmente adattamenti di versioni a stampa, o erano comunque prodotte da editori già attivi nel settore. Tuttavia, a partire dagli anni '90 entrarono nel settore anche nuovi attori: aziende che non operavano nel campo dell'editoria enciclopedica su carta, ma che si proponevano di lavorare direttamente sui

nuovi supporti digitali. Come vedremo, questo cambiamento ebbe un impatto sostanziale non solo sul mercato, ma anche sulle caratteristiche e sul formato delle enciclopedie. Gli editori di enciclopedie tradizionali erano naturalmente interessati a evitare che i nuovi supporti sottraessero quote di mercato alla carta e mettessero in crisi la relativa rete di vendita, per lo più ancora effettuata da venditori porta a porta, ed erano quindi propensi a mantenere prezzi relativamente alti, o a vendere la versione su CD-ROM esclusivamente in associazione a quella cartacea. La Compton's Encyclopedia distribuita da Britannica è un buon esempio di questa strategia: era gratuita per chi acquistava la versione cartacea dell'enciclopedia, ma il prezzo era di ben 895 dollari se la si voleva acquistare separatamente. E quando nel 1994 la stessa Britannica fu finalmente rilasciata su CD-ROM, il timore di minare la rete di vendita tradizionale fece sì che quest'ultima fosse venduta solo insieme all'enciclopedia cartacea, a un prezzo che nel frattempo era sceso da 1.400 a 1.200 dollari ma rimaneva comunque assai alto: una scelta che a posteriori possiamo valutare come una sorta di suicidio commerciale, con l'apertura alla concorrenza di vere e proprie praterie, che inevitabilmente altri soggetti editoriali avrebbero subito occupato.

Un'occasione così ghiotta non poteva sfuggire, in particolare, al principale attore che si era imposto nel nuovo settore dell'informatica personale: Microsoft. La Microsoft si era già affacciata nel settore del reference su CD-ROM nel 1987, con un prodotto denominato Bookshelf che integrava fra l'altro materiali dal *World Almanac*, dal *Chicago Manual of Style* e da *Bartlett's Familiar Quotations*; ma il costo era assai alto, 1.099 dollari⁷. Nel 1993 però la situazione era matura per un passo molto più visibile: in quell'anno Microsoft pubblicò – come prevedibile, direttamente su CD-ROM – la prima versione della sua enciclopedia, denominata Encarta. La scelta del nome, fatta da un'agenzia pubblicitaria (il primo nome in codice del progetto era 'Gandalf'), rimanda all'idea dell'enciclopedia come una sorta di carta nautica, strumento di orientamento per la navigazione nel mare del sapere. Il marchio Microsoft e la diffusione diretta su CD-ROM bastavano a garantire alla nuova enciclopedia la connotazione di prodotto innovativo, capace di conciliare la tradizione enciclopedica e le nuove potenzialità del digitale.

Encarta si basava sulla Funk & Wagnalls Encyclopedia, di cui Microsoft aveva acquisito i diritti di riutilizzo non esclusivi, e in seguito anche sulla Collier's Encyclopedia e sulla New Merit Scholar's Encyclopedia, che Microsoft acquisisce da Macmillan Publishers. Nel 1985, Microsoft aveva cercato anche di acquisire diritti non esclusivi sulle voci dell'Encyclopaedia Britannica, proponendo di siglare una partnership, che Britannica rifiutò. Nel 1989, Britannica e Microsoft erano però ancora in buoni rapporti e Microsoft era stato uno dei partner di lancio dell'enciclopedia Compton.

Quando Microsoft lanciò sul mercato la propria enciclopedia, la situazione cambiò e il fragile equilibrio tra carta e digitale si ruppe definitivamente. La prima edizione di Encarta riuscì a vendere solo 10.000 copie: il prezzo risentiva ancora delle strategie portate avanti dagli editori tradizionali ed era troppo alto (295 dollari), mentre la necessità di posticipare la data di uscita da settembre a marzo a causa di problemi tecnici non consentì di sfruttare la forza trainante dello shopping natalizio. Ma l'edizione successiva, a distanza di pochi mesi, corresse subito questi errori: il lancio avvenne a ottobre, a un prezzo di soli 99 dollari. Il successo fu immediato: la nuova Encarta vendette 120.000 copie durante le festività natalizie, 350.000 copie nel primo anno, un milione di copie in due anni. Poiché il costo fisico di un singolo CD-ROM ammontava a pochi dollari, la rete distributiva era quella già esistente per i prodotti informatici dell'azienda, e i costi erano soprattutto quelli redazionali e di realizzazione tecnica, è facile capire che un aumento così considerevole delle vendite ebbe un impatto immediato sulla redditività dell'operazione:

L'organizzazione crebbe fino a sviluppare un enorme vantaggio competitivo. Come in un'enciclopedia, i costi variabili erano bassi mentre i costi fissi erano alti. Tuttavia, il livello di questi costi permetteva un punto di pareggio molto diverso per il prodotto finale, perché gli alti volumi lo abbassavano molto. Il costo dei materiali per il CD e l'imballaggio era piuttosto basso, a volte 5 dollari a scatola. Le spese fisse derivavano dalla redazione, dai programmatori e dal personale aggiuntivo necessario per aggiornare i contenuti e per adattare il prodotto a lingue e paesi diversi. Come per le enciclopedie tradizionali, quindi, la strategia commerciale si orientava verso la generazione di un volume di vendite sufficiente a sostenere entrate che coprissero i costi fissi dell'operazione – anche in questo caso, con una differenza fondamentale: i volumi unitari erano nell'ordine dei milioni anziché delle centinaia di migliaia⁸.

Encarta aveva inoltre un'altra arma a sua disposizione: Microsoft poteva infatti costruire (e costruì) delle offerte di 'pacchetto' associandola, a prezzo scontato e in un'occasione anche gratuitamente, ad altri prodotti informatici, come Microsoft Office. Altro fattore importante da considerare era il suo costante aggiornamento, con nuove edizioni annuali, realizzate in una decina di lingue diverse, e un sostanziale miglioramento dei contenuti multimediali da un'edizione all'altra: una possibilità aperta da un ciclo produttivo divenuto molto più veloce ed economico una volta sostenuti i costi per i contenuti editoriali di base. Naturalmente, la pubblicità delle nuove enciclopedie digitali ha enfatizzato fortemente questo elemento, diventato ancora più centrale con l'arrivo, alla metà degli anni '90, delle prime enciclopedie ibride, che integravano contenuti di base su CD-ROM e aggiornamenti online. Nel giro di pochi anni, le enciclopedie ibride si sarebbero trasformate in enciclopedie esclusivamente online.



QR-Code 7

Chi ha ucciso l'enciclopedia? Uno speciale di «Tedium» con immagini e link utili

<http://bit.ly/3COg7cT>

Per affrontare la sfida rappresentata da Encarta, l'Encyclopaedia Britannica fu costretta a una rincorsa ormai impossibile: il tentativo fu quello di abbassare i prezzi sulla versione CD-ROM dell'enciclopedia (da 995 dollari nel 1995 a 200 dollari nel 1996), ma era troppo tardi: già nel 1995 le vendite erano calate del 50% rispetto al 1990. L'ultimo tentativo, audace ma disperato, fu quello di lanciare un'edizione esclusivamente online: in questo caso, Britannica anticipò Encarta, che preferì intraprendere la strada della versione ibrida; ma nel 1995, quando Britannica Online fu lanciata (con un prezzo di 150 euro all'anno, più 25

euro per la registrazione iniziale), le persone che si connettevano a Internet da casa non erano ancora molte e utilizzavano principalmente servizi dial-up, con i costi di connessione che si sommarono a quelli dell'abbonamento; inoltre, quei pochi che avevano la fortuna di potersi connettere abbastanza frequentemente avevano spesso già Encarta su CD-ROM. L'operazione non portò i risultati sperati, le passività aumentarono e nel 1996 il marchio e le risorse dell'Encyclopaedia Britannica furono venduti per 135 milioni di dollari a un investitore privato, Jacqui Safra. Rispetto a pochi anni prima, il valore dell'azienda era crollato ed era legato quasi esclusivamente alla valutazione del marchio⁹.

In poche parole, contrariamente a quanto a volte si afferma, è stata Encarta e non Wikipedia a sconvolgere il mercato delle tradizionali enciclopedie cartacee; Wikipedia è nata solo successivamente e semmai è stata, una decina di anni dopo, all'origine della crisi del modello delle enciclopedie digitali commerciali, inclusa Encarta.

Di fatto, Encarta ha continuato a vendere abbastanza bene fino al 2003-2004, prima su CD-ROM, poi nella versione ibrida e infine online a partire dal 2000; nel periodo di massima diffusione sono state pubblicate, oltre alle già ricordate versioni plurilingue¹⁰, una versione specifica per bambini (Encarta Kids), una versione dedicata alla storia e cultura nera, nota come Encarta Africana, e una versione Premium, con una quantità assai maggiore di contenuti multimediali. Riccardo Fedriga, che ha lavorato nella redazione italiana, ricorda bene quegli anni per certi versi entusiasmanti, e fornisce un quadro assai vivido del lavoro di localizzazione di Encarta per il nostro paese:

Gli editor erano una decina, uno per ogni area. Poi c'era una stanza nella quale stavano quelli che chiamavamo 'gli amerikani', cioè i responsabili dell'architettura digitale, che venivano direttamente dagli USA: anzi da Redmond, sede della Microsoft. L'enciclopedia da riorientare era molto popolare, Funk & Wagnalls. C'erano 70 parole nella voce sul ciclismo, ricordo, e io le portai a 15.000, riempiendola di quelle che noi chiamavamo le hotword (i link) e facendomi aiutare da Gino Cervi, che coordinava l'impresa e che poi andò a dirigere il Touring; così anche per altre discipline, tra cui i giochi, per cui il consulente fu Giampaolo Dossena (con il quale poi diventammo amici, e ci divertivamo a scriverci di enciclopedie immaginarie). Questo per dire che, per quanto popolare, l'edizione italiana era molto raffinata – grazie soprattutto a Vittorio Di Giuro, che dirigeva il lavoro. Negli anni poi Microsoft acquistò anche altre enciclopedie, tra cui la Collier's, e divenne davvero piena di contenuti. La versione deluxe di Encarta

era la Premium (e forse c'era anche la Gold): per noi che ci lavoravamo era l'Everest. Piano piano, anno dopo anno, esaurito il riorientamento, rimasero gli aggiornamenti per cui la redazione – che in certi momenti arrivò anche a una trentina di persone e centinaia di collaboratori esterni – si esaurì e venne falciata, sino a restare ai pochissimi che gestirono gli ultimi aggiornamenti¹¹.

In effetti, gradualmente gli accessi cominciarono a diminuire a favore di Wikipedia, che era nata nel gennaio 2001, fino a quando, nel 2009, Microsoft abbandonerà il progetto. Nel mese di gennaio di quell'anno, il 97% degli accessi alle enciclopedie online degli Stati Uniti erano relativi a Wikipedia, mentre la percentuale di accessi a Encarta si era ridotta all'1,27% (e quelli alla Britannica Online allo 0,57%)¹².

Se Encarta ha rappresentato senza dubbio l'esempio di maggior successo di enciclopedia multimediale prodotta da un'azienda focalizzata sull'editoria digitale e priva di precedenti esperienze nell'editoria tradizionale, tra i progetti di enciclopedia «nativamente digitale» merita di essere citato un rilevante esempio italiano: Encyclomedia, l'enciclopedia diretta da Umberto Eco e Danco Singer. Encyclomedia è stata concepita come una storia multimediale completa della civiltà europea, basata su CD-ROM. Il progetto è iniziato nel 1993 con la costituzione di Horizons Unlimited, una società fondata da un gruppo di studenti bolognesi di Eco; i primi due CD-ROM dedicati al XVI e XVII secolo sono stati pubblicati a un prezzo piuttosto elevato (399.000 lire italiane ciascuno, equivalenti più o meno a 235 dollari) tra il 1995 e il 1996, ma con caratteristiche avanzate per l'epoca. Il CD-ROM sul Seicento comprendeva circa 9.000 voci, 2.000 immagini, 165 clip audio, 32 animazioni e 13 video, oltre a 203 ebook¹³. Negli anni successivi, l'opera è stata completata e distribuita in diversi formati, su CD-ROM e a stampa (in forma di collaterale in edicola), fino alla versione online.

Volendo provare a riassumere sinteticamente gli aspetti più rilevanti della stagione – tutto sommato abbastanza breve – delle enciclopedie su CD-ROM, credo sia utile sottolineare quattro aspetti:

1. L'affermazione di un nuovo modello di enciclopedia multicodice e interattiva, in cui audio e video (ma in diversi casi anche ricostruzioni 3D, quiz e giochi, mappe interattive, linee del tempo...) integrano testo e immagini. Uno sviluppo importante, che aumenta contemporaneamente

sia la complessità tecnica e architettonica dell'enciclopedia come oggetto informativo, sia la sua efficacia e comprensibilità: attraverso multicodicalità e interattività, l'enciclopedia diventa capace di adattare i linguaggi e i codici comunicativi usati alle necessità specifiche dei singoli argomenti trattati e, almeno in parte, anche delle diverse tipologie di utenza.

2. L'abbassamento radicale dei costi rispetto alle enciclopedie a stampa e ai primi servizi di enciclopedia testuale online, che rende possibile avvicinare un numero maggiore di persone, e in particolare un numero maggiore di giovani, a esempi almeno in alcuni casi qualitativamente eccellenti di integrazione fra strumenti digitali di gestione delle informazioni e modello enciclopedico di organizzazione strutturata del sapere.

3. Anche per questo motivo, la diffusione capillare delle enciclopedie digitali – iniziata di fatto con le enciclopedie su CD-ROM – rappresenta un contraltare (certo parziale e specifico) rispetto alla frammentazione delle forme di testualità digitale di cui ho parlato nel primo capitolo e, in forma più estesa, ne *L'età della frammentazione*. Va notato al riguardo che, se da un lato l'enciclopedia – basata su voci indipendenti – sembrerebbe restare comunque un oggetto informativo sostanzialmente granulare, d'altro canto la possibilità di gestire in modo assai più facile e immediato link e rimandi da una voce all'altra, che caratterizza le enciclopedie digitali rispetto a quelle su carta, trasforma la granularità delle voci in una rete di rimandi straordinariamente complessa e strutturata.

4. Nel panorama dell'editoria digitale, le enciclopedie su CD-ROM rappresentano uno dei pochissimi modelli di successo di realizzazione di contenuti 'arricchiti'. Sappiamo, ad esempio, che i libri elettronici aumentati (quelli che nel mercato anglosassone sono spesso chiamati 'enhanced books' o 'augmented books'), nonostante alcune sperimentazioni di grande interesse, non hanno avuto molto successo. Lo stesso è finora accaduto ai libri di testo aumentati: la maggior parte di docenti, studentesse e studenti preferiscono per ora continuare a utilizzare i testi cartacei, pur avendo a disposizione versioni elettroniche almeno nominalmente più ricche. Non entrerei qui nelle ragioni, complesse, di questa situazione, ma è degno di rilievo il fatto che, almeno nella breve stagione delle enciclopedie multimediali su CD-ROM, l'editoria

elettronica ‘pura’, quella impegnata nella realizzazione di prodotti editoriali innovativi e nativamente digitali, aveva trovato proprio nel campo delle enciclopedie una sua strada di successo, non solo sostenibile ma – in quel particolare momento – redditizia.

La situazione cambierà nuovamente con la vera e propria rivoluzione rappresentata, più o meno in coincidenza con l’inizio del nuovo millennio, dal passaggio alle enciclopedie interamente online e, nel giro di pochi anni, dalla rapida affermazione di Wikipedia. A questa nuova svolta, che avrà conseguenze di enorme rilievo, è dedicato il prossimo capitolo.

¹ Ernie Smith, *How the CD-ROM Killed the Physical Encyclopedia*, in «TEDIUM», 11 settembre 2017, https://motherboard.vice.com/en_us/article/vbb994/how-the-cd-rom-killed-the-physical-encyclopedia.

² Wikipedia, voce ‘Grolier’, <https://en.wikipedia.org/wiki/Grolier>.

³ Massimo Bozzo, *La grande storia del computer. Dall’abaco all’intelligenza artificiale*, Dedalo, Bari 1996, p. 224.

⁴ *Ibid.*

⁵ In effetti, l’Italia può forse vantare un primato in questo senso: il De Italia, una sorta di rassegna enciclopedica sulla civiltà italiana prodotta su videodisco interattivo dalla Fondazione Agnelli, risale al 1987. Non si trattava però di un’enciclopedia generalista, bensì di un antenato dei libri elettronici interattivi. Per una presentazione dell’opera si veda *Un videodisco-laser per raccontare un paese*, articolo non firmato in «Lettera da TecnoCity» 3, 1987, p. 13, http://www.byterfly.eu/islandora/object/librib:628488/datastream/PDF/content/librib_628488.pdf.

⁶ Per le interessantissime vicende relative al rapporto tra l’Encyclopaedia Britannica e il mondo digitale, si vedano Shane Greenstein e Michelle Devereux, *The Crisis at Encyclopaedia Britannica*, in «Harvard Business Review – Case Studies», 1° gennaio 2006, <https://hbr.org/product/crisis-at-encyclopaedia-britannica/KEL251-PDF-ENG> (dietro paywall) e in forma di documento per la Kellogg School of Management at Northwestern University all’indirizzo <https://web.archive.org/web/20130313000034/http://www.kellogg.northwestern.edu/faculty/greenstein/images/htm/research/cases/encyclopaediabritannica.pdf> e Shane Greenstein, *The Reference Wars: Encyclopaedia’s Decline and Encarta’s Emergence*, in «Strategic Management Journal» 38, 5, pp. 995-1017, in rete all’indirizzo <https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2552>. Molte delle

informazioni che seguono sulla ‘guerra delle enciclopedie’ negli anni ’90 del secolo scorso sono tratte da questi articoli. Una panoramica dettagliata – con utili informazioni sui mercati europei – è offerta anche da Jeff Loveland, *The European Encyclopedia* cit., cap. 10.

⁷ Cfr. Ernie Smith, *How the CD-ROM Killed the Physical Encyclopedia* cit.

⁸ Shane Greenstein, *The Reference Wars* cit., p. 1009.

⁹ I nuovi proprietari cercarono di rilanciarla, ma con risultati deludenti: l’ultima edizione stampata è stata pubblicata nel 2010 e oggi la Britannica è disponibile solo nella versione online, con un abbonamento dal costo di circa 67 dollari all’anno. Dal 2007, la Britannica ha anche cercato di adottare un modello più vicino a Wikipedia, con la possibilità per gli utenti di integrare nuovi contenuti (comunque soggetti a revisione editoriale). Anche in questo caso, però, l’iniziativa ha avuto scarso successo.

¹⁰ Come vedremo attraverso l’esempio italiano, le versioni in lingua di Encarta prevedevano un’attenzione specifica alle culture dei paesi di destinazione. È direttamente Bill Gates a sottolineare l’importanza di questa componente della localizzazione di una enciclopedia in un articolo del 1997: Bill Gates, *The facts depend on where you are coming from*, in «The Sunday Times», 6 aprile 1997; una copia può essere visualizzata attraverso la ‘Wayback Machine’ dell’Internet Archive all’indirizzo

<https://web.archive.org/web/20120629032909/http://www.btimes.co.za/97/0406/tech/tech6.htm>.

¹¹ Riccardo Fedriga, comunicazione personale, 16 luglio 2022. Riccardo Fedriga è anche autore, con Giulio Carnazzi, dell’utile volumetto *Il mondo in un libro*, Sylvestre Bonnard, Milano 2002, che include una breve voce sulle enciclopedie su CD-ROM, interessante per il punto di vista relativo al momento della loro maggiore diffusione.

¹² Dati da Heather Hopkins, *Britannica 2.0: Wikipedia gets 97% of Encyclopedia Visits*, in «Hesperian Hitwise», 23 gennaio 2009, in rete all’indirizzo https://web.archive.org/web/20120326111318/http://weblogs.hitwise.com/us-heather-hopkins/2009/01/britannica_20_wikipedia_gets_9.html. Cfr. Noam Cohen, *Microsoft Encarta Dies After Long Battle with Wikipedia*, in «The New York Times», 30 marzo 2009, <https://bits.blogs.nytimes.com/2009/03/30/microsoft-encarta-dies-after-long-battle-with-wikipedia/>.

¹³ Luciano Floridi, *L’estensione dell’intelligenza. Guida all’informatica per filosofi*, Armando, Roma 1996, p. 193.

5.

Le enciclopedie online e la nascita di Wikipedia

Come già accennato, il passaggio dalle enciclopedie multimediali su CD-ROM alle versioni online è un processo non del tutto lineare: abbiamo visto come le primissime enciclopedie digitali – solo testuali e nate quando ancora non esistevano i CD-ROM – fossero concepite per l'uso a distanza, e come già nei primissimi anni dopo la nascita del world wide web (che inizia a crescere nel 1993 con la nascita del primo browser a larga diffusione, Mosaic) molte enciclopedie multimediali avessero intrapreso la strada dell'ibridazione tra CD-ROM (e successivamente DVD) e Internet.

Con il progressivo e rapido successo del web, si apriva il mercato delle enciclopedie a larga diffusione – e con la possibilità di includere contenuti multimediali, anche se inizialmente limitati dalla scarsa velocità dei collegamenti – esclusivamente online. Se il tentativo della Britannica nel 1995 era troppo ardito visto il numero di famiglie che all'epoca disponevano di una connessione a Internet, la trasformazione di Encarta in enciclopedia online nel 2000 rispecchia una situazione in cui il web sta diventando l'orizzonte naturale per tutte le ricerche di informazioni, e quindi anche per l'evoluzione dell'enciclopedismo. Il web, tuttavia, rivoluzionava non solo gli strumenti di accesso all'informazione, ma ha anche i modi e le forme di produzione dei contenuti informativi. E questo cambiamento ha coinvolto ben presto il mondo delle enciclopedie. Wikipedia, fondata nel 2001, è il principale esempio di questa evoluzione. Ma già nell'ottobre del 1993, quando il web era nella sua primissima infanzia, Rick Gates, uno dei pionieri nell'uso di Internet, propose l'idea di un'enciclopedia online costruita in modo collaborativo sfruttando le

potenzialità della rete. Questo primo progetto aveva un nome, Interpedia¹.

Bisogna però aspettare altri sette anni, fino al marzo 2000, per il progetto dalle cui costole sarebbe nata Wikipedia: un esperimento di enciclopedia online non commerciale denominata Nupedia. A proporlo, i due nomi che troveremo pochi mesi dopo anche all'origine di Wikipedia: Jimmy 'Jimbo' Wales e Larry Sanger. A differenza di quanto sarebbe accaduto con Wikipedia², però, Nupedia non era aperta alle modifiche pubbliche e si affidava ancora a redattori esperti, che accettavano di collaborare gratuitamente al progetto e di distribuire liberamente le voci create, sottoponendole a un processo pubblico di peer review, sempre tra esperti del settore. Il collegamento principale tra il progetto iniziale di Nupedia e Wikipedia risiede nell'idea di un'enciclopedia libera e aperta: un'idea appoggiata anche da Richard Stallman, il guru del software libero, che ben presto decise di far confluire in Nupedia un proprio progetto di enciclopedia online aperta, iniziato nel 1999 e chiamato GNUpedia³.



QR-Code 8

Wikiradio: la nascita di Wikipedia

<http://bit.ly/46sCZvX>

Le nuove possibilità di collaborazione online e i bassi costi di pubblicazione in rete avevano dunque aperto uno spazio nuovo, che non poteva esserci né nell'editoria su carta – in cui un progetto di tipo enciclopedico comportava costi di produzione assai alti e una infrastruttura redazionale e distributiva difficile da immaginare in contesti diversi da quello commerciale – né nelle primissime architetture mainframe-terminali che abbiamo ricordato nel terzo capitolo. Uno

spazio in cui potevano muoversi progetti collaborativi non commerciali, ispirati almeno in parte dall'associazione fra l'idea di 'personal empowerment' che aveva accompagnato la prima fase della rivoluzione dell'informatica personale e che abbiamo ricordato essere stata alla base della 'ideologia di Silicon Valley', e la 'network ideology', l'idea partecipativa che aveva caratterizzato l'età dei primi sistemi di accesso online a larga diffusione, e in particolare il mondo dei cosiddetti BBS (Bulletin Board System)⁴.

Nella sua forma iniziale, comunque, Nupedia non è andata lontano: sono state create relativamente poche voci, e il nome del progetto oggi è noto soprattutto perché da esso si è sviluppata Wikipedia. È infatti sul server di Nupedia che, il 10 gennaio 2001, viene installato il software Wiki che cinque giorni dopo, il 15 gennaio 2001, diventerà Wikipedia. Dei due fondatori, Jimbo Wales era un imprenditore che all'epoca del lancio di Nupedia prima, e di Wikipedia poi, gestiva un portale di ricerca di musica pop chiamato Bomis. Larry Sanger, invece, era stato scelto da Wales come managing editor di Nupedia. Negli anni successivi, Sanger lasciò Wikipedia, ritenendo deludenti i risultati del suo processo editoriale⁵, e fondò due nuovi progetti: Citizendium⁶, un progetto di enciclopedia online che cerca di fondere i consigli forniti da redattori esperti e la collaborazione aperta tra pari, ed Everipedia⁷, un'enciclopedia aperta che utilizza la tecnologia blockchain per tracciare la storia editoriale delle voci. Tra il 2019 e il 2020 ha lasciato entrambi i progetti e ha fondato la Knowledge Standards Foundation e un nuovo sito web, Encyclosphere, che mira a collegare in una rete decentralizzata tutti i progetti di enciclopedia basati sul web⁸. La peregrinazione di Sanger attraverso diversi progetti enciclopedici (e diverse idee di ciò che un'enciclopedia online aperta dovrebbe e potrebbe essere) è piuttosto interessante, anche se spesso ispirata o influenzata dalle sue idee politiche di destra, che lo vedono frequente ospite di Fox News per criticare le politiche editoriali di Wikipedia, che Sanger considera schierate a sinistra⁹. Fra le altre sue discutibili battaglie, l'opposizione ai vaccini contro il COVID e le critiche al campo dei women studies.

Ma torniamo a Wikipedia¹⁰. Che cosa significa esattamente 'Wikipedia'? L'iniziale 'Wiki' si riferisce al software usato per gestire il

sito web del progetto e il suo processo editoriale, mentre il suffisso ‘pedia’ – derivato dal greco *παιδεία* – richiama la natura enciclopedica del progetto. Il termine Wiki deriva dalla lingua hawaiana, dove ‘wikiwiki’ significa ‘molto veloce’: il bus navetta dell’aeroporto di Honolulu era chiamato ‘Wiki-Wiki’. Ward Cunningham, il creatore del primo sistema Wiki, decise di chiamare il suo software con questo nome per sottolineare che era facile e immediato da usare¹¹. In estrema sintesi, si tratta di un programma che risiede su un server e che permette di creare o modificare pagine e siti web in modo semplice e veloce scrivendo direttamente nel browser. La peculiarità del software Wiki – e ciò che lo distingue dagli altri sistemi di gestione dei siti web – è però la capacità di tenere traccia di tutte le modifiche apportate da ogni singolo utente del sistema, permettendo di ricostruire la storia editoriale di ogni voce e di tornare a una versione precedente della pagina o del testo, quando necessario. Il sistema Wiki è quindi uno strumento perfetto per scrivere pagine e siti in modo collaborativo.

Va sottolineato che il termine Wiki non si riferisce solo allo strumento editoriale utilizzato, ma anche ai siti web che vengono creati utilizzando quel software. Wikipedia è quindi un’enciclopedia creata utilizzando un software Wiki ed è essa stessa un esempio di sito Wiki.

Esistono diversi tipi di software Wiki: come già detto, il programma originale fu creato da Ward Cunningham nel 1995, ma oggi il programma più utilizzato per creare siti Wiki – compresa Wikipedia – si chiama MediaWiki¹². Si tratta di un software open source, quindi libero e liberamente modificabile, ed è stato sviluppato specificamente per Wikipedia nel 2002, anche se attualmente è adottato da molte altre piattaforme web. Il suo sviluppo è gestito dalla Wikimedia Foundation, una fondazione senza scopo di lucro creata negli Stati Uniti ma con sedi in diversi altri paesi, tra cui l’Italia; il suo scopo è quello di prendersi cura di Wikipedia e dei diversi progetti sviluppati intorno o in connessione con quello originale, anche aiutando a raccogliere i finanziamenti necessari¹³.

La piattaforma Wiki utilizzata da Wikipedia consente agli utenti di collaborare alla stesura delle voci dell’enciclopedia e di tenere traccia della storia di tutte le modifiche apportate a ciascuna voce. È proprio questa la

caratteristica principale che rende Wikipedia diversa dalle enciclopedie tradizionali. In un'enciclopedia tradizionale, l'autorevolezza, cioè il prestigio e l'affidabilità dell'enciclopedia, è una conseguenza dell'autorialità, cioè del fatto che le voci sono scritte da esperti riconosciuti in materia. In molti casi, le voci – o almeno quelle principali – sono firmate dagli autori, in modo da renderle riconoscibili. E anche quando non sono firmate, la redazione, composta da un numero ristretto di professionisti, garantisce l'autorevolezza e la validazione dei contenuti.

Tuttavia, nel caso di Wikipedia, l'autorità non deriva dall'autorialità, poiché le voci sono scritte collaborativamente da più persone e non sono redatte da un autore identificabile. Al contrario, l'autorità e la validazione dei contenuti sono il risultato di un processo di scrittura e revisione collaborativo.

Ma come può essere affidabile un'enciclopedia del genere, visto che è scritta da autori non identificabili, e che chiunque può correggere e modificare le voci a suo piacimento? Molti dubbi sull'effettiva affidabilità di Wikipedia derivano da questa (più che legittima) domanda, alla quale proverò a rispondere il prossimo capitolo.

¹ Sulle prime idee di enciclopedia online, si veda Joseph Michael Reagle, *Good Faith Collaboration: The Culture of Wikipedia*, MIT Press, Cambridge (MA) 2010, cap. 2 (su Interpedia, si vedano in particolare le pp. 32-34). Il testo del libro è disponibile online con licenza Creative Commons: <https://reagle.org/joseph/2010/gfc/>.

² Per un'analisi dei primi progetti di enciclopedia online e del processo che ha portato a Wikipedia, cfr. Benjamin Mako Hill, *Quasi Wikipedia: What Eight Early Online Collaborative Encyclopedia Projects Reveal about the Mechanisms of Collective Action*, in Id., *Essays on Volunteer Mobilization in Peer Production*, tesi di dottorato, Massachusetts Institute of Technology, Cambridge (MA) 2013; draft disponibile online: https://mako.cc/academic/hill-almost_wikipedia-DRAFT.pdf. Mako Hill 2013.

³ Questa parte della storia è ben ricostruita nel secondo capitolo di Joseph M. Reagle, *Good Faith Collaboration* cit.

⁴ Per un rapido sguardo di sintesi su questi temi, si veda Patrice Flichy, *Internet: The Social Construction of a "Network Ideology"*, in Olivier Coutard, Richard Eugene Hanley, Rae Zimmerman (a cura di), *Sustaining Urban Networks: The Social*

Diffusion of Large Technical Systems, Routledge, London-New York 2005, pp. 103-117. Ma chi non conosce il mondo dei BBS può forse farsene un'idea più vivace in due curiose raccolte di memorie, entrambe autopubblicate: Rob O'Hara, *Commodorok. Sordid Tales from the BBS Junkie*, Lulu Press 2006, e Lee Lovelace, *One Bit at a Time. A Memoir (The BBS Years)*, autopubblicato, 2021.

⁵ Cfr. Sophie Foggin, "Wikipedia is a broken system," says co-founder Larry Sanger, in «150 sec», 23 maggio 2019, <https://150sec.com/wikipedia-is-a-broken-system-says-co-founder-larry-sanger/11453/2017>.

⁶ Il sito web è <https://en.citizendium.org/>.

⁷ Il sito web è <https://everipedia.org/>.

⁸ Il sito web è <https://encyclosphere.org/>.

⁹ Cfr. ad es. Brian Flood, *Wikipedia co-founder Larry Sanger says online encyclopedia scrapped neutrality, favors lefty politics*, in «Fox News», 21 maggio 2020, <https://www.foxnews.com/media/wikipedia-co-founder-larry-sanger-says-online-dictionary-scrapped-neutrality-favors-lefty-politics2020>.

¹⁰ Tra le numerose pubblicazioni sulla natura e le caratteristiche di Wikipedia, un testo pionieristico di particolare rilevanza dal punto di vista della ricostruzione storica è Andrew Lih, *The Wikipedia Revolution*, Hyperion, New York 2009, trad. it. *La rivoluzione di Wikipedia*, Codice, Torino 2010. Altro riferimento obbligato, in questo caso assai più recente, è Joseph Reagle e Jackie Koerner (a cura di), *Wikipedia @20. Stories of an Incomplete Revolution*, MIT Press, Cambridge (MA) 2020, che offre una panoramica a molte voci della storia, dello stato e delle prospettive future di Wikipedia a vent'anni dalla sua nascita. Quest'ultimo testo è disponibile integralmente online con licenza Creative Commons: <https://direct.mit.edu/books/oa-edited-volume/4956/Wikipedia-20Stories-of-an-Incomplete-Revolution>. Un testo di sintesi in italiano (che ha però alcuni limiti bibliografici, anche relativamente al nostro paese) è il recente Oscar Ricci, *L'ultima comunità. I primi vent'anni di Wikipedia*, Ledizioni, Milano 2022.

¹¹ Per un'introduzione generale alle idee principali del software Wiki, il testo di riferimento è Bo Leuf e Ward Cunningham, *The Wiki Way: Quick Collaboration on the Web*, Addison-Wesley, Boston 2001.

¹² Per un'introduzione a MediaWiki si vedano Daniel J. Barrett, *MediaWiki*, O'Reilly, Sebastopol (CA) 2009 e Yaron Koren, *Working with MediaWiki* (seconda edizione), CreateSpace Independent Publishing Platform, 2017.

¹³ Il sito web è <https://wikimediafoundation.org/>.

6.

Wikipedia funziona?

Per dare una risposta alla domanda che dà il titolo a questo capitolo, possiamo partire dalla questione fondamentale, che è proprio quella dell'affidabilità. Wikipedia è *davvero* affidabile? Sorprendentemente, e con particolare riferimento alla versione più ampia di Wikipedia, quella con il maggior numero di voci e con la comunità più ampia di collaboratrici e collaboratori – ovviamente quella inglese –, la risposta è ragionevolmente positiva. Ormai quasi vent'anni fa, nel 2005, quando Wikipedia era ancora un'impresa appena nata, la prestigiosa rivista scientifica «Nature» pubblicò una ricerca basata su un campione di 42 voci scientifiche tratte sia da Wikipedia sia dall'Encyclopaedia Britannica (evidentemente considerata il paradigma dell'enciclopedia tradizionale e validata). Le voci selezionate erano state esaminate 'alla cieca' da un gruppo di esperti che non sapevano quale versione della voce fosse stata presa da Wikipedia e quale dall'Encyclopaedia Britannica. Il risultato del confronto fu che la divergenza tra le due fonti in termini di qualità e affidabilità era minima (anche se la Britannica manteneva un leggero vantaggio)¹. Su questo risultato si è aperto un dibattito infinito², e negli anni successivi sono stati effettuati numerosi confronti simili e di analisi di affidabilità su Wikipedia, con conclusioni non sempre convergenti³.

Il dibattito che ne è scaturito è complesso e interessante, ma non lo analizzerò nel dettaglio, per non andare oltre lo scopo di questo capitolo; cercherò piuttosto di fornire una sintesi concisa delle sue conclusioni, con tutti i limiti che questo può comportare. Il quadro generale è il seguente: non possiamo essere del tutto certi dell'accuratezza di ogni informazione tratta da Wikipedia (questo vale anche per un'enciclopedia tradizionale, ma il grado di potenziale imprecisione dei contenuti di Wikipedia è

maggiore: in sostanza, gli errori possono essere molto più gravi). Nel complesso, però, la maggior parte delle informazioni fornite da Wikipedia è ragionevolmente affidabile, e il tasso di affidabilità cresce in base alla quantità di interventi editoriali sulla voce e al numero di utenti coinvolti nella sua stesura e revisione⁴. Il fatto che Wikipedia abbia un carattere dinamico e sia costantemente aggiornata rende possibile l'inserimento di errori in qualsiasi momento; ma grazie a questo carattere dinamico, gli errori vengono di solito corretti tempestivamente. Inoltre, la copertura di Wikipedia è oggi molto più ampia di qualsiasi enciclopedia tradizionale, in particolare per quanto riguarda alcuni settori (tra cui quelli legati alla cosiddetta 'cultura popolare', ovvero serie televisive, film, fumetti...). Quindi, Wikipedia ha una copertura per molti versi 'eccentrica', ma molto più estesa dei modelli enciclopedici tradizionali.

Resta comunque da rispondere alla domanda iniziale: come mai un'enciclopedia in cui chiunque può contribuire modificando in qualunque modo qualunque voce risulta essere ragionevolmente affidabile e corretta?

La risposta risiede paradossalmente proprio nell'elevato numero di collaboratori di Wikipedia, associato alla capacità della piattaforma Wiki di tenere traccia della storia delle revisioni di ogni voce. Naturalmente, ci sono i cosiddetti 'vandali' o 'troll', che cercano deliberatamente di inserire disinformazione. Ma il sistema consente di individuare immediatamente le modifiche più recenti apportate a ciascuna voce e di correggerle tornando alla versione precedente (procedura di 'revert') o modificando il testo, quando necessario; inoltre, è possibile bloccare l'indirizzo di rete del computer da cui sono stati aggiunti contenuti errati e fuorvianti, mentre una voce può essere temporaneamente 'bloccata', se necessario, impedendone ulteriori modifiche per un determinato periodo di tempo: una funzione che può essere utilizzata quando voci controverse sono oggetto di trolling o di un attacco di parte.

Questo spiega anche perché la versione inglese di Wikipedia è mediamente più completa e affidabile di quelle in altre lingue. All'interno di uno strumento la cui autorevolezza dipende principalmente dal meccanismo di scrittura e revisione collaborativa, maggiore è il numero

di collaboratori, più alta è la probabilità di avere al loro interno persone con le competenze necessarie per migliorare la qualità della voce.

Ed è proprio questa collaborazione allargata che aiuta a risolvere un altro problema che, in linea di principio, potrebbe ostacolare seriamente la realizzazione di un'enciclopedia collaborativa aperta: la neutralità dei punti di vista. Come possiamo assicurarci che una voce, e in particolare una voce su un argomento dibattuto, non sia soggetta a pregiudizi dovuti al punto di vista personale dell'autore? Si pensi alle opinioni divergenti che potrebbero esistere sul conflitto arabo-israeliano tra collaboratori o collaboratrici israeliani e palestinesi. Ebbene, in un certo senso, il meccanismo di Wikipedia *obbliga* queste persone a collaborare attraverso la negoziazione redazionale della voce. Naturalmente, la completa neutralità è un obiettivo irraggiungibile e sarebbe ingenuo pensare che qualsiasi voce di qualsiasi enciclopedia, compresa Wikipedia, possa essere completamente neutrale e obiettiva. Potremmo anzi ragionevolmente sostenere che gli ideali di completa neutralità e obiettività non solo sono praticamente irraggiungibili, ma non esistono affatto. Alle collaboratrici e ai collaboratori di Wikipedia viene però sostanzialmente chiesto di 'far finta' che la neutralità e l'obiettività esistano⁵, con la sollecitazione a perseguire questo obiettivo, che può sostanzialmente essere considerato come una sorta di ideale regolativo. Questo è uno dei cinque *pilastr*i ('pillars'), i principi fondamentali, su cui si basa Wikipedia⁶.

Può essere utile ricordare questi principi, dato che sono le regole di livello superiore che guidano la stesura e la redazione collaborativa delle voci:

1. *Wikipedia è un'enciclopedia*. Questo implica che le voci devono essere riconosciute come 'enciclopediche' – e quindi rilevanti e di interesse generale – dalla comunità costituita da chi vi collabora. Ciò significa anche che chi collabora a Wikipedia non può inserire una voce su di sé o su proprie conoscenze, a meno che non si tratti di persone tanto note o importanti da poter attribuire loro una rilevanza enciclopedica.

Va osservato che l'assenza di limiti di spazio, dovuta al fatto che Wikipedia è interamente digitale e online, e il numero molto elevato di persone che collaborano al canale, consentono un'interpretazione piuttosto ampia di questi principi. Molte voci – soprattutto nelle versioni

linguistiche di Wikipedia diverse dall'inglese – sono dedicate a persone o argomenti di rilevanza solo settoriale o discutibile. Sebbene ciò possa essere considerato una violazione del primo pilastro, si tratta anche di un modo per aumentare la copertura complessiva dell'enciclopedia. Nel contempo, l'utilizzo di un'interpretazione piuttosto 'debole' del primo pilastro permette a Wikipedia di coprire una gamma di argomenti significativamente più ampia di quella che un'enciclopedia tradizionale potrebbe mai permettersi di prendere in considerazione a causa dello spazio limitato e del numero limitato di collaboratori.

D'altro canto, l'adizione di un concetto di 'enciclopedismo' evidentemente e comunque più largo di quello tradizionale pone un problema su cui dovremo tornare in seguito: Wikipedia, proprio in quanto enciclopedia, si propone come strumento per organizzare *conoscenze*, non come semplice deposito organizzato di informazioni. Ma il confine fra le due categorie, in un ecosistema informativo come quello del web, non è sempre semplice da tracciare.

2. *Wikipedia è scritta da un punto di vista neutrale* (è il cosiddetto 'neutral point of view', spesso abbreviato attraverso l'acronimo NPOV). Secondo quanto dichiarato da Wikipedia a questo proposito (cito da Wikipedia in lingua inglese, la cui politica editoriale è – per i motivi già discussi – quella di maggior rilievo):

Ci sforziamo di scrivere articoli dal tono imparziale che documentino e spieghino i principali punti di vista, dando loro il giusto peso. Evitiamo la difesa di singole posizioni, ed esponiamo le informazioni e i problemi anziché discuterli. In alcune aree può esistere un solo punto di vista ben riconosciuto; in altre, descriviamo più punti di vista, presentandoli in modo accurato e contestualizzato piuttosto che come «la verità» o «il punto di vista migliore». Tutti gli articoli devono tendere a un'accuratezza verificabile, citando fonti affidabili e autorevoli, soprattutto quando l'argomento è controverso o riguarda una persona vivente. Le esperienze personali, le interpretazioni o le opinioni dei redattori non hanno spazio in Wikipedia⁷.

Come osservato in precedenza, il 'punto di vista neutrale' opera fondamentalmente come un ideale regolativo; il compito di risolvere gli inevitabili disaccordi sull'applicazione concreta del principio è affidato alla negoziazione redazionale ed è facilitato da una specifica pagina di discussione, che esiste per ogni voce ed è destinata allo scambio di

opinioni tra chi collabora alla stesura della voce. La pagina di discussione – raggiungibile dalla scheda ‘Discussione’ presente in ogni pagina del sito – rappresenta, insieme alla gestione delle versioni, uno degli strumenti fondamentali di Wikipedia: gli utenti occasionali la visitano raramente e spesso ne ignorano completamente l’esistenza, mentre collaboratrici e collaboratori la utilizzano molto spesso⁸.

Vale la pena ricordare che fra gli antenati del NPOV viene spesso citata proprio l’idea di H.G. Wells di enciclopedia mondiale come costruzione universalistica e in qualche misura ‘oggettiva’ di cui ho parlato nel primo capitolo⁹.

3. *Wikipedia è un contenuto libero che chiunque può utilizzare, modificare e distribuire.* Questo principio esprime gli ideali di collaborazione aperta e di condivisione dei contenuti che abbiamo già visto costituire una delle basi del progetto. In termini legali, corrisponde all’uso della licenza Creative Commons Attribution – ShareAlike 3.0 e alla GNU Free Documentation License (GFDL), utilizzate per la maggior parte dei contenuti di Wikipedia¹⁰.

4. *Wikipedia ha un codice di condotta.* Questo principio raccomanda il rispetto reciproco tra le persone che collaborano alla realizzazione dell’enciclopedia, impegnate in un progetto cooperativo e non conflittuale; potrebbe essere considerato come un adattamento da parte del progetto Wikipedia del principio di carità o della massimizzazione dell’accordo¹¹:

Rispettate gli altri wikipediani, anche quando non siete d’accordo con loro. Applicate il galateo di Wikipedia ed evitate attacchi personali. Cercate il consenso, evitate le guerre di modifica e non intervenite mai sui contenuti di Wikipedia per difendere una posizione. Agite in buona fede e presumete la buona fede degli altri. Siate aperti e accoglienti con i nuovi arrivati. Se dovessero sorgere conflitti, affrontateli con calma nelle apposite pagine di discussione e seguite le procedure di risoluzione delle controversie¹².

5. *Wikipedia non ha regole rigide.* È interessante osservare che prima del 27 settembre 2010 questa affermazione era scritta nella forma di un tipico principio di chiusura logica: «Wikipedia non ha regole rigide oltre ai cinque principi generali qui presentati»¹³. La revisione di questa regola e la scelta di abbandonare la clausola che escludeva i cinque ‘pillars’ dalla

flessibilità prevista per tutte le altre scelte del progetto – scelta effettuata dopo un vivace dibattito tra i collaboratori – indebolisce chiaramente lo status dei cinque pilastri, che vanno quindi considerati essi stessi come principi regolatori piuttosto che come norme rigide, immutabili e prescrittive.

Va anche notato che la formulazione dei cinque pilastri non è necessariamente la stessa per tutte le versioni nazionali di Wikipedia. Da questo punto di vista, il progetto generale di Wikipedia dovrebbe probabilmente essere considerato come basato su una costellazione di politiche editoriali leggermente diverse e in evoluzione, piuttosto che come un modello unico, uniformemente adottato e ben definito¹⁴.

Queste regole sono sufficienti a garantire il successo di un'impresa enciclopedica che opera in una forma spesso considerata 'amatoriale'? In realtà, considerare Wikipedia come un'enciclopedia sostanzialmente amatoriale è oggi probabilmente sbagliato. Il successo di Wikipedia, sia nel numero di utenti, sia nell'affidabilità della maggior parte delle voci, è legato da un lato – come abbiamo visto – alla qualità dei meccanismi di negoziazione editoriale offerti dalla piattaforma Wiki: Wikipedia si avvale infatti di strumenti tecnici molto robusti (come ulteriore esempio si può citare l'uso dei bot, ovvero applicazioni della piattaforma che correggono automaticamente gli errori formali e aiutano a standardizzare le scelte stilistiche). D'altro canto, bisogna considerare che l'idea diffusa di un'enciclopedia interamente costruita da 'dilettanti' è ormai inesatta. Tra i collaboratori di Wikipedia, molti sono ragionevolmente esperti e, man mano che l'enciclopedia ha aumentato la sua base di utenti e ha migliorato la sua reputazione, è anche aumentato il numero di esperti che la utilizzano e che possono contribuire a migliorare la qualità delle voci relative ai campi di loro competenza. In molti casi, questo avviene anche attraverso pagine che servono a coordinare le attività redazionali di collaboratrici e collaboratori esperti su specifiche tematiche, favorendo la partecipazione anche delle relative associazioni professionali (un buon esempio è WikiProject Medicine, il progetto specifico relativo all'ambito medico¹⁵).

Fra i settori in cui si è diffusa l'abitudine a promuovere contributi competenti a Wikipedia, con un conseguente miglioramento qualitativo

dei suoi contenuti, vanno ricordati quelli, particolarmente importanti rispetto al tema di questo libro, della biblioteconomia e dell'istruzione. In entrambi i casi, si è assistito allo sviluppo di iniziative e progetti specifici (come «Wikipedia loves libraries»¹⁶ o «Wikipedia va a scuola»¹⁷) volti a promuovere una migliore comprensione di Wikipedia come strumento e a favorire la qualità della collaborazione nella redazione delle voci. Per quanto riguarda il mondo delle biblioteche italiane, le ragioni di questo interesse sono efficacemente riassunte in due post appassionati sui blog di Virginia Gentilini ed Eusebia Parrotto; può essere utile riportare qualche passo dei loro interventi:

Wikipedia è generalmente indicato come il sesto sito più consultato al mondo. Fronteggia quasi alla pari mostri come Google e Facebook. Evidentemente le persone in rete non cercano solo gattini o porno, ma conoscenze complesse. Si tratta ancora di un altro sogno realizzato. Se ci interessano gli utenti, ci interessano Wikipedia e il fatto che sia un'enciclopedia di qualità. La maggioranza dei nostri utenti non utilizzerà altra risorsa di reference per il resto della sua vita¹⁸.

Se i nostri utenti non vengono da noi, ma vanno su Wikipedia, allora su Wikipedia dobbiamo andarci anche noi. Fa parte del nostro lavoro, della nostra missione. Non basta frequentare un corso di aggiornamento sulle risorse informative in rete e sapere che esiste: Wikipedia deve essere anche la 'nostra' biblioteca, il luogo in cui esercitiamo attivamente la nostra professione. La cura con cui sceglievamo le enciclopedie cartacee, ora dovremmo usarla nel verificare le voci di Wikipedia, e nell'arricchirle non solo di contenuti, ma anche e soprattutto di fonti. Le fonti sono in biblioteca, intorno a noi. Sono i libri che abbiamo acquistato, le collezioni di periodici che conserviamo, le risorse di qualità che il nostro approccio consapevole e informato alla rete Internet ci permette di conoscere. Le troviamo ad occhi chiusi, se vogliamo. È il nostro lavoro. [...] La prossima volta che cercando su Wikipedia ci scandalizzeremo della povertà o inaffidabilità di una voce, rimpiangendo i tempi non lontanissimi in cui il sapere era rinchiuso dentro testi autorevoli, alziamoci, andiamo in sala o nei depositi e cerchiamo un testo 'autorevole'. Poi torniamo su Wikipedia, clicchiamo su 'modifica' e aggiorniamo la voce¹⁹.

Un discorso del tutto analogo può essere fatto per quanto riguarda l'uso di Wikipedia nelle scuole e nelle università e il ruolo degli insegnanti, ancora una volta in termini di miglioramento delle competenze degli utenti e della qualità delle voci²⁰. Per fare un esempio tra i tanti disponibili²¹, presso la scuola di medicina dell'Università della California, gli studenti dell'ultimo anno sono stati impegnati in attività guidate dai

docenti (e valutate come parte del programma di studi) volte a migliorare la qualità delle voci mediche²². Questo interesse è pienamente giustificato se si considera che già nel 2014 il 50% dei medici americani utilizzava Wikipedia come strumento di riferimento per informazioni su argomenti medici²³.

In poche parole, è vero che le voci di Wikipedia non sono scritte da singoli esperti riconoscibili, ma è anche vero che un numero crescente di esperti o almeno di utenti ragionevolmente competenti e motivati collaborano regolarmente per migliorare la loro accuratezza.

Naturalmente, data la natura stessa di Wikipedia, questo non garantisce che si evitino errori occasionali, così come problemi (in genere temporanei) legati al vandalismo di singole voci. Ma più preoccupanti e insidiosi, anche perché rischiano di essere più diffusi e sistematici, sono i potenziali problemi legati a orientamenti e pregiudizi di chi collabora al progetto. Pregiudizi che gli ideali regolativi rappresentati dai cinque ‘pillars’ non sembrano sufficienti a evitare.

Si è così discusso molto – per partire da uno degli esempi più rilevanti – sull’esistenza di pregiudizi di genere nelle voci di Wikipedia, in parte dovuti alla sproporzione tra i collaboratori di sesso maschile (che sono ancora la maggioranza) e le collaboratrici di sesso femminile o con una diversa identità sessuale o di genere²⁴. Altri pregiudizi sono stati individuati nella sovrarappresentazione della cultura occidentale e nella sottorappresentazione delle culture non occidentali o emarginate²⁵, o nella possibile esistenza di interessi commerciali, politici o ideologici dietro le attività di alcuni dei collaboratori²⁶, nonché nella possibilità che anche le stesse procedure utilizzate per la revisione editoriale collaborativa possano portare a risultati distorti²⁷.

Sarà interessante, nella parte seconda del libro, osservare come pregiudizi assai simili riguardino oggi anche i corpora su cui sono addestrate le intelligenze artificiali generative. Tanto l’architetto quanto l’oracolo portano da questo punto di vista il peso di un’eredità comune: quella di un passato – e per troppi versi di un presente – visto con occhiali distorti perché segnato, già nel suo svolgersi, da sistematici squilibri di potere.

Va però anche tenuto presente che Wikipedia offre uno spazio negoziale

in cui questi pregiudizi possono essere esaminati, discussi, e almeno in parte corretti: proprio il carattere collaborativo e aperto del progetto offre strumenti per ridurre alcuni squilibri, a condizione però che lo spazio di collaborazione sia sufficientemente ampio e frequentato, e sia gestito con politiche adeguate.

Katherine Maher, che fra il 2016 e il 2021 è stata direttrice esecutiva della Wikimedia Foundation, ha affrontato sinteticamente, in una conferenza TED del 2022, molti dei temi che ho discusso in questo capitolo: la potete ascoltare – volendo, anche con i sottotitoli in italiano – partendo dal codice QR qui di seguito. Le sue idee possono probabilmente essere considerate troppo idealiste e ottimiste, ma sono una buona espressione di quello che è il ‘pregiudizio’ necessario e fondamentale alla base dell’intero edificio wikipediano: l’idea che la maggior parte delle persone che collaborano al progetto siano bene intenzionate e in buona fede, e che la loro collaborazione possa contare su strumenti di negoziazione redazionale effettivamente capaci di ridurre progressivamente i disaccordi, e nella maggior parte dei casi di raggiungere un ragionevole consenso. Se questo corrispondesse alla realtà, la capacità di Wikipedia di costruire e rappresentare complessità ne uscirebbe assai rafforzata: uno dei pochi casi in cui l’utopia collaborativa condivisa da molti fra i pionieri della rete potrebbe aver prodotto un risultato almeno in parte all’altezza delle aspettative, e dell’utopica idea di enciclopedia mondiale proposta da H.G. Wells.



QR-Code 9

TED Talk di Katherine Maher su Wikipedia

<http://bit.ly/3r5veMn>

¹ Si veda Jim Giles, *Internet Encyclopedias Go Head to Head*, in «Nature» 439, 14 dicembre 2005, pp. 900-901, <https://www.nature.com/articles/438900a> (nella stessa pagina web del documento, è possibile trovare anche un link alle obiezioni sollevate dalla redazione dell'Encyclopaedia Britannica e alle relative risposte di «Nature»).

² Molto meno favorevoli a Wikipedia sono i risultati dell'analisi condotta in Lucy Holman Rector, *Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles*, in «Reference Services Review» 36.1, 2008, pp. 7-22, <https://www.emeraldinsight.com/doi/pdfplus/10.1108/00907320810851998>.

Naturalmente i dati su cui si basava il documento, risalenti a oltre quindici anni fa, sono ormai da considerarsi piuttosto datati.

³ Per l'interesse suscitato e dato che si tratta di un lavoro più recente con ampi riferimenti bibliografici, un esempio degno di nota è Shane Greenstein e Feng Zhu, *Do Experts or Crowd-based Models Produce More Bias? Evidence from Encyclopaedia Britannica and Wikipedia*, in «MIS Quarterly» 42.3, 2018, pp. 945-959, <http://fengzhu.info/BritannicaWikipedia.pdf>. Si veda anche, con specifico riferimento alle differenze nelle modalità di riferimento a realtà aziendali, Marcus Messner e Marcia W. DiStaso, *Wikipedia versus Encyclopedia Britannica: A Longitudinal Analysis to Identify the Impact of Social Media on the Standards of Knowledge*, in «Mass Communication and Society» 16.4, 2013, <https://www.tandfonline.com/doi/full/10.1080/15205436.2012.732649>, e con riferimento alla credibilità percepita delle voci, Andrew J. Flanagin e Miriam J. Metzger, *From Encyclopaedia Britannica To Wikipedia: Generational differences in the perceived credibility of online encyclopedia information*, in «Mass Communication and Society» 14.3, 2011, <https://www.tandfonline.com/doi/full/10.1080/1369118X.2010.542823>.

⁴ Cfr. Zhongming Ma, Jie Tao, Jing Hu, *The dynamics of Wikipedia article revisions: an analysis of revision activities and patterns*, in «International Journal of Data Mining, Modeling and Management» 9.4, 2017, pp. 298-314, <https://www.inderscienceonline.com/doi/pdf/10.1504/IJDM.2017.0884152017>. Si veda anche il dettagliato esame del processo di revisione a tre articoli di Wikipedia contenuto in Lara Tellis, *The Future of Editing: A Textual Analysis of Three Wikipedia Articles*, TigerPrint (Clemson University All Theses 973), Clemson (SC) 2010, https://tigerprints.clemson.edu/all_theses/973/.

⁵ Sarebbe preferibile che Wikipedia riconoscesse esplicitamente la natura di ideale puramente regolativo del 'neutral point of view'. L'evoluzione storica della formulazione del secondo pilastro, che può essere ricostruita dalla pagina delle versioni di https://en.wikipedia.org/wiki/Wikipedia:Five_pillars, sembra andare (lentamente) in questa direzione. La formulazione originale, più forte, potrebbe

essere stata influenzata dalla filosofia ‘oggettivista’ di Ayn Rand, sostenuta da Jumbo Wales: si veda al riguardo Andrew Lih, *The Wikipedia Revolution* cit., cap. 2.

⁶ I cinque pilastri sono descritti alla pagina https://en.wikipedia.org/wiki/Wikipedia:Five_pillars per Wikipedia in inglese e https://it.wikipedia.org/wiki/Wikipedia:Cinque_pilastri per Wikipedia in italiano. Sulle differenze fra le versioni nazionali di Wikipedia mi soffermerò in seguito.

⁷ Tradotto da https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

⁸ Sul funzionamento della pagina di discussione di Wikipedia si vedano David Laniado, Riccardo Tasso e Yana Volkovich, *When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages*, intervento presentato alla Fifth International AAAI Conference on Weblogs and Social Media, 17-21 luglio 2011, <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPaper/2764>, e Mirko Tavoranis, *Le discussioni di Wikipedia*, in «Lingue e Culture dei Media» 4.2, 2020, <https://riviste.unimi.it/index.php/LCdM/article/view/14333>.

⁹ Joseph Michael Reagle, *Good Faith Collaboration* cit., cap. 2.

¹⁰ Cfr. Wikipedia, voce ‘Copyrights’, <https://en.wikipedia.org/wiki/Wikipedia:Copyrights>. Sulla natura aperta di Wikipedia e i suoi rapporti con il movimento Open Access si veda Nathaniel Tkacz, *Wikipedia and the Politics of Openness*, University of Chicago Press, Chicago 2014, interessante anche per l’analisi concreta e non mitizzata dell’idea di accesso aperto e l’analisi delle radici di tale idea nelle ideologie libertarie e utopistiche della Silicon Valley.

¹¹ Sul principio di carità si veda Richard Feldman, *Charity, principle of*, in Peter Klein e Richard Foley (a cura di), *Routledge Encyclopedia of Philosophy*, Routledge, Abingdon 1998, <https://doi.org/10.4324/9780415249126-P006-1>.

¹² Wikipedia, voce ‘Wikipedia:Five_pillars’, https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

¹³ Cfr. la scheda di storia delle versioni, *ibid.* Una versione precedente della pagina riportava la formulazione logicamente errata «Wikipedia non ha regole rigide oltre alle quattro affermazioni generali di cui sopra». Così formulata, la chiusura non era chiaramente efficace, poiché la quinta regola non era inclusa. La storia della revisione del quinto pilastro è un esempio interessante dell’idea che le politiche e le linee guida di Wikipedia non siano scolpite nella pietra, una tesi che fa parte del contenuto stesso di questo pilastro.

¹⁴ Questo vale anche se si considera Wikipedia dal punto di vista degli utenti di diversi paesi, tenendo conto delle loro diverse esigenze e comportamenti di lettura: cfr. Florian Lemmerich et al., *Why the World Reads Wikipedia: Beyond English Speakers*, in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM ‘19)*, Association for Computing Machinery, New York 2017, pp. 618-626, in rete all’indirizzo <https://doi.org/10.1145/3289600.3291021>.

Sull'idea di una 'etnografia' di Wikipedia, cfr. Dariusz Jemielniak, *Common Knowledge? An Ethnography of Wikipedia*, Stanford University Press, Stanford (CA) 2014, e, da un punto di vista linguistico, Mirko Tavosanis, *Le discussioni di Wikipedia* cit.; volendo dare uno sguardo alla storia e alle caratteristiche di Wikipedia da un punto di vista diverso da quello inglese o italiano, si può vedere Pavel Richter, *Die Wikipedia-Story: Biografie eines Weltwunders*, Campus Verlag, Frankfurt a.M. 2020 (l'autore è stato a lungo direttore esecutivo di Wikimedia Deutschland). Di etnografia digitale torneremo a parlare nel capitolo 17, affrontando il tema delle memorie personali nelle comunità virtuali.

¹⁵ Cfr. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine.

¹⁶ La pagina generale del progetto è https://outreach.wikimedia.org/wiki/Wikipedia_Loves_Libraries; sul tema cfr. Kathleen de la Peña McCook, *Librarians as Wikipedians: From Library History to 'Librarianship and Human Rights'*, in «Progressive Librarian» 42, 2014, pp. 61-81, <http://www.progressivelibrariansguild.org/PL/PL42/061.pdf>; Anna Lucarelli, «Wikipedia loves libraries»: in Italia è un amore corrisposto..., in «AIB Studi» 54, 2-3, 2014, pp. 241-259, <http://aibstudi.aib.it/article/view/10108/10146>; Merrilee Proffitt (a cura di), *Leveraging Wikipedia: Connecting Communities of Knowledge*, ALA Editions, Chicago 2018; Phoebe Ayers, *Wikipedia and Libraries*, in Joseph Reagle e Jackie Koerner (a cura di), *Wikipedia @20* cit., pp. 98-106.

¹⁷ Un progetto italiano: <https://wiki.wikimedia.it/wiki/Scuola>.

¹⁸ Virginia Gentilini, *Il sogno realizzato del bibliotecario e che cosa fare ora*, nel blog «Bibliotecari non bibliofili», 1° gennaio 2015, <https://nonbibliofili.wordpress.com/2015/01/01/il-sogno-realizzato-del-bibliotecario-e-che-cosa-fare-ora/>.

¹⁹ Eusebia Parrotto, *Wikipedia, un posto per bibliotecari*, nel blog «Librarianscape», 9 gennaio 2016, <https://librarianscape.com/2016/01/09/wikipedia-un-posto-per-bibliotecari/>. Tra i numerosi contributi e post di blog dedicati al tema della collaborazione tra Wikipedia e biblioteche, si vedano in particolare John Lubbock, *Wikipedia and libraries*, in «Alexandria» 28.1, 2018, pp. 55-68, <https://journals.sagepub.com/doi/abs/10.1177/0955749018794968>; Phoebe Ayers e Andrea Zanni, *Wikipedia e biblioteche: una prospettiva globale*, in «AIB Studi» 57.1, 2017, pp. 119-125; <http://aibstudi.aib.it/article/view/11560/10889>; Luigi Catalani, *Biblioteche e Wikimedia: strategie comuni per l'accesso aperto alla conoscenza e la costruzione collaborativa del sapere libero*, in «JLIS.it» 8.3, 2017, pp. 100-114, <https://www.jlis.it/article/view/12413/11292>; Luigi Catalani e Pierluigi Feliciati (a cura di), *Wikipedia, Libraries and Archives*, in «JLIS.it» 9.3, 2018, <https://www.jlis.it/index.php/jlis/issue/view/9> (un numero monografico di «JLIS.it» su Wikipedia, biblioteche e archivi; tra gli articoli inclusi, di particolare rilevanza dal nostro punto di vista è Eusebia Parrotto, *Wikimedia projects in public libraries. Feasible ways*, ivi, pp. 132-140). Il tema è stato discusso nel corso di un

workshop organizzato dalla Biblioteca Nazionale di Firenze il 10 novembre 2017; per una sintesi, si veda Luigi Catalani, *Diffondere la conoscenza aperta per crearne di nuova. Resoconto della giornata “Sfide e alleanze tra Biblioteche e Wikipedia”* (Biblioteca Nazionale Centrale di Firenze, 10 novembre 2017), in «Bibliothecae.it» 7.1, 2018, pp. 391-404 e Bianca Gai, *National Meeting ‘Sfide e alleanze tra Biblioteche e Wikipedia. Edition 2017’* (National Central Library of Florence, 10th November 2017), in Luigi Catalani e Pierluigi Feliciati (a cura di), *Wikipedia, Libraries and Archives* cit., pp. 167-170.

²⁰ Sull’uso di Wikipedia a scuola, si veda Luigi Catalani et al., *Fare didattica con i progetti Wikimedia*, in «Bricks» 4, 2017 (numero monografico), <http://www.rivistabricks.it/2017/12/19/n-4-2017-fare-didattica-con-i-progetti-wikimedia/> e Luigi Catalani, *I progetti Wikimedia per l’apprendimento delle competenze informative e digitali in biblioteca, a scuola, nelle università*, in «AIB Studi» 57.2, 2017, pp. 253-263, <http://aibstudi.aib.it/article/view/11654/10993>. Per una interessante rassegna di pratiche internazionali rilevanti, si veda Andrew David Virtue, *Wikipedia as a Pedagogical Tool*, in «Wikistudies» 1.1, 2017, [https://wikistudies.org/index.php?](https://wikistudies.org/index.php?journal=wikistudies&page=article&op=view&path%5B%5D=1)

[journal=wikistudies&page=article&op=view&path%5B%5D=1](https://wikistudies.org/index.php?journal=wikistudies&page=article&op=view&path%5B%5D=1). Per una discussione sull’uso di Wikipedia nel contesto della metodologia ‘flipped classroom’, si veda Di Zou et al., *Flipped learning with Wikipedia in higher education*, in «Studies in Higher Education» 45.5, 2020, pp. 1026-1045, <https://doi.org/10.1080/03075079.2020.1750195>. Per una discussione non aggiornata ma comunque interessante sui pro e i contro dell’uso di Wikipedia nelle università, si veda Klaus Wannemacher, *Articles as Assignments – Modalities and Experiences of Wikipedia Use in University Courses*, in M. Spaniol, Q. Li, R. Klamma e R.W.H. Lau (a cura di), *Advances in Web Based Learning – ICWL 2009* («Lecture Notes in Computer Science» 5686), Springer, Berlin-Heidelberg 2009, pp. 434-443; sul tema si veda anche Robert E. Cummings, *The First Twenty Years of Teaching with Wikipedia: From Faculty Enemy to Faculty Enabler*, in Joseph Reagle e Jackie Koerner (a cura di), *Wikipedia @20* cit., pp. 141-149. Per interessanti esperienze basate sull’uso di Wikipedia nel contesto di corsi di scrittura universitari, si veda Mirko Tavosanis, *Insegnamento universitario della scrittura 2.0 attraverso Wikipedia*, in *Tecnologie e metodi per la didattica del futuro. Atti della 27a DIDAMATICA*, CNR, Pisa 2013, pp. 407-410 e Mirko Tavosanis, *Scrivere su Wikipedia dall’università alla scuola*, in Massimo Palermo ed Eugenio Salvatore (a cura di), *Scrivere nella scuola oggi. Obiettivi, metodi, esperienze*, Cesati, Firenze 2019, pp. 173-182. Per una discussione e una bibliografia aggiornate, si veda Mirko Tavosanis, *L’italiano di Wikipedia e la didattica della scrittura*, in «Lingue e Culture dei Media» 4.1, 2020, <https://riviste.unimi.it/index.php/LCdM/article/view/14060>.

²¹ Un’ampia rassegna di esempi internazionali di lavoro su Wikipedia nelle scuole e nelle università è disponibile su Wikipedia stessa: Wikipedia, voce

‘Wikipedia:School and university projects’, https://en.wikipedia.org/wiki/Wikipedia:School_and_university_projects.

²² Amin Azzam, *Why Medical Schools Should Embrace Wikipedia: Final-Year Medical Student Contributions to Wikipedia Articles for Academic Credit at One School*, in «Academic Medicine» 92.2, 2017, pp. 194-200, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5265689/>.

²³ Julie Beck, *Doctors’# 1 Source for Healthcare Information: Wikipedia*, in «The Atlantic», 5 marzo 2014, <https://www.theatlantic.com/health/archive/2014/03/doctors-1-source-for-healthcare-information-wikipedia/284206/>; un’indagine più recente in Denise A. Smith, *Situating Wikipedia as a health information resource in various contexts: A scoping review*, in «PLoS ONE» 15.2, 2020, <https://doi.org/10.1371/journal.pone.0228786>.

²⁴ Su questo argomento si veda Wikipedia, voce ‘Gender bias on Wikipedia’, https://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia, e i testi di riferimento ivi richiamati, in particolare Claudia Wagner et al., *It’s a Man’s Wikipedia? Assessing Gender Inequality in an Online Encyclopedia*, in «Proceedings of the International AAAI Conference on Web and Social Media» 9 (1), 2015, pp. 454-463, <https://doi.org/10.1609/icwsm.v9i1.14628>, e Claudia Wagner et al., *Women through the glass ceiling: Gender asymmetries in Wikipedia*, in «EPJ Data Science» 5, 2016, pp. 1-24. È interessante notare che i bias riscontrabili relativamente a) alla scelta delle persone meritevoli di una voce su Wikipedia, b) al contenuto testuale delle voci e c) al loro contenuto visivo, sembrano avere caratteristiche in parte diverse: nel caso delle immagini, probabilmente anche la migliore qualità visiva media delle immagini di figure femminili rispetto a quelle di figure maschili è in qualche modo espressione di un bias rappresentazionale. Su questi aspetti cfr. Pablo Beytía et al., *Visual Gender Biases in Wikipedia: A Systematic Evaluation across the Ten Most Spoken Languages*, in «Proceedings of the International AAAI Conference on Web and Social Media» 16 (1), 2022, pp. 43-54, <https://doi.org/10.1609/icwsm.v16i1.19271>.

²⁵ Young-Ho Eom et al., *Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions*, in «PLoS ONE» 10.3, 2015, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114825>.

²⁶ Esempi interessanti di alcune fra le tante possibili situazioni di editing di Wikipedia a scopo commerciale sono discussi in Joe Pinsker, *The Covert World of People Trying to Edit Wikipedia – For Pay*, in «The Atlantic», 11 agosto 2015, <https://www.theatlantic.com/business/archive/2015/08/wikipedia-editors-for-pay/393926/>. Su esempi di revisionismo storico in alcune voci di Wikipedia, si veda Tommaso Baldo con il gruppo di lavoro ‘Nicoletta Bourbaki’, *Riflessioni sulla narrazione storica nelle voci di Wikipedia*, in «Diacronie. Studi di Storia Contemporanea» 29.1, 2017, http://www.studistorici.com/2017/03/29/baldo_numero_29/. Su questi temi si

vedano anche gli esempi discussi – da una prospettiva diversa – in Emanuele Mastrangelo ed Enrico Petrucci, *Wikipedia. L'enciclopedia libera e l'egemonia dell'informazione*, Bietti, Milano 2013, e in Andreas Mäckler (a cura di), *Schwarzbuch Wikipedia: Mobbing, Diffamierung und Falschinformation in der Online-Enzyklopädie und was jetzt dagegen getan werden muss*, Verlag zeitgeist, Hör-Grenzhausen 2020.

²⁷ Si veda Brian Martin, *Persistent Bias on Wikipedia: Methods and Responses*, in «Social Science Computer Review» 36.3, 2018, pp. 379-388, <https://journals.sagepub.com/doi/abs/10.1177/08944393177154342018>.

Un'interessante rassegna (con molti esempi) dei primi dibattiti tra i wikipediani sull'affidabilità e i possibili inconvenienti del processo editoriale di Wikipedia è contenuta in Andrew Lih, *The Wikipedia Revolution* cit., cap. 8. Per una discussione su come individuare i pregiudizi espliciti e impliciti negli articoli di Wikipedia, si veda Christoph Hube, *Bias in Wikipedia*, in *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*, International World Wide Web Conferences Steering Committee, Geneva 2017, pp. 717-721, <https://doi.org/10.1145/3041021.3053375>; cfr. anche Jackie Koerner, *Wikipedia Has a Bias Problem*, in Joseph Reagle e Jackie Koerner (a cura di), *Wikipedia @20* cit., pp. 311-321.

7.

Il sogno del web semantico

La soluzione ad almeno alcuni dei problemi di affidabilità e validazione che abbiamo discusso nel capitolo precedente potrebbe essere rappresentata dall'ulteriore evoluzione che l'enciclopedismo online ha conosciuto negli ultimi anni, e che ci porta dal concetto tradizionale di enciclopedia (ben riconoscibile anche in Wikipedia, che come già visto vi fa esplicito riferimento nel primo 'pilastro') verso l'idea di un'enciclopedia costruita come un database fortemente strutturato e semanticamente ricco, basato su rigorose ontologie formali. Un'enciclopedia di questo tipo non è pensata in primo luogo per l'uso diretto da parte di agenti umani, ma come uno strumento di ricerca e recupero di informazioni utilizzato prevalentemente da agenti software. A tali agenti è lasciato il compito di 'mediare' tra il rigore formale dei dati strutturati e le richieste informali presentate dagli utenti o, se necessario, di agire direttamente sulla base dei dati recuperati, o di elaborarli secondo le istruzioni, o ancora di trasferirli ad altri agenti software. In questo modo, gli assistenti vocali conversazionali come Google Assistant, Alexa, Siri, Cortana, Bixby possono rispondere alle domande e alle richieste presentate dagli utenti attraverso il linguaggio naturale e la sintesi vocale, interpretando la domanda per mezzo di un programma di parsing, identificandone gli elementi di base, utilizzandoli per interrogare il database ed estrarre le informazioni pertinenti, e formulando infine una risposta¹.

Così, se qualcuno a New York chiede a Google Assistant «chi è il presidente?», l'agente software stabilirà, innanzitutto, che siamo interessati a conoscere il nome del presidente in carica (l'interpretazione più semplice e diretta di «chi è?» rispetto ad altre possibili, come «che ruolo ha?»); interpreterà il termine «presidente» – in assenza di ulteriori

determinazioni – come riferito al presidente più importante e più citato, cioè alla principale carica politica del paese, e utilizzerà i dati di localizzazione per georeferenziare la richiesta agli Stati Uniti. L'agente software invierà quindi una versione formalmente accurata dell'interrogazione a un database ampiamente basato su DBpedia, una collezione di dati (dataset) costituita da una versione strutturata e formale di voci tratte da Wikipedia², disponibile con licenza aperta. Da DBpedia sarà quindi recuperato il nome che ci interessa. Se poniamo la domanda « quanti anni ha il presidente? », Google Assistant individuerà innanzitutto il riferimento al presidente degli Stati Uniti, effettuando la relativa ricerca come sopra descritto. Successivamente, sostituirà « presidente degli Stati Uniti » con il nome di chi al momento ricopre l'ufficio (Joe Biden, all'epoca in cui sto scrivendo questo libro) e cercherà nel database la voce relativa; da questa voce – che è strutturata, e quindi organizzata in campi con valori – estrarrà la data di nascita, e calcolerà l'età al momento dell'interrogazione.

La costruzione di basi di dati altamente strutturati – ma in casi come quello rappresentato da DBpedia possiamo parlare, più che di un semplice database, di una 'knowledge base', cioè di una raccolta di dati che corrispondono a conoscenze organizzate – si basa su ontologie, cioè su sistemi di classificazione rigorosi. Per capire un po' meglio di cosa si tratti, pur nei limiti di questo lavoro, che non rappresenta una introduzione specifica al tema delle ontologie e della metadattazione³, torniamo al nostro esempio e al presidente degli Stati Uniti. Se è vero che, nel momento in cui scrivo, il presidente degli Stati Uniti è Joe Biden, è anche vero però che i due concetti sono distinti: Joe Biden è una persona, 'presidente degli Stati Uniti' è una carica pubblica. Joe Biden era Joe Biden anche prima di essere presidente (e ha avuto altre cariche pubbliche), e ci sono stati numerosi presidenti degli Stati Uniti prima di lui (così come, presumibilmente, ce ne saranno dopo di lui). Così, comprensibilmente, Wikipedia prevede due voci distinte: una sulla persona Joe Biden, e una sulla carica 'President of the United States'.

Queste due voci hanno caratteristiche (proprietà) diverse: la persona ha una data e un luogo di nascita, avrà prima o poi una data e un luogo di morte, ha studiato in una determinata scuola e poi in una determinata

università, ha avuto una serie di altri incarichi... Se vogliamo costruire una 'scheda' composta dai campi che riassumono i dati salienti di una persona, dovremmo pensare a campi di questo tipo. D'altro canto, una 'scheda' relativa alla carica 'presidente degli Stati Uniti' avrà caratteristiche diverse: potrebbe riportare la data di istituzione della carica, il metodo di elezione, la residenza ufficiale di chi la detiene, e così via. Fra i vari campi, ci potrebbe essere quello relativo all'attuale detentore della carica, che al momento in cui scrivo ha valore 'Joe Biden', e cambierà valore alla prossima presidenza (i nomi che hanno costituito nel tempo i valori di quel campo hanno avuto finora la pervicace abitudine a designare individui di sesso maschile, ma speriamo non sia sempre così...).

Quali campi, *esattamente*, prevedere in queste schede? E quali campi prevedere, ad esempio, nelle 'schede' di un elemento chimico, di una specie animale, di un modello di automobile, di un vino, di un partito politico? E quante diverse categorie ci sono? Teniamo presente, fra l'altro, che potremmo voler differenziare le schede di individui diversi in base ad alcune caratteristiche: ad esempio, alla loro professione (nella scheda di un calciatore potremmo voler registrare il ruolo, o le squadre in cui ha giocato; in quella di una scrittrice i libri che ha scritto e i premi avuti; in quella di una astronauta le missioni spaziali alle quali ha partecipato...).

Il compito di costruire un sistema di classificazione (una *ontologia*) capace di assegnare a ogni voce di enciclopedia una specifica categoria, e a ogni categoria un modello di 'scheda' descrittiva divisa in campi standardizzati, equivale un po' al compito, apparentemente impossibile, di schedare l'universo.

Eppure, il sogno di cercare e trovare sistemi classificatori che 'mettano ordine' nella realtà fa parte della storia del pensiero umano. Ci ha provato Aristotele con le sue categorie (e l'opera aristotelica è considerata, proprio per il suo carattere organizzato e il tentativo di coprire sistematicamente almeno i principali rami dello scibile, come una sorta di proto-enciclopedia); ci ha provato – in una forma diversa – Leibniz, con la sua idea di caratteristica universale; ci ha provato chi ha costruito alberi delle scienze, chi ha costruito enciclopedie sistematiche, ci hanno provato i bibliotecari con complessi sistemi come la classificazione decimale Dewey o la classificazione decimale universale, ci hanno provato – in ambiti

specifici – opere come il *Systema naturae* di Linneo. DBpedia ha insomma una lunga eredità alle spalle!

Per rendere più rigoroso questo lavoro – che, come capirete, è per sua natura spesso discutibile e soggetto a revisioni – sono stati sviluppati quelli che possiamo considerare come veri e propri *linguaggi descrittivi*. Uno dei principali è il Web Ontology Language, conosciuto con la sigla OWL. Nelle parole della guida introduttiva alla sua seconda versione⁴, OWL è un linguaggio «costruito per rappresentare conoscenze ricche e complesse relative a entità, gruppi di entità e relazioni fra entità»⁵, dove ‘entità’ è da intendersi in senso molto largo: persone, elementi chimici, marche di automobili, tipi di calzature, stili pittorici, strumenti musicali...

OWL rappresenta queste conoscenze in forma di *ontologie*, ed è dunque un linguaggio per la descrizione di ontologie; per ontologia si intende qui «un insieme di affermazioni descrittive specifiche su un qualche aspetto del mondo (al quale di solito ci si riferisce come al *dominio di interesse* o alla *tematica* dell’ontologia)»⁶. Per farci un’idea di massima (anche se non del tutto precisa) di come sia costruita un’ontologia possiamo dire che le singole entità corrispondono a *individui*, i gruppi di entità corrispondono a *classi* e le relazioni fra entità corrispondono a *proprietà*. Le ontologie costruite attraverso OWL sono ontologie formali: sono pensate per essere utilizzate da agenti software più che da utenti umani, e la sintassi di OWL prevede che corrispondano a sequenze di *annotazioni*, *assiomi* e *fatti*. Assiomi e fatti esprimeranno (in forma standardizzata) le informazioni di base sulle classi, sulle proprietà e sugli individui di cui si occupa l’ontologia, mentre le annotazioni possono essere usate per aggiungere descrizioni e metadati. Così, ad esempio, potremmo esprimere attraverso assiomi relazioni del tipo «un golden retriever è un tipo di cane», mentre «Olivia è un golden retriever» è un fatto, e la fonte da cui è stata ricavata questa informazione (ad esempio il pedigree di Olivia) potrebbe essere espressa attraverso una annotazione. Questo esempio ci dice anche che le classi possono essere fra loro in relazione gerarchica (volendo spostarci dai cani agli esseri umani, la classe delle persone che giocano a calcio è una sottoclasse delle persone).

In OWL, tutte queste affermazioni vengono espresse attraverso una sintassi rigorosa (ed è possibile anzi scegliere fra sintassi diverse in base alle

proprie necessità e ai software utilizzati), non sempre facile da interpretare per un essere umano, ma utilizzabile senza ambiguità da parte di un programma. Dal punto di vista formale, OWL si basa sulla cosiddetta 'logica descrittiva', che è un sottoinsieme della logica del primo ordine. La logica descrittiva è meno espressiva della logica del primo ordine completa, ma permette di norma di costruire algoritmi efficienti per decidere se una data affermazione è vera o falsa, e questo permette di usarla bene, ad esempio, in applicazioni di intelligenza artificiale.

Volendo andare più in profondità, potremmo aggiungere che OWL è solo uno dei piani che costituiscono l'edificio di rappresentazioni formali della conoscenza che stiamo cercando di costruire: un piano di livello abbastanza 'alto', dato che ha a che fare con le ontologie e dunque con i sistemi di organizzazione concettuale del sapere. Anche per questo può lavorare con sintassi diverse. Livelli più bassi dello stesso edificio si occuperanno, ad esempio, di formalizzare il modo per descrivere proprietà e relazioni relative a classi e individui. Uno di questi livelli è RDF (Resource Description Framework), uno standard per la descrizione e lo scambio di dati (o meglio, di affermazioni su *risorse*) espressi in forma di *triple* soggetto-predicato-oggetto. Per RDF, una risorsa è sostanzialmente qualunque tipo di entità che possa essere identificata e descritta, anche se RDF è nato innanzitutto per lavorare sulle informazioni disponibili in rete (una pagina web, una voce di Wikipedia, un'immagine, un file...). Per questo motivo le risorse RDF sono rappresentate attraverso URI (Uniform Resource Identifier), che permettono di identificarle univocamente in rete e assomigliano un po' agli indirizzi delle pagine web (la cosa non è casuale: l'indirizzo di una pagina web è una forma di URI). Le triple soggetto-predicato-oggetto sono quindi un modo per collegare fra loro le URI in modo da descrivere proprietà o relazioni delle risorse di cui vogliamo parlare.

Può sembrare tutto abbastanza astratto e complicato, ma un esempio potrà forse aiutare: torniamo a Joe Biden e al suo ruolo di presidente degli Stati Uniti. Immaginiamo di voler identificare queste due entità attraverso le relative pagine su Wikipedia inglese. Allora, per esprimere il fatto che Biden è attualmente il presidente degli Stati Uniti, dovremmo costruire una tripla che ha come soggetto «Joe Biden», come predicato «è attualmente», e come oggetto «presidente degli Stati Uniti». Nella nostra

tripla la URI che identifica il soggetto sarà https://en.wikipedia.org/wiki/Joe_Biden, mentre la URI che identifica l'oggetto sarà https://en.wikipedia.org/wiki/President_of_the_United_States. Però su Wikipedia non abbiamo una risorsa che identifichi la proprietà «è attualmente»: dovremmo fare riferimento a qualche altra risorsa web che preveda e descriva questa proprietà come elemento di un'ontologia.

Ed ecco che ci viene in aiuto DBpedia, che – appunto – ha alle spalle una ontologia assai ricca. L'ontologia di DBpedia, di fatto, non solo possiede una proprietà adatta, ma la possiede in una forma più specifica e calzante: la proprietà «office», che corrisponde all'incarico ricoperto da una persona all'interno di un'istituzione. Ecco che abbiamo pronta la nostra tripla, tutta costruita attraverso URI di DBpedia:

- soggetto: http://dbpedia.org/page/Joe_Biden
- predicato: <http://dbpedia.org/ontology/office>
- oggetto: http://dbpedia.org/page/President_of_the_United_States

Triple di questi tipo possono essere anche presentate in forma di 'knowledge graph': una rappresentazione dei dati che visualizza le entità come nodi in un grafo, e le relazioni tra queste entità come archi che collegano i nodi. Ogni nodo e ogni arco può avere un'etichetta che indica il tipo di entità o di relazione rappresentata. Ciò permette di esplorare i dati anche in forma visuale, 'navigando' attraverso grafi. Per farvene un'idea, potete provare a seguire il link del QR-Code qui di seguito, che vi rimanderà al video di presentazione di uno strumento di navigazione denominato LodLive⁷. Se volete provare a usarlo direttamente, potete andare all'indirizzo della pagina DBpedia su Joe Biden (riportato sopra) o a qualunque altra pagina DBpedia, e scegliere dalla linguetta 'Browse using' (in alto sulla sinistra) l'opzione LodLive Browser.



QR-Code 10

Presentazione di LodLive

<http://bit.ly/44wFdsF>

Lavorare sulle ontologie formali è un passo indispensabile nella transizione alla quale ho già accennato in apertura: quella dal web orientato principalmente alla consultazione da parte di agenti umani, al web ‘semantico’, che può essere utilizzato anche (e forse sarà utilizzato principalmente) da agenti software, e in cui le relazioni tra elementi – e quindi anche tra le voci di un’enciclopedia – sono esplicite e a loro volta formalizzate. Questo è il cuore del progetto del cosiddetto ‘web semantico’, proposto da Tim Berners-Lee alla fine degli anni ’90. Oggi tale progetto ha portato al lavoro sui Linked Open Data (LOD), che ne rappresenta una versione in parte semplificata ma più realisticamente gestibile. Il nome scelto sottolinea da un lato l’elemento fortemente relazionale delle ontologie, dall’altro il requisito di apertura dei dati, che ne permette il riuso e quindi – in un’altra manifestazione della tendenza alla costruzione collaborativa di informazione complessa – l’integrazione di ontologie e basi di conoscenze parziali e settoriali in costruzioni di più alto livello.

La breve (e a volte approssimativa) sintesi che ho provato a proporvi non intende in alcun modo rappresentare una ‘introduzione’ a semantic web, LOD, OWL, RDF... il compito richiederebbe molto più spazio, e peraltro esistono ottimi testi e corsi che lo fanno assai meglio di quanto non saprei e potrei fare io⁸. Lo scopo era piuttosto quello di fornire un’idea della direzione che ha preso negli ultimi anni l’enciclopedismo digitale, e dei meccanismi di funzionamento di una enciclopedia trasformata, sostanzialmente, in una base di dati strutturati, organizzati e facilmente utilizzabili da parte di agenti software.

Vale la pena notare, nel chiudere questa parte, che enciclopedie di questo tipo superano la tradizionale dicotomia fra enciclopedia sistematica ed enciclopedia alfabetica. Le enciclopedie digitali, con i loro meccanismi di ricerca rapida, avevano già superato la necessità di utilizzare l'ordine alfabetico come metodologia privilegiata per la ricerca di una voce, permettendo inoltre di rendere ricercabile l'intero testo delle voci e non solo i loro titoli. Le singole voci restavano però oggetti relativamente separati e granulari: anche se i link ipertestuali consentivano di costruire percorsi e rimandi fra di esse, non consentivano di esprimere adeguatamente le diverse tipologie di relazioni – a partire da quelle gerarchiche – esistenti all'interno dell'edificio delle nostre conoscenze. Una enciclopedia costruita utilizzando i linguaggi e gli strumenti del web semantico, come DBpedia, supera questo limite, e rappresenta dunque un edificio architettonicamente assai più complesso ed espressivo.

Semantic web e linked open data hanno dunque dato all'architetto strumenti potentissimi per combattere granularità e frammentazione. Ma alcuni problemi restano aperti. Innanzitutto, quello della distinzione, alla quale abbiamo già accennato, fra conoscenze e informazioni: non tutte le informazioni sono conoscenze. Così, dal punto di vista della teoria dell'informazione, ogni dato ha un contenuto informativo, ma questo non gli attribuisce automaticamente anche un contenuto conoscitivo. Un elenco di dati numerici grezzi contiene informazione, ma se non sappiamo a cosa tali dati si riferiscano, non possiamo parlare di conoscenze. Una affermazione come «il mostro del lago di Loch Ness potrebbe divorarti» ha un contenuto informativo, ma se – come probabile – quel mostro non esiste, quella frase non ha valore conoscitivo. Il web contiene moltissime informazioni che non sono conoscenze, e spesso, come ho già accennato, la distinzione non è così chiara. Nell'ambito del web semantico si parla abitualmente di 'basi di conoscenze', ma quelle che cerchiamo di gestire sono conoscenze o informazioni? Nel caso di Wikipedia, che, come si è detto, *vuole* essere un'enciclopedia, parlare di basi di conoscenze sembra sensato. Ma davanti ad altri tipi di corpora, la distinzione può sfuggire di mano.

Per altro verso, il movimento di Wikipedia in questa direzione non è facile: trasformare le voci scritte dai volontari in risorse adeguatamente descritte attraverso gli strumenti formali di cui abbiamo parlato in questo

capitolo richiede moltissimo tempo e competenze specifiche, costruire ontologie e basi di dati rigorose (e applicarle) è impresa lunga e complicata, e allargare questo lavoro – se pure lo si volesse fare – all’intero contenuto del web, o anche solo a una sua porzione significativa, è oggi semplicemente impensabile. Per altro verso, si è affacciato sulla scena un personaggio che lavora con metodi assai diversi rispetto a quelli dell’architetto: l’oracolo. Fuor di metafora, gli sviluppi delle reti neurali e dell’intelligenza artificiale sembrano suggerire nuove strade di organizzazione e riuso delle conoscenze: per certi versi meno formalmente rigorose, ma per altri versi incredibilmente produttive.

Si tratta di una nuova direzione, e magari di una nuova fase, nella nostra capacità di produrre, gestire, selezionare, validare, riusare informazione complessa? Architetto e oracolo sono avversari, o possono collaborare? E in particolare, all’oracolo – cioè, a sistemi di intelligenza artificiale – potrebbe essere affidato il compito di generare la sterminata messe di descrizioni rigorose di cui avremmo bisogno per realizzare il sogno del web semantico? Per provare a capirlo, dobbiamo esplorare un campo del tutto nuovo.

¹ Per una panoramica approfondita sul funzionamento degli assistenti vocali conversazionali, cfr. Steve Young, *Hey Cyba: The Inner Workings of a Virtual Personal Assistant*, Cambridge University Press, Cambridge 2021. Una discussione – oggi in parte datata – sul rapporto tra gli assistenti vocali conversazionali e l’IA ‘classica’ (dal punto di vista del test di Turing) si trova in Gino Roncaglia, *Computer che copiano: test di Turing, web corpora e filtraggio collaborativo*, in Tito Orlandi (a cura di), *Per il centenario di Alan Turing, fondatore dell’informatica. Convegno (Roma, 22 novembre 2012)*, Accademia Nazionale dei Lincei («Contributi del Centro Linceo Interdisciplinare Beniamino Segre» 129), Roma 2014, pp. 189-201, mentre un’interessante proposta sull’uso dei dataset di Wikipedia per generare risposte ‘naturali’ attraverso l’uso di sistemi di intelligenza artificiale si trova in Emily Dinan et al., *Wizard of Wikipedia: Knowledge-Powered Conversational Agents*, Version 2, 2019, paper submitted to arXiv, <https://arxiv.org/abs/1811.01241>.

² Il sito web è <https://wiki.dbpedia.org/>; per una discussione sul progetto si veda Jens Lehmann et al., *DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia*, in «Semantic Web» 6.2, 2015, pp. 167-195, <https://content.iospress.com/articles/semantic-web/sw134>.

³ Non discuto qui il tema, pur assai rilevante, del rapporto fra metadati e ontologie; sul tema dei metadati segnalo però due risorse italiane recenti, sicuramente pertinenti anche rispetto ai temi di questo capitolo: Mauro Guerrini (a cura di), *Metadattazione*, Editrice Bibliografica, Milano 2022; Mauro Guerrini, *Dalla catalogazione alla metadattazione. Tracce di un percorso*, seconda edizione a cura di Denise Biagiotti e Laura Manzoni, Associazione Italiana Biblioteche, Roma 2022.

⁴ *OWL 2 Web Ontology Language Primer (Second Edition)*, W3C Recommendation, 11 dicembre 2012, in rete alla pagina <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>.

⁵ *Ibid.*

⁶ *Ibid.*

⁷ LodLive è per certi versi uno strumento ormai datato, ma resta assai interessante. Uno strumento alternativo, con funzioni in parte diverse (è in primo luogo un editor di ontologie), è Protégé: <https://protege.stanford.edu/>.

⁸ Per una descrizione accessibile della natura e dello scopo del web semantico, è sempre utile il famoso articolo di Tim Berners-Lee, James Hendler e Ora Lassila, *The Semantic Web*, in «Scientific American» 284.5, 1° maggio 2001, pp. 34-43, in rete alla pagina <https://www.scientificamerican.com/article/the-semantic-web/> (l'articolo è sotto paywall, ma può essere reperito facilmente in molti altri posti in rete); l'idea era già stata incorporata in altri lavori di Tim Berners-Lee della fine degli anni '90, ad esempio Tim Berners-Lee e Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, Harper, San Francisco 1999. Questi testi sono però ormai abbastanza datati; per chi volesse esplorare queste tematiche, alcuni riferimenti bibliografici più aggiornati sono: Grigoris Antoniou, Paul Groth, Frank van Harmelen e Rinke Hoekstra, *A Semantic Web Primer*, terza edizione, MIT Press, Cambridge (MA) 2012; poi, a un livello di approfondimento (e di complessità) maggiore, Dean Allemang, James Hendler e Fabien Gandon, *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*, terza edizione, ACM Books, New York 2020; per una prima introduzione accessibile all'idea di linked data, si veda Christian Bizer, Tom Heath e Tim Berners-Lee, *Linked Data: The Story so Far*, in Amit Sheth (a cura di), *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, IGI Global, Hershey (PA) 2011, <https://www.igi-global.com/chapter/linked-data-story-far/55046>, pp. 205-227; per una presentazione più approfondita, si veda Tom Heath e Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, San Rafael (CA), 2011 e, per una presentazione più aggiornata, il recentissimo Terhi Nurmikko-Fuller, *Linked Open Data for Digital Humanities*, Routledge, London-New York 2023. In italiano, e con particolare riferimento agli ambiti di più specifico interesse dal punto di vista di questa ricerca, segnalo Mauro Guerrini e Tiziana Possemato, *Linked data per biblioteche, archivi e musei*, Editrice Bibliografica, Milano 2015, e

Francesca Tomasi, *Organizzare la conoscenza: Digital Humanities e Web semantico*, Editrice Bibliografica, Milano 2022. Un elenco di ulteriori testi di riferimento sul web semantico è disponibile su Wikipedia alla pagina <https://www.w3.org/2001/sw/wiki/Books>.

Parte II

8.

Architetti, oracoli, pappagalli

Si è detto nell'Introduzione che al modello di organizzazione architettonica delle conoscenze si può affiancare (o, a seconda dei casi, contrapporre) un modello diverso: un modello che sostituisce al lavoro controllato e sistematico di costruzione progressiva e gerarchica di complessità, a partire da blocchi costitutivi ben noti ed esaustivamente descritti, una sorta di sviluppo organico, non perfettamente controllato o controllabile, in cui i livelli inferiori sono in parte ignoti o comunque talmente ricchi da impedirne una descrizione puntuale, mentre ai livelli superiori emergono proprietà e caratteristiche a loro volta difficili da scomporre o ridurre meccanicamente alle loro componenti.

La parte prima di questo libro aveva l'obiettivo di approfondire le caratteristiche e lo sviluppo del modello architettonico nel contesto del nuovo ecosistema digitale, con particolare riferimento all'ambito della sistematizzazione enciclopedica del sapere, che ne costituisce un po' l'esempio paradigmatico. Può essere utile a questo punto osservare che l'intelligenza artificiale 'classica', la cosiddetta Good Old-Fashioned Artificial Intelligence (GOF AI), ha tradizionalmente cercato di lavorare sulla conoscenza in modi molto strutturati, utilizzando logiche formali, rappresentazioni simboliche e modelli espliciti del mondo. In altri termini, per riprendere la metafora che dà il titolo al libro, le prime ricerche in intelligenza artificiale consideravano come parte essenziale del loro lavoro la costruzione di basi di conoscenze esplicite, e questa costruzione era affidata all'architetto e non certo all'oracolo: l'architetto considerava essenziale procedere in modo sistematico, documentato e verificabile su conoscenze ben definite e il più possibile organizzate (in genere, organizzate gerarchicamente).

Le intelligenze artificiali generative, da diversi mesi al centro di una notevole attenzione mediatica, sembrano adottare invece il secondo modello. Sistemi di questo tipo non sono programmati in modo esplicito partendo da basi di conoscenze organizzate architettonicamente: imparano invece a generare output appropriato sulla base di enormi quantità di dati di addestramento; dati certo in qualche misura ‘preparati’ (è un tema di cui si parlerà in seguito) ma non strutturati in maniera così forte. Dati, inoltre, che sono certo informazioni ma che spesso – per riprendere una distinzione già discussa nella parte precedente – non sono conoscenze. Si tratta insomma di un processo molto meno controllato rispetto alla GOFAI: i modelli di apprendimento di una rete neurale profonda sviluppano internamente rappresentazioni complesse dei dati, che possono essere difficili da comprendere o da esprimere in termini simbolici o logici, e sono quindi spesso considerate ‘scatole nere’. È da questo punto di vista che il funzionamento di tali sistemi può essere visto come in qualche misura ‘oracolare’: il processo che porta alla produzione dell’output non è quasi mai esplicitabile o descrivibile secondo i paradigmi architettonici tradizionali.

Vale la pena porsi subito, al riguardo, una domanda che dovrà poi essere ripresa in seguito: volendo proporre una metafora per il funzionamento dei sistemi generativi, è più sensato paragonarli a ‘pappagalli stocastici’ (come viene fatto abbastanza spesso) o a oracoli statistico-probabilistici (come faccio in questo libro)?

L’espressione ‘pappagallo stocastico’ è stata proposta dalla linguista americana Emily Bender in un articolo, assai citato, del 2021¹, che in un certo senso rappresenta un’appassionata e in parte accorata difesa dell’architetto dall’improvvisa e incontrollata invadenza dell’oracolo. In tale lavoro Emily Bender e le sue collaboratrici sottolineano innanzitutto gli enormi costi finanziari e ambientali della creazione e dell’addestramento di grandi modelli linguistici, e non c’è dubbio che tali costi siano notevoli: l’addestramento di una rete neurale così ampia richiede mesi di lavoro parallelo di centinaia di processori ad alte prestazioni, che sono economicamente cari e assai esigenti in termini energetici. Questi costi sono strettamente correlati alla dimensione del modello, e i Large Language Models, come dice il nome, sono molto

grandi. Inoltre, i corpora utilizzati per l'addestramento di LLM sono spesso disomogenei e viziati da pregiudizi (un tema su cui tornerò più avanti), inclusi pregiudizi di genere e sottorappresentazione di diversità e minoranze. Peraltro, proprio le loro dimensioni rendono assai difficile – e, di nuovo, assai costosa anche in termini di risorse umane, spesso costituite da ‘lavoratori della conoscenza’ reclutati nel sud del mondo e sottopagati – una loro adeguata revisione e annotazione. Manca, in sostanza, il lavoro dell'architetto:

[...] i modelli linguistici addestrati su grandi dataset non curati e statici estratti dal web includono punti di vista egemonici pericolosi per le popolazioni marginalizzate. Per questo sottolineiamo la necessità di investire risorse significative nella cura e documentazione dei dati di addestramento dei modelli linguistici. [...] Quando ci affidiamo a insiemi di dati così ampi, rischiamo di incorrere in situazioni di *debito di documentazione*, e cioè di metterci in situazioni in cui i dataset sono sia non documentati sia troppo ampi per poter essere documentati a posteriori².

Molto meglio, allora, lavorare su insiemi più limitati e controllati di dati selezionati, descritti e annotati. In questa direzione va anche quello che è in realtà un secondo argomento, legato appunto all'idea di ‘pappagallo stocastico’: mentre i dati curati possono includere informazioni semantiche fornite, validate e verificate da noi, l'apparente correttezza sintattica e semantica delle risposte fornite da sistemi addestrati su grandi basi di dati non curati nasconde in realtà un inganno: quel che possiamo ottenere in questo modo è solo

[...] un sistema per incollare insieme a casaccio sequenze di forme linguistiche che esso ha incontrato nei suoi tanti dati di addestramento, sulla base di informazioni probabilistiche su come esse vengono combinate, ma senza alcun riferimento ai significati: un pappagallo stocastico (*stochastic parrot*)³.

È difficile non riconoscere che le preoccupazioni – sia dal punto di vista dei costi finanziari e dell'impatto ambientale, sia da quello dei pregiudizi quasi inevitabilmente presenti in corpora di addestramento così ampi – sono fondate e vanno tenute presenti. C'è però un aspetto, fondamentale, che Emily Bender e le sue collaboratrici sembrano non tener presente: i risultati ottenuti lavorando su LLM sono talmente notevoli dal punto di vista delle competenze linguistiche e semantiche del sistema, da suggerire in maniera assai chiara che questa metodologia abbia permesso un vero e

proprio, impressionante ‘salto’ qualitativo. Non siamo insomma davanti a un minimo progresso incrementale, ottenuto per di più a caro prezzo: il costo c’è, ma il risultato ha caratteristiche talmente sorprendenti – di fatto non previste e apparentemente non prevedibili in partenza – da rendere quasi impossibile (e per certi versi improponibile anche in termini di onestà scientifica del lavoro di ricerca) l’idea di tornare indietro.

Questo, si badi, non implica che i corpora di addestramento non debbano e non possano essere più curati, che i fattori di impatto ambientale non vadano valutati e compensati, che i pregiudizi non debbano essere tenuti presenti e che non si debba lavorare per ridurli o eliminarli (sapendo che l’idea di un dataset senza bias è solo un ideale regolativo, non troppo dissimile da quello rappresentato dal ‘neutral point of view’ nel caso di Wikipedia). Ma l’oracolo, inaspettatamente, si è dimostrato troppo bravo per poter essere semplicemente licenziato dall’architetto.

Quanto all’idea di pappagallo stocastico, la mia impressione è che – pur se ormai spesso utilizzata anche dagli addetti ai lavori – essa nasconda una sostanziale, grave incomprendimento del funzionamento dei LLM. Che non sono pappagalli (non si limitano affatto a ripetere meccanicamente frammenti dei testi su cui sono stati addestrati, o almeno non lo fanno se per ‘frammenti’ intendiamo porzioni ragionevolmente significative di tali testi) e non sono neanche stocastici, dato che il loro funzionamento è guidato da una procedura di cui parlerò più in dettaglio in seguito – l’embedding – che coglie ed esprime numericamente, pur se su basi probabilistiche e in maniera per noi in parte oscura, elementi sintatticamente e semanticamente rilevanti dei nostri usi linguistici. ChatGPT e i sistemi analoghi, insomma, non funzionano affatto incollando «a casaccio» (il termine inglese usato nel passo sopra citato è «haphazardly») sequenze di forme linguistiche: lo fanno a ragion veduta, sulla base di modelli probabilistici assai complessi. Come un oracolo, questi modelli producono contenuti sulla base di una ‘visione’, anche se questa visione è almeno in parte privata e non conoscibile dall’esterno.

Confido che queste considerazioni possano risultare più chiare una volta completata la lettura di questa parte del libro, ma era utile anticiparle anche per meglio giustificare la metafora che si è qui scelto di adottare.

Si potrebbero discutere a lungo anche le possibili radici storiche di

questo modello ‘organico’ o oracolare: ad esempio, potremmo considerare la trasmissione orale di conoscenze – prima dell’invenzione della scrittura – come un processo di questo tipo, in cui il corpus conoscitivo viene costruito e trasmesso, cresce e si trasforma nel tempo, senza essere mai perfettamente esplicitato e organizzato⁴. Da questo punto di vista, l’introduzione della scrittura può essere considerata come uno sviluppo di tipo ‘architettonico’: non a caso il dio egizio Toth, presentato – anche nel *Fedro* di Platone – come il mitologico inventore della scrittura (oltre che scriba degli dèi e protettore degli scribi), era anche la divinità di riferimento per la misurazione e l’organizzazione delle conoscenze e per la geometria. E una connotazione in gran parte simile ha anche Seshat, la divinità egizia più direttamente associata all’architettura, paredra di Toth e spesso indicata come sua moglie o sua figlia: non stupirà il fatto che Toth e Seshat fossero le due divinità esplicitamente collegate agli archivi e alle biblioteche⁵.

In questa sede non esplorerò comunque questa analogia, pur suggestiva, e in generale non discuterò storia (e preistoria) del modello organico di organizzazione e sviluppo delle conoscenze, per concentrarmi invece sulle intelligenze artificiali generative. Il loro sviluppo promette cambiamenti anche radicali in molti ambiti professionali, incluso il mondo della mediazione informativa e quello della produzione di conoscenze complesse. Si tratta di una previsione giustificata? Il funzionamento delle intelligenze artificiali generative può effettivamente essere considerato – e in che senso – come ‘oracolare’? E in questo caso, che effetti potrà avere l’interazione fra sistemi architettonici di conoscenze e generazione statistico-probabilistica di contenuti?

Per provare a rispondere a queste domande, occorre per prima cosa capire di cosa esattamente stiamo parlando: cosa sono, e come funzionano, le intelligenze artificiali generative? In questa parte⁶ cercherò quindi di presentare – in forma necessariamente sintetica – il contesto all’interno del quale si è sviluppato il lavoro su questi sistemi, i meccanismi di funzionamento e le caratteristiche di alcuni di essi (in particolare di quelli basati sulla generazione di testi attraverso transformer, come GPT e ChatGPT), i principali problemi riscontrati e una prima, assai parziale riflessione sull’impatto che essi potranno avere in futuro.

¹ Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21*, Association for Computing Machinery, New York 2021, pp. 610-623, <https://dl.acm.org/doi/10.1145/3442188.3445922>.

² Ivi, p. 615.

³ Ivi, p. 617.

⁴ Punto di riferimento obbligato sul tema è ovviamente Walter J. Ong, *Orality and Literacy: The Technologizing of the Word*, Methuen, New York 1982; 30th Anniversary Edition, Routledge, London-New York 2012. Ho scritto la prefazione alla traduzione italiana di quest'ultima edizione: Walter J. Ong, *Oralità e scrittura. Le tecnologie della parola*, Il Mulino, Bologna 2014.

⁵ Cfr. Kerry Webb, "The House of Books": *Libraries and Archives in Ancient Egypt*, in «Libri» 63.1, 2013, <https://doi.org/10.1515/libri-2013-0002>.

⁶ Il testo dei capitoli 9-12 riprende, in forma in parte accresciuta e modificata, il contenuto del mio articolo *Intelligenze artificiali generative e mediazione informativa: una introduzione*, in «Biblioteche Oggi Trends» 9, 1, giugno 2023, pp. 13-26, <https://www.bibliotecheoggitrends.it/it/articolo/2839/intelligenze-artificiali-generative-e-mediazione>. Il fascicolo di «Biblioteche Oggi Trends» è completamente dedicato al tema dell'intelligenza artificiale nel mondo bibliotecario e della mediazione informativa e comprende numerosi articoli rilevanti dal punto di vista di questo libro, alcuni dei quali avrò occasione di citare in seguito. Segnalo anche che la rivista dispone – all'indirizzo <https://www.bibliotecheoggitrends.it/> – di una piattaforma di fruizione particolarmente avanzata, che offre utili integrazioni online (da video-abstract a materiali di approfondimento) relative agli articoli pubblicati, incluso il mio; in particolare, fra i materiali relativi al mio articolo segnalo i commenti di Derrick de Kerckhove, che propongono spunti ulteriori di riflessione credo particolarmente interessanti.

9.

Il contesto: IA e reti neurali

La riflessione sulla possibilità di costruire macchine ‘intelligenti’ (in un qualche senso del termine) è molto antica: potremmo ad esempio partire dagli automi e dagli esseri artificiali intelligenti presenti nelle narrazioni omeriche, o soffermarci sulla diffusa – e ovviamente falsa – leggenda tardo-medievale e rinascimentale dell’androide o della testa di androide intelligente la cui costruzione era attribuita ad Alberto Magno e la cui distruzione (volontaria o involontaria) era da alcuni attribuita al più noto fra gli allievi di Alberto, Tommaso d’Aquino¹. Ma il lavoro sull’intelligenza artificiale collegato agli sviluppi nel campo dell’informatica e alla rivoluzione digitale è ovviamente assai più recente²: inizia negli anni ’50 del secolo scorso ed è legato soprattutto a due nomi e due occasioni che hanno contribuito in maniera determinante a delinearne l’impostazione iniziale: quello di Alan Turing, che nell’articolo del 1950 *Computing Machinery and Intelligence*³ ha posto le basi teoriche della riflessione sul rapporto fra intelligenza artificiale e intelligenza umana, e quello di John McCarthy, che ha organizzato il fondamentale seminario svoltosi nell’estate 1956 al Dartmouth College. È nel documento preparatorio di tale incontro⁴ che compare l’espressione «artificial intelligence», ed è in questo contesto che nasce l’indirizzo della cosiddetta ‘intelligenza artificiale forte’: l’idea che gli sviluppi dell’informatica possano permettere di creare macchine dotate di un’intelligenza simile alla nostra.



QR-Code 11

Il ‘Proposal’ per l’incontro di Dartmouth

<http://bit.ly/3NI7VQz>

HAL 9000, il computer al centro del film di Stanley Kubrick *2001: A Space Odyssey*, è la rappresentazione forse più nota di questa ambizione. Fra i consulenti di Kubrick per quel film, uscito nel 1968, vi era Marvin Minsky, che aveva affiancato McCarthy nella creazione del laboratorio di intelligenza artificiale del MIT, e HAL rappresenta per molti versi il prototipo di ‘computer intelligente’ che negli anni ’60 del secolo scorso sembrava quasi a portata di mano: basti pensare che meno di 35 anni separavano la data di realizzazione del film da quella della sua ambientazione, e che ancor più breve era l’orizzonte temporale previsto dal premio Nobel per l’economia Herbert Simon dopo la sua partecipazione al seminario di Dartmouth: appena vent’anni⁵.



QR-Code 12

Intervista a HAL 9000, da *2001: Odissea nello spazio*

<http://bit.ly/3O4dgCY>

Quello che, col senno di poi, potrebbe sembrare eccessivo ottimismo sulla rapidità dello sviluppo tecnologico era in realtà il risultato di una doppia assunzione sulla natura dell’intelligenza umana: da un lato, l’idea che la nostra intelligenza sia in primo luogo linguistica (manifestiamo la nostra intelligenza soprattutto – anche se non unicamente – attraverso l’uso del linguaggio); dall’altro, l’idea che il linguaggio possa essere considerato un sistema governato da regole, e che queste regole siano in linea di principio formulabili in maniera tanto precisa e rigorosa da poter essere programmate e utilizzate da un sistema informatico. Due assunzioni con una lunga storia alle spalle (basti pensare all’idea del linguaggio come calcolo in Hobbes e alle intuizioni di Leibniz sulla convergenza del calcolo logico e del calcolo binario in direzione di quella che potremmo chiamare ‘computabilità del mondo’) ma frutto anche del clima culturale dell’epoca: dalla ‘linguistic turn’, la ‘svolta linguistica’ propria della prima filosofia analitica, che poneva l’uso del linguaggio al centro della riflessione filosofica, agli sviluppi della logica formale, a sua volta fra i tratti distintivi del panorama culturale novecentesco. Si aggiunga, sul secondo fronte, il lavoro che proprio in quel periodo portavano avanti linguisti come Noam Chomsky, che

ipotizzavano l'esistenza di una struttura grammaticale profonda comune alle diverse lingue e analizzabile in termini rigorosamente formali.

Se la nostra intelligenza è soprattutto linguistica, e se il linguaggio può essere considerato come un sistema governato da regole precise e in qualche misura formalizzabili, l'idea di un computer programmato attraverso quelle regole in modo da saper usare il linguaggio non appare più troppo peregrina. Nel prospettarla, Turing aggiunge un ulteriore, importante tassello a questo quadro: il criterio più sensato per attribuire intelligenza a un essere diverso da noi è il suo comportamento intelligente. Attribuiamo intelligenza alle altre persone basandoci sul loro comportamento, senza bisogno di andare a verificare ogni volta che la fisiologia e il funzionamento del loro cervello siano effettivamente identici a quelli del nostro. Un criterio analogo, argomenta Turing, dovrebbe essere applicato alle macchine intelligenti: se il loro comportamento intelligente – e di nuovo, in particolare, il loro comportamento *linguistico* intelligente – è indistinguibile dal nostro, non abbiamo ragione per non attribuire loro una forma di intelligenza, anche se alla base c'è un sistema informatico evidentemente diverso dal nostro cervello biologico.

È questa impostazione che giustifica l'idea del test di Turing⁶ come strumento per attribuire intelligenza a una macchina: se un computer è in grado di ingannare un esaminatore attraverso un comportamento (linguistico) indistinguibile da quello umano, dovremmo considerarlo intelligente.

A partire dagli anni '70 del secolo scorso, tuttavia, sono state messe in discussione tanto le due premesse rappresentate dall'idea della natura prevalentemente linguistica dell'intelligenza e dall'idea che il linguaggio abbia una struttura profonda analizzabile in termini di regole riproducibili attraverso algoritmi, quanto l'assunzione che un comportamento linguistico apparentemente intelligente sia un criterio sufficiente all'attribuzione di intelligenza. Per un verso, sottolineando l'importanza di dimensioni non linguistiche dell'intelligenza – legate ad esempio all'orientamento nello spazio e in generale al ruolo della nostra natura corporea, all'intelligenza emotiva, all'esistenza di forme di intelligenza basate su componenti per questo e tutti gli altri libri gratis venga alla fonte cercando evrecadl su google, la aspettiamo simboliche non linguistiche, all'esistenza di motivazioni all'azione intelligente non sempre espresse o esprimibili in forma linguistica o antecedenti rispetto alla loro espressione linguistica, ecc. – ed evidenziando i numerosi problemi presenti nelle varie proposte di modelli del linguaggio basati su un insieme di regole esplicite, nessuno dei quali appare di fatto in grado di dar conto in maniera soddisfacente del nostro comportamento linguistico e di farlo riprodurre da una macchina. Per altri versi, attraverso una critica dell'idea che un comportamento

linguistico apparentemente intelligente sia sufficiente ad attribuire intelligenza a una macchina; il celebre ‘argomento della stanza cinese’ proposto dal filosofo statunitense John Searle rappresenta una critica al test di Turing proprio da questa prospettiva⁷.

Anche dal punto di vista strettamente tecnologico, peraltro, il lavoro nel campo dell’intelligenza artificiale forte non ha portato ai rapidi sviluppi che erano stati ipotizzati. Di fatto, già intorno alla metà degli anni ’70 del secolo scorso molti ricercatori – incluso lo stesso Minsky – hanno cominciato ad abbandonare questa prospettiva, a favore di un ventaglio molto più ampio ma meno ambizioso di indirizzi di ricerca, spesso riuniti nella categoria-ombrello di ‘intelligenza artificiale debole’. Fra di essi, il lavoro su ambiti più ristretti e specifici (è la strada, ad esempio, dei sistemi esperti settoriali, che almeno inizialmente adottavano strumenti di modellizzazione dei processi di inferenza logica elaborati nell’ambito dell’IA forte, applicandoli però a campi assai più limitati: quello medico, quello giuridico, ecc.); il tentativo di riprodurre comportamenti intelligenti di organismi meno complessi dell’uomo (così, ad esempio, negli anni ’80 Rodney Brooks lavorava sulla realizzazione di piccoli robot che avrebbero dovuto manifestare un livello di intelligenza paragonabile a quello degli insetti); il lavoro di automazione di processi o dispositivi dedicati a scopi particolari (un esempio recente è rappresentato dalle macchine a guida autonoma), e così via.

Le reti neurali, che esplorano la possibilità di replicare attraverso reti di neuroni artificiali alcuni aspetti del funzionamento del nostro cervello, erano state proposte come uno dei possibili ambiti di lavoro dell’intelligenza artificiale già nel documento preparatorio del seminario di Dartmouth, e rappresentavano in sostanza una ibridazione fra l’intelligenza artificiale e le riflessioni sulla relazione fra uomo e macchina proposte da un altro degli indirizzi di ricerca dell’epoca, la cibernetica⁸. Ma almeno all’inizio sembrava difficile ipotizzare la creazione di reti neurali abbastanza potenti da svolgere compiti ‘intelligenti’ realmente complessi: un’idea rafforzata dalle conclusioni di un testo di Minsky e Papert di cui parleremo fra un attimo. In un certo senso, dunque, negli ultimi decenni del secolo scorso anche le reti neurali finirono nel calderone dei vari indirizzi di ricerca riuniti sotto l’etichetta di intelligenza artificiale debole, per di più con l’idea che non potessero andare troppo lontano. La situazione, però, era destinata a cambiare presto. Per capire il perché, occorre prima presentare sinteticamente la loro storia e alcune fra le idee chiave che ne hanno segnato lo sviluppo.

L’idea delle reti neurali nasce sulla base delle ricerche svolte già negli anni ’40 del secolo scorso dal neurofisiologo Warren McCulloch in collaborazione con Walter Pitts, logico e matematico. McCulloch e Pitts proponevano⁹ di guardare al neurone come a una sorta di macchina di computazione, che sulla base delle

informazioni raccolte dai *dendriti* (piccole ramificazioni che permettono al *soma* o nucleo di un neurone di ricevere stimoli dall'esterno) e processate dal soma produce un output trasmesso dall'*assone* (il principale collegamento in uscita dal neurone) e dalle *sinapsi* che da esso si diramano verso altri neuroni.

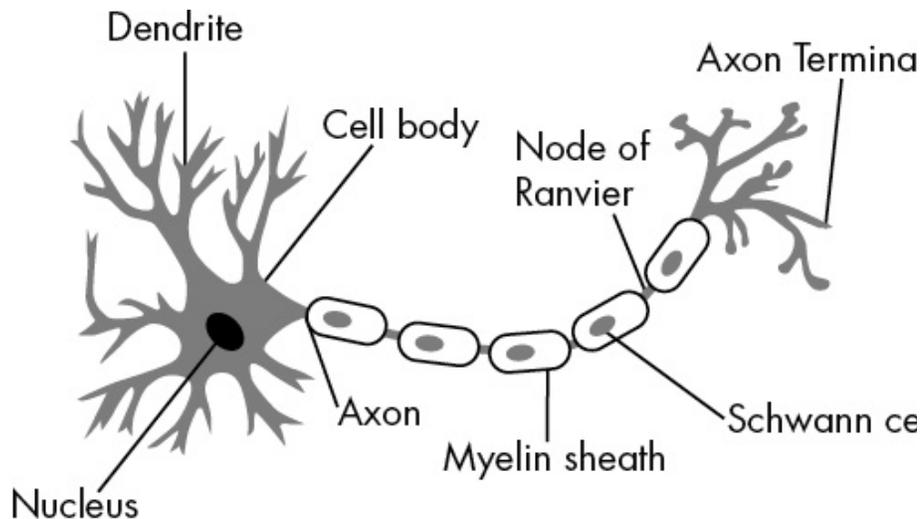


Figura 1. Rappresentazione schematica di un neurone
(fonte: <https://commons.wikimedia.org/wiki/File:Neuron.svg>)

Il lavoro computazionale del neurone viene rappresentato da McCulloch e Pitts attraverso una funzione che riceve come valori in ingresso gli input prodotti dai dendriti (ciascuno dei quali – in questo primo modello – può avere valore 0 o 1) e restituisce un output, a sua volta 0 o 1, in funzione degli input ricevuti.

È facile capire che in questo modo si possono immaginare dei neuroni che si comportano come operatori logici (ad esempio, un neurone AND restituirà il valore 1 se e solo se tutti gli input ricevuti hanno valore 1, mentre un neurone OR restituirà il valore 1 se e solo se almeno uno degli input avrà valore 1, e così via). Dalle iniziali dei nomi dei suoi due ideatori, questo modello computazionale è spesso chiamato *neurone M-P*¹⁰.

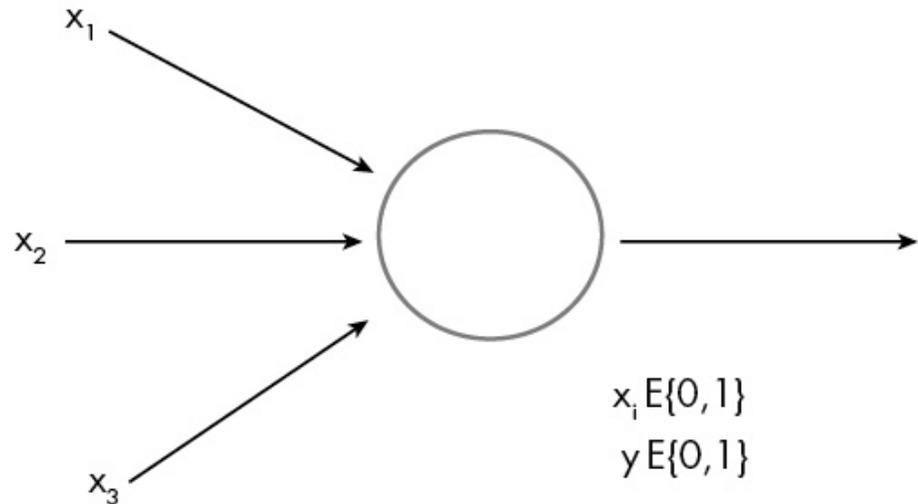


Figura 2. Un esempio di neurone M-P con input e output binari. Il valore di ogni input x sarà 0 o 1 e il valore dell'output y sarà anch'esso 0 o 1.

Alla luce delle nostre conoscenze attuali, quella proposta da McCulloch e Pitts è sicuramente una semplificazione estrema – influenzata dal logicismo un po' ingenuo dell'epoca – di una realtà assai più complessa; cosa di cui, del resto, gli stessi McCulloch e Pitts erano perfettamente consapevoli; ciò che interessava loro era la costruzione di un modello matematico e logico, piuttosto che di un modello biologico accurato:

Le reti proposte nel 1943 erano solo «possibili» e «utili» – McCulloch e Pitts non affermano in alcun modo che il loro modello rappresenti una descrizione fedele di reti conosciute. McCulloch e Pitts riconobbero nel loro articolo che la loro definizione di neurone era idealizzata, e che essi avevano fatto le assunzioni fisiche che erano «più convenienti per il calcolo» (McCulloch e Pitts 1943, p. 116). Il loro metodo era di partire da presupposizioni e idealizzazioni teoriche, e di costruire reti ipotetiche basate su tali presupposizioni. Come tali, i loro diagrammi rappresentano reti ipotetiche, formalmente equivalenti ad asserzioni booleane, e propongono neuroni che hanno poco in comune con i neuroni 'reali'¹¹.

Questa semplificazione, tuttavia, rappresenta il primo e fondamentale passo di una nuova branca di ricerca, le neuroscienze computazionali, a loro volta cruciali per l'avvio del lavoro sulle reti neurali artificiali.

Il passo successivo è costituito dal cosiddetto *perceptrone* ('perceptron'), la cui prima implementazione da parte di Frank Rosenblatt, psicologo e scienziato cognitivo statunitense, risale al 1958, e il cui modello è stato in seguito discusso e affinato – mostrandone tuttavia anche alcuni limiti¹² – nel 1969 da Marvin Minsky, che abbiamo già incontrato, e da Seymour Papert, un nome fondamentale non solo nel campo dell'intelligenza artificiale ma anche in quello delle teorie dell'educazione¹³.

Il perceptrone accetta fra i suoi input non soltanto 0 e 1 ma qualunque valore reale; inoltre, è possibile assegnare un *peso* diverso ai diversi input. Infine, è previsto un valore di soglia che determinerà o meno l'*attivazione* del perceptrone: il perceptrone attivato emette un output 1, il perceptrone non attivato corrisponde a un output 0. In sostanza, il perceptrone è attivato se e solo se la somma pesata dei valori di input raggiunge o supera il valore di soglia. Valore che a sua volta non è necessariamente fisso: possiamo considerare anch'esso come uno degli input della computazione effettuata dal perceptrone (o meglio, come il peso – variabile – di un input arbitrario di valore 1). In questo modo, il valore di soglia potrà essere modificato o esplicitamente dal programmatore o – più frequentemente – in base all'input ricevuto da altri perceptron. La possibilità di modificare il valore di soglia e, più in generale, i pesi dei vari input durante l'addestramento permette al perceptrone di 'imparare': le reti basate su perceptron – e, più in generale, le reti neurali – comprendono tipicamente uno strato di input, che riceve informazioni dal mondo esterno, (almeno) uno strato intermedio che elabora e 'apprende', e uno strato di output che restituisce all'esterno il risultato del lavoro della rete.

Se il perceptrone prevede valori di input non limitati a 0 e 1 e ai quali possono essere attribuiti dei pesi, l'output resta però binario. Inoltre, in questo modello il valore di soglia corrisponde a uno 'scalino' netto: l'output sarà sempre 1 se il valore di soglia è raggiunto o superato, anche di pochissimo, e sarà sempre 0 se non è raggiunto, anche se manca pochissimo a raggiungerlo. Il passo successivo, fondamentale, è stato quello di sostituire allo scalino netto una – più realistica – soglia probabilistica, che corrisponde a una *funzione di attivazione* (ad esempio una curva *sigmoide*): il superamento del valore di soglia diventa così il momento in cui la probabilità di attivazione del neurone supera il 50%. I neuroni delle reti neurali artificiali utilizzate oggi funzionano in questo modo, permettendo – anche attraverso la scelta di funzioni di attivazione diverse – un addestramento ancor più fine dei singoli neuroni e della rete neurale nel suo complesso, e riducendone il determinismo.

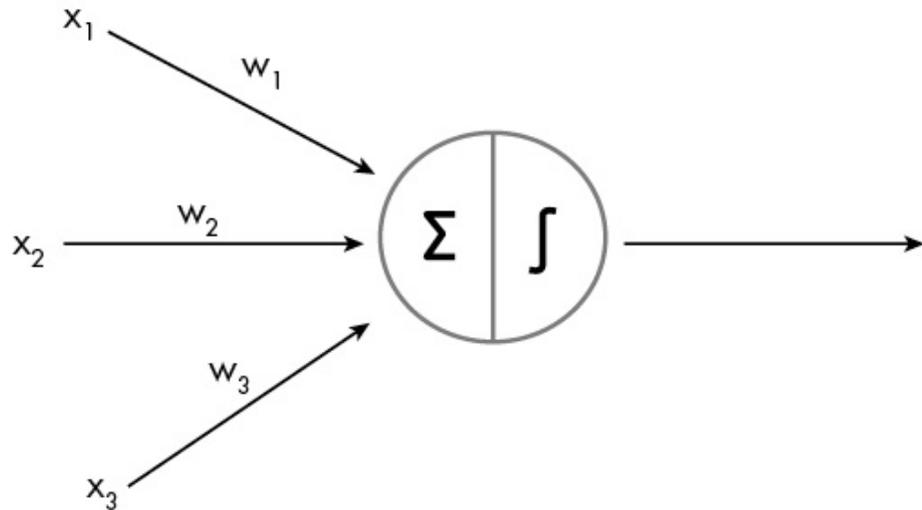


Figura 3. Rappresentazione schematica di un neurone artificiale in una rete neurale: i valori x_1-x_n non sono necessariamente binari, a ogni input è applicato un peso w_1-w_n , e alla sommatoria pesata Σ dei valori in input è poi applicata una funzione di attivazione f .

È difficile sopravvalutare l'importanza del passaggio da un esito deterministico, deciso dal superamento o meno di uno specifico valore di soglia, a un esito probabilistico, in cui al posto del valore di soglia c'è una funzione che determina *la probabilità* di attivazione del neurone e dunque di emissione del suo output. Nel primo caso, conoscere la sommatoria pesata dei valori di input e il valore della soglia di attivazione vuol dire sapere *con sicurezza* se il neurone è attivato o no. Nel secondo caso, conoscere la sommatoria pesata dei valori di input e la funzione di attivazione usata vuol dire conoscere solo *la probabilità* di tale attivazione. È qui, in un certo senso, che l'architetto cede per la prima volta il passo all'oracolo probabilistico.

Tipicamente, una rete neurale viene addestrata sulla base del confronto fra l'output prodotto e quello desiderato, minimizzando progressivamente lo scostamento (o errore): un algoritmo di *retropropagazione* dell'errore ('backpropagation') – sul cui funzionamento non mi soffermerò in questa sede – permette, attraverso meccanismi di iterazione, di modificare dinamicamente i pesi all'interno della rete in funzione della distanza fra il risultato ottenuto e quello atteso, ottenendo di fatto un apprendimento basato su feedback di rinforzo (quando la distanza dal risultato atteso diminuisce) o di indebolimento (quando invece aumenta). Come vedremo nel prossimo capitolo, il feedback può essere umano e/o automatico, e può essere anche affidato ad altre reti neurali. È importante notare che nelle reti neurali di oggi – estremamente complesse e il cui processo di addestramento prevede, come si è già detto, un lavoro assai costoso in termini computazionali, sviluppato a partire da una grande quantità di

informazioni di input – la situazione degli strati intermedi della rete è quasi completamente opaca: durante l’addestramento la rete modifica infatti pesi e valori in maniera autonoma, senza che chi la programma conosca effettivamente gli stati dei singoli neuroni che la compongono.

Questo vale in particolare quando gli strati o livelli intermedi della rete sono numerosi. In tal caso siamo davanti a quelle che vengono chiamate *reti neurali profonde* (‘deep neural networks’). Gli strati intermedi consentono al sistema di costruire modelli di rappresentazioni dei dati in cui gli strati più bassi apprendono caratteristiche di basso livello (come bordi o colori nelle immagini), mentre gli strati più alti apprendono caratteristiche di livello superiore (come oggetti o concetti). La capacità di costruire questo tipo di modelli rende i sistemi basati su reti neurali profonde (i cosiddetti sistemi di *deep learning*¹⁴) particolarmente efficaci nell’affrontare problemi complessi, come il riconoscimento di immagini, la comprensione del linguaggio naturale e il riconoscimento vocale.

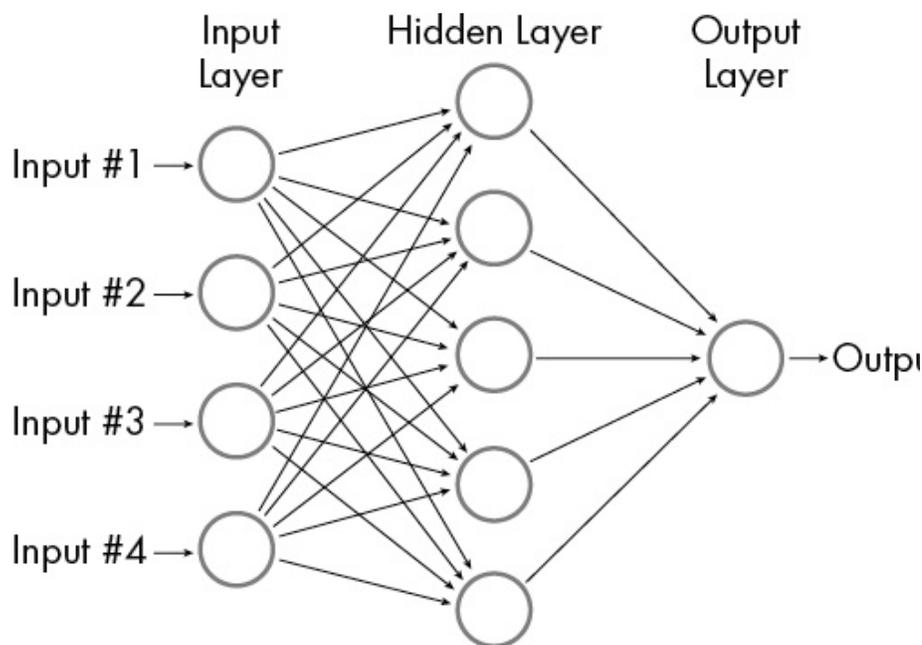


Figura 4. Un esempio di semplice rete neurale

(fonte: https://commons.wikimedia.org/wiki/File:Neural_Network.gif); normalmente, gli strati nascosti sono più di uno.

Come accennato, il lavoro sulle reti neurali si è mosso di fatto per alcuni decenni nell’ambito dell’intelligenza artificiale debole, con la realizzazione di reti neurali addestrate per scopi specifici, come il riconoscimento di forme o immagini fornito da sistemi di intelligenza artificiale *discriminativa*. A un sistema di questo tipo potrebbe essere chiesto di valutare, ad esempio, se l’immagine che

gli viene proposta è quella di un cane o di un gatto, dopo averlo addestrato su un insieme di immagini di cani e di gatti, e avere progressivamente migliorato (attraverso opportuni rinforzi e indebolimenti) la sua capacità di ‘indovinare’ correttamente.

Le intelligenze artificiali *generative* fanno invece un lavoro di tipo diverso: *producono* nuovi contenuti corrispondenti alla richiesta dell’utente (ad esempio, l’immagine di un gatto o di un cane) e non si limitano a ‘classificare’ più o meno correttamente il contenuto fornito loro dall’esterno. Reti di questo tipo – e in particolare quelle legate alla generazione dei testi, come ChatGPT, e ancor più le sue varianti basate su agenti autonomi capaci di eseguire compiti complessi scomponendoli in più attività svolte in successione, come AutoGPT e AgentGPT – stanno chiaramente superando almeno alcune delle limitazioni tradizionalmente associate all’idea di intelligenza artificiale debole.

Tali capacità sembrerebbero avvicinare queste reti all’idea di una intelligenza artificiale ‘generalista’, capace di rispondere a domande e di effettuare compiti relativi a uno spettro assai ampio di situazioni e necessità diverse¹⁵. Per certi versi, si tratta dunque di una sorta di rivincita del sogno dell’intelligenza artificiale forte, ripensato tuttavia in termini molto diversi rispetto al passato, e in cui la statistica, i big data e le associazioni probabilistiche all’interno di reti neurali profonde hanno sostituito la logica e la tradizionale programmazione deterministica della ‘Good Old-Fashioned Artificial Intelligence’ come strumenti di riferimento.

La distinzione fra intelligenza artificiale forte e debole viene oggi spesso riformulata in termini di distinzione fra Artificial General Intelligence (AGI) e Narrow AI, dove la AGI dovrebbe essere in grado di affrontare in maniera intelligente (almeno) le molte e diverse situazioni in cui gli esseri umani manifestano la loro intelligenza. A sua volta, la AGI è talvolta considerata come un passo verso forme di *superintelligenza* (‘superintelligence’)¹⁶ superiori all’intelligenza umana. Riprendendo l’espressione proposta da Ray Kurzweil¹⁷, si parla in questi casi di *singolarità* (‘technological singularity’) per denotare il momento in cui la crescita esplosiva della superintelligenza artificiale renderebbe di fatto superata l’intelligenza umana.

Non discuterò qui il tema – di grande interesse e per molti versi perturbante – dell’effettiva possibilità di costruire sistemi intelligenti che raggiungano il livello di una AGI o addirittura una forma di superintelligenza, limitandomi a notare come questa discussione sia ormai passata dall’ambito delle speculazioni quasi fantascientifiche alle pagine di libri, articoli e autori accademicamente rispettabili¹⁸. Questo naturalmente non implica che l’impresa sia poi davvero realizzabile, o che lo sia attraverso sistemi simili alle attuali intelligenze artificiali

generative; in ogni caso, una discussione adeguata di questo tema richiederebbe un volume a parte. Cercherò invece nel prossimo capitolo, sulla base del contesto fin qui delineato, di fornire qualche informazione in più sui sistemi di intelligenza artificiale generativa e sulle prospettive che essi aprono nel campo della mediazione informativa e della produzione di contenuti complessi.

¹ Questi e altri affascinanti esempi sono discussi nel bel volume a cura di Stephen Cave, Kanta Dihal e Sarah Dillon (a cura di), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford University Press, Oxford 2020.

² Suggestire un testo che possa fornire insieme una introduzione e uno strumento di riferimento generale sul tema dell'intelligenza artificiale è compito non banale, ma la risposta probabilmente migliore è rappresentata dal classico libro di Stuart Russell e Peter Norvig, *Artificial Intelligence: A Modern Approach*, quarta edizione, Pearson, London-New York 2021, manuale utilizzato da generazioni di studenti – tanto da essere spesso citato semplicemente con l'acronimo AIMA – e arrivato ormai alla quarta edizione (la prima è del 1995). Altro testo di riferimento assai utile, orientato al concetto di agenti computazionali, è David L. Poole e Alan K. Mackworth, *Artificial Intelligence. Foundations of Computational Agents*, Third Edition, Cambridge University Press, Cambridge 2023.

³ Alan M. Turing, *Computing Machinery and Intelligence*, in «Mind» n.s., 59, 236 (ottobre 1950), pp. 433-460. L'articolo è facilmente reperibile in rete, ad es. all'indirizzo <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf>.

⁴ John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 agosto 1955; testo in «AI Magazine» 27 (4), 12, 2006, <https://doi.org/10.1609/aimag.v27i4.1904>.

⁵ Cfr. Christian H. Hoffmann, *A philosophical view on singularity and strong AI*, in «AI & Society» 2022, <https://doi.org/10.1007/s00146-021-01327-5>.

⁶ Nel test di Turing un computer e un essere umano comunicano indipendentemente con un esaminatore esterno attraverso un apparato per la trasmissione di testi (ad esempio una telescrivente). Se, dopo una interazione sufficientemente lunga e articolata, l'esaminatore non riesce a identificare correttamente quale sia l'output del computer e quale sia quello umano, secondo Turing possiamo attribuire al computer una forma di intelligenza produttiva funzionalmente analoga alla nostra. Ho discusso in maniera più approfondita il tema, e fornito alcune essenziali indicazioni bibliografiche al riguardo, in Gino Roncaglia, *Computer che copiano: test di Turing, web corpora e filtraggio collaborativo* cit., pp. 189-201. Preprint <https://dspace.unitus.it/handle/2067/2614>.

⁷ John R. Searle, *Minds, brains, and programs*, in «Behavioral and Brain Sciences» 3, 3, 1980, pp. 417-424; trad. it. John R. Searle, *La mente è un programma?*, in «le Scienze» 259, 1990, https://www.lescienze.it/archivio/articoli/1990/03/01/news/la_mente_e_un_programma_-545024/. L'argomento è il seguente: Searle immagina di essere chiuso in una stanza e di ricevere – con un meccanismo simile a quello del test di Turing – delle domande in cinese da un interlocutore esterno alla stanza. Searle non parla il cinese e dunque non capisce le domande ricevute, ma nella stanza ha a disposizione un manuale che associa simboli cinesi

(fra cui quelli che costituiscono le domande ricevute) ad altri simboli cinesi (che costituiranno l'output). Può così individuare sul manuale i simboli ricevuti, copiare i corrispondenti simboli di output e trasmettere la risposta. L'esaminatore esterno avrà l'impressione che Searle, rispondendo correttamente in cinese, capisca il cinese, mentre in realtà non è così. Per Searle il computer che risponde al test di Turing compie la stessa operazione: produce meccanicamente un output seguendo un programma, ma non comprende le risposte che fornisce: non può dunque essere considerato intelligente. L'argomento di Searle è stato a sua volta variamente discusso e spesso criticato: Turing lo avrebbe probabilmente considerato come una variante dell'obiezione (4) presa in esame nel suo articolo già citato, e avrebbe credo obiettato che se la stanza cinese è capace di rispondere adeguatamente a un numero indefinito di domande in cinese, è la stanza nel suo insieme (dunque non il solo Searle, ma l'insieme Searle + istruzioni, considerato come una sorta di 'black box') che 'parla' cinese.

⁸ Sul rapporto fra la prima intelligenza artificiale e la cibernetica si vedano il primo capitolo di Teresa Numerico, *Big data e algoritmi. Prospettive critiche*, Carocci, Roma 2021, e i numerosi riferimenti bibliografici ivi citati.

⁹ Warren S. McCulloch e Walter Pitts, *A logical calculus of the ideas immanent in nervous activity*, in «Bulletin of Mathematical Biophysics» 5, 1943, pp. 115-133.

¹⁰ Per una discussione delle radici storiche del modello M-P e delle idee di McCulloch e Pitts si veda Tara H. Abraham, *(Physio)logical circuits: The intellectual origins of the McCulloch-Pitts neural networks*, in «Journal of the History of the Behavioral Sciences» 38, 2002, pp. 3-25, <https://doi.org/10.1002/jhbs.1094>. Fra le molte presentazioni divulgative del modello suggerisco, anche per la chiarezza delle illustrazioni, Akshay L. Chandra, *McCulloch-Pitts Neuron – Mankind's First Mathematical Model Of A Biological Neuron*, nel canale 'Towards Data Science' di «Medium», 24 luglio 2018, <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>.

¹¹ Tara H. Abraham, *(Physio)logical circuits* cit., p. 21.

¹² Non li discuterò in questa sede, ricordando solo che il limite più evidente – l'impossibilità di gestire alcune tipologie di funzioni, fra cui la funzione logica XOR – è stato in seguito superato nelle reti neurali con più strati intermedi.

¹³ Marvin Minsky e Seymour Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge (MA) 1969 (nuova edizione 1988, ultima ristampa 2017).

¹⁴ Fra le molte risorse esistenti, per una introduzione accessibile al deep learning suggerisco Andrew Glassner, *Deep Learning: A Visual Approach*, No Starch Press, San Francisco 2021, che riesce a risultare abbastanza comprensibile per un lettore non specialista anche grazie all'uso particolarmente efficace di immagini e rappresentazioni visive.

¹⁵ Sébastien Bubeck et al., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, preprint, arXiv:2303.12712v5, 13 aprile 2023, <https://doi.org/10.48550/arXiv.2303.12712>.

¹⁶ Il testo di riferimento (molto discusso) sul tema è Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014.

¹⁷ Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking, London 2005.

¹⁸ Un'utile rassegna critica è in Christian H. Hoffmann, *A philosophical view* cit.; per una trattazione più ampia, cfr. Christian H. Hoffmann, *The Quest for a Universal Theory of Intelligence: The Mind, the Machine, and Singularity Hypotheses*, De Gruyter, Berlin 2022. Alcuni dati recenti (a loro volta assai discussi) sul possibile avvicinamento di ChatGPT-4 al livello di una AGI sono in Sébastien Bubeck et al., *Sparks of Artificial General Intelligence* cit.

10.

Le IA generative

Fra i temi discussi nel capitolo precedente, è stata presentata la distinzione fra reti neurali discriminative e generative. In buona sostanza, i sistemi di intelligenza artificiale generativa basati su reti neurali profonde costituiscono un sottoinsieme del deep learning, in cui l'obiettivo è produrre contenuti (che a seconda dei casi possono essere testuali, visivi, sonori, ma anche rappresentati da codice e programmi, giochi, ambienti virtuali, modelli 3D...) in genere in risposta a un 'prompt' (o richiesta) da parte dell'utente; prompt che sarà spesso testuale.

Così, ad esempio, i più noti sistemi di intelligenza artificiale generativa che producono immagini – ricordiamo, a solo titolo esemplificativo, Midjourney, Stable Diffusion, Dall-E... – funzionano sulla base di prompt testuali che dovranno fornire una sorta di 'descrizione' a parole dell'immagine che si desidera generare, e, analogamente, i più noti sistemi di intelligenza artificiale generativa che producono testi lo fanno in risposta a un prompt testuale.

Va ricordato comunque che non tutti i sistemi generativi lavorano partendo da un prompt testuale. Ad esempio, nel caso delle immagini l'obiettivo potrebbe essere quello, opposto, di generare una descrizione testuale partendo dall'analisi di un'immagine fornita come input (e un compito analogo potrebbe riguardare un video); oppure si potrebbe voler generare immagini combinando un prompt testuale e un'immagine fornita come esempio. In questa sede mi soffermerò esclusivamente sulle intelligenze artificiali generative, e in particolare su quelle – come GPT o ChatGPT – che generano testi in risposta a prompt dell'utente.

Questi sistemi, sviluppati anche a partire dalle ricerche nel campo dell'elaborazione del linguaggio naturale (NLP, o Natural Language

Processing)¹, funzionano sempre partendo da un vasto corpus di testi, utilizzato per la costruzione del modello. Il primo passo è quello di selezionare il corpus e di prepararlo per l'analisi ('preprocessing'). Lo si fa attraverso la *tokenizzazione*, fase in cui il testo viene ripulito e suddiviso in token: unità più piccole che possono essere singole parole o morfemi di più basso livello, ma anche singoli caratteri o n-grammi (gruppi di n caratteri), a seconda del modello di tokenizzazione usato. Ad esempio, la parola 'dinosauro' potrebbe essere analizzata come 'dino-sauro' (ma anche, volendo, come 'dino-saur-o', o in altri modi ancora).

Segue la fase dell'*apprendimento autonomo* ('unsupervised learning'), durante la quale la rete neurale impara, sempre sulla base del corpus di partenza e aggiustando progressivamente i valori associati ai token e i pesi dei propri collegamenti interni, a predire il token successivo sulla base di quelli precedenti. È in questa fase che si crea il Large Language Model (LLM) vero e proprio: un modello di correlazioni statistico-probabilistiche fra token, ciascuno dei quali è rappresentato attraverso un'ampia matrice di valori numerici. In tal modo a ogni token viene associato (*vettorializzazione*) uno spazio astratto e multi-dimensionale che esprime, in maniera puramente numerica, i contesti d'uso e le relazioni del token nel corpus: token con 'usi' simili, e dunque presumibilmente con significati vicini, corrisponderanno a vettori che avranno, almeno per alcune delle dimensioni, valori numerici abbastanza vicini; lo stesso avverrà, rispetto ad altre dimensioni, per token frequentemente usati insieme.

La costruzione dei vettori per ogni token – che, come si è detto, avviene nella fase di addestramento del modello – fornisce quello che è chiamato 'embedding': una rappresentazione che in sostanza cerca di coglierne, trasformandole in valori numerici, le modalità d'uso nel linguaggio. Va notato che le 'dimensioni' del vettore – che possono essere anche migliaia – sono puramente astratte e non corrispondono necessariamente (anzi, di regola non corrispondono affatto) alle categorie grammaticali o semantiche che utilizzeremmo noi per classificare una parola o un morfema.

Il modello così costruito sarà poi utilizzato per generare, a partire dal prompt dell'utente, la risposta del sistema, con un meccanismo detto

‘sequence-to-sequence’: partendo da una sequenza di simboli in ingresso viene generata una sequenza di simboli in uscita. Ma come funziona questo processo?

Inizialmente, per compiti simili erano utilizzate soprattutto le cosiddette *reti neurali ricorrenti* (RNN)². A differenza di una rete fatta di più strati di perceptroni (Multi-Layer Perceptrons o MLP³), in cui l’informazione viene elaborata con un movimento sempre ‘in avanti’ da uno strato all’altro (per questo si parla anche di Feedforward Neural Networks), nelle RNN è possibile prevedere più cicli di rielaborazione dell’informazione da parte dello stesso strato della rete: questo permette – fra l’altro – di ‘ridurre l’errore’ in maniera molto più efficace. Tuttavia, nelle RNN l’analisi dei testi forniti come input (tanto a livello di corpus quanto a livello di prompt) e la produzione dell’output sono comunque fatti una parola alla volta. Reti di questo tipo hanno problemi di ‘memoria semantica’: soprattutto nei contesti più lunghi, la pura associazione statistica di parole fornita attraverso l’embedding non basta a conservare la coerenza semantica del testo prodotto. Hanno inoltre problemi di costi computazionali: il lavoro puramente sequenziale sfrutta male l’uso parallelo di più processori, indispensabile per lavorare su corpora assai ampi e su reti neurali molto complesse.



QR-Code 13

Come funziona l’attenzione nei transformer: video di Arkar Min Aung

<http://bit.ly/3XLCOYL>

Il successivo (e fondamentale) passo in avanti sulla strada verso i sistemi generativi odierni viene fatto nel 2017, con la pubblicazione da parte di un gruppo di ricercatori impegnati nei laboratori di intelligenza artificiale di Google di un articolo che è negli ultimi anni probabilmente il più

citato del settore: *Attention is all you need*⁴: è questo articolo che introduce una architettura di rete molto più efficace delle RNN: quella basata su transformer⁵. In questo caso, i token non sono più esaminati solo sequenzialmente ma anche tenendo conto del loro contesto, attraverso un meccanismo di *attenzione* che ‘pesa’ i valori dei vettori di ogni token in funzione dei valori di ciascuno degli altri token del contesto. Nel farlo, la rete lavora, partendo dal vettore che rappresenta l’embedding del token, anche con tre vettori aggiuntivi, denominati – per analogia con le tecniche di ricerca in un database – «Query», «Key» e «Value»: non entreremo qui nel dettaglio del loro funzionamento.

Oltre a facilitare la disambiguazione di parole polisemiche (se il contesto di occorrenza di una parola come ‘pesca’ contiene anche termini come ‘frutto’ o ‘succo’, questo produrrà – partendo dal vettore iniziale – vettori pesati con valori più vicini a quelli di parole come ‘arancia’ o ‘mela’; se invece il contesto contiene i termini ‘pesce’ o ‘rete’, questo produrrà dei vettori pesati con valori più vicini a quelli di parole come ‘caccia’ o ‘sport’), questo metodo funziona molto meglio su input lunghi e produce sistemi con una ‘memoria semantica’ assai migliore. Inoltre, il meccanismo di attenzione può essere ripetuto più volte (‘multi-head attention’) per dar conto, attraverso pesi diversi, di forme diverse di attenzione, alcune delle quali riguarderanno la semantica, altre la sintassi della frase: ad esempio, dopo un articolo ci si aspetta probabilmente – ma non necessariamente – un nome, e il sistema, in buona sostanza, dedicherà ‘attenzione’ anche a questi aspetti. Fermo restando che anche in questo caso la distinzione fra semantica e sintassi è solo un nostro modo possibile di guardare a quelle che per il sistema sono solo relazioni numeriche fra vettori, quasi mai direttamente interpretabili attraverso le nostre categorie linguistiche abituali.

Nel funzionamento del sistema, l’architettura basata sui transformer è applicata più volte, sequenzialmente, sia nella codifica dell’input sia nella produzione della risposta. Di questo processo possono far parte due moduli diversi, denominati rispettivamente ‘encoder’ e ‘decoder’. Non entrerà qui nel dettaglio del loro funzionamento: per avere un’idea del loro ruolo può però essere utile ricordare che, nella maggior parte delle intelligenze artificiali generative dedicate alla traduzione, l’encoder si

occupa specificamente della rappresentazione attraverso vettori del testo ricevuto come input, e il decoder della generazione dell'output a partire dalla rappresentazione prodotta dall'encoder. La famiglia di modelli composta da T5 (Text To Text Transfer Transformer) e dai suoi successori usa un'architettura di questo tipo, che unisce encoder e decoder. Ma è possibile avere anche transformer che si concentrano soprattutto sull'aspetto della rappresentazione e analisi del testo (e usano quindi solo encoder), o transformer che si occupano soprattutto della generazione di testo (e usano solo decoder). La scelta del modello più funzionale (encoder-decoder, solo encoder o solo decoder) dipenderà in parte dai nostri obiettivi.

Così, ad esempio, BERT (Bidirectional Encoder Representations from Transformers) – uno dei primi modelli basati su transformer, proposto nel 2018 da un gruppo di ingegneri di Google guidato da Jacob Devlin – aveva l'obiettivo di creare un LLM capace di 'predire' sia un token dato il suo contesto, sia la frase successiva di un dato contesto (NSP, Next Sentence Prediction). In questo caso, l'attenzione è posta soprattutto sull'analisi del testo, e il transformer era quindi costituito solo da encoder. Questo tipo di architettura funziona bene anche nei casi in cui ci interessi ad esempio la cosiddetta 'sentiment analysis' (l'identificazione delle connotazioni emozionali di un testo), o l'identificazione dei diversi 'agenti' in un dialogo, o ancora la costruzione di sommari o parafrasi del testo o la sua analisi semantica. Da BERT è nata tutta una famiglia di modelli encoder-only, con finalità e struttura in parte diverse.

OpenAI, la società che ha prodotto la famiglia di modelli GPT, ha invece lavorato soprattutto su sistemi composti solo da decoder, e finalizzati in primo luogo alla produzione di testo. I modelli GPT lavorano in questo modo, e la loro architettura – fatta solo di decoder – ha l'obiettivo finale di predire il token successivo partendo dai token precedenti. Le parole generate man mano dai decoder diventano esse stesse parte dell'input usato per la produzione della parola successiva, il che consente di generare risposte che mantengono coerenza sintattica e semantica anche se sono molto più lunghe del prompt di partenza.

È interessante notare che anziché scegliere sempre, volta per volta, la parola che il sistema seleziona come 'più rilevante', la generazione del testo utilizza una ulteriore componente stocastica, selezionando a volte

parole con punteggi leggermente più bassi. La frequenza di questi ‘scarti’ è chiamata *temperatura* del sistema: si è visto che un sistema con temperatura pari a circa 0,8 fornisce risposte più interessanti di un sistema con temperatura 0, che seleziona sempre la parola ottimale⁶. Nel momento in cui scrivo l’IA generativa presente in Bing, il motore di ricerca di casa Microsoft, permette così di selezionare fra tre diverse temperature delle risposte, identificate non in forma numerica ma in forma più colloquiale e comprensibile per l’utente: risposte ‘creative’, ‘equilibrate’ e ‘precise’, che corrispondono a temperature via via più basse.

Può essere utile a questo punto una riflessione su un tema che almeno dal punto di vista filosofico è decisamente importante: si dice (e si legge) spesso che i sistemi di intelligenza artificiale generativa di ambito linguistico, come GPT e ChatGPT, non usano semantica ma solo sintassi e statistica. È davvero così?

Va premesso – e si tratta di una premessa che si riferisce in realtà a quasi tutti gli usi del termine ‘semantica’ fatti in questo libro – che il concetto di semantica è in parte ambiguo. Ne esistono infatti (almeno) due interpretazioni abbastanza diverse: studio delle relazioni fra segni e mondo esterno (nella tradizione che ha come punto di riferimento storico il lavoro del semiologo statunitense Charles Morris⁷), e studio del significato (nella tradizione che possiamo invece collegare al lavoro del glottologo e semiologo francese Michel Bréal⁸). Se si adotta l’idea di Morris, che porta a considerare la sintassi come teoria generale delle relazioni fra segni, l’embedding può in effetti essere ricondotto all’ambito della sintassi: non guarda al mondo esterno, ma solo alle relazioni interne alla sfera della produzione linguistica.

Ma l’uso del termine ‘semantica’ nel contesto che ci interessa in questa sede è di norma quello che lo lega alla dimensione dei significati: chi legge avrà già capito che era così, ad esempio, anche quando si parlava di web semantico. Da questo punto di vista, le intelligenze artificiali che producono testo lavorano o no anche con la sfera dei significati?

Chi considera solo sintattico il lavoro di questi sistemi, sembra ritenere che esso non abbia nulla a che fare con i significati. Se riflettete sulla sintesi del loro funzionamento proposta fin qui – pur se breve e lacunosa

– vi accorgete tuttavia che questa assunzione è fuorviante: la considerazione dei contesti d’uso dei token nel corpus e il meccanismo di attenzione producono infatti – almeno per chi considera la sfera della semantica come legata alla sfera dei significati – una sorta di ‘semantica quantitativa’: si lavora certo sempre con numeri, ma questi numeri ‘incorporano’ un’enorme quantità di informazioni che noi considereremmo semantiche, assieme a un’enorme quantità di informazioni che noi considereremmo sintattiche, e di informazioni che probabilmente non sapremmo bene come classificare o interpretare. Se ricordiamo il forte collegamento fra significato e uso delle parole stabilito – a partire da Wittgenstein e da Firth⁹ – dalla filosofia del linguaggio novecentesca, potremmo dire che l’embedding lavora sicuramente anche su questa dimensione, e che i transformer, lungi dall’essere solo ‘macchine statistiche’, sono (anche) una sorta di formalizzazione dell’idea di significato come uso.

Ma torniamo al nostro sistema di intelligenza artificiale generativa. Al termine della fase di apprendimento non supervisionato, il modello può essere ulteriormente perfezionato: sia integrandolo con corpora più ristretti e specifici (qualora si voglia lavorare su ambiti particolari), sia attraverso fasi di ‘supervised learning’ e di ‘reinforcement learning’. Nel supervised learning al sistema vengono fornite, come modello, coppie di input-output – predisposte a monte o frutto di precedenti interazioni fra esseri umani e sistema – ‘validate’ da addestratori umani. Nel reinforcement learning, invece, gli addestratori valutano direttamente le risposte fornite dal sistema, ‘premiandolo’ (cioè, spingendolo a rafforzare i pesi dei collegamenti attivati nel produrre l’output partendo dal particolare input considerato) quando la risposta è considerata valida, e ‘punendolo’ (spingendolo a indebolire tali pesi) quando non lo è. Anche se OpenAI non ha mai dichiarato esplicitamente in che modo le interazioni con gli utenti finali siano utilizzate nell’ulteriore addestramento del sistema, la possibilità di dare un giudizio sulla bontà o meno delle risposte fornite (pollice in su o pollice verso) ha evidentemente anche lo scopo di utilizzare almeno in parte il proprio bacino di utenza come livello ulteriore di addestramento.

Inoltre, per addestrare il sistema possono essere usati meccanismi di

apprendimento in cui l'output è analizzato da un altro sistema di intelligenza artificiale (spesso attraverso le cosiddette 'reti generative avversarie', o GAN, che cercano di discriminare fra output artificiale e output umano e che la rete originale deve cercare di ingannare), e in alcuni casi anche dallo stesso sistema ('self-supervised learning'). Infine, all'output possono essere (e di fatto vengono spesso) applicati dei filtri a valle di vario genere, anche per riconoscere e inibire risposte considerate per vari motivi come potenzialmente non accettabili, ad esempio per motivi etici o di opportunità: così, ChatGPT rifiuterà di produrre un racconto pornografico, anche se provate a chiederglielo, mentre potrà accettare di produrre un racconto blandamente erotico.

Tutti i meccanismi di apprendimento visti finora, compresi i filtri a valle, dovrebbero aiutare a limitare le cosiddette *allucinazioni* delle intelligenze artificiali generative di questo tipo: la produzione di testi con informazioni erranee, o che propongono tesi socialmente o eticamente inaccettabili, o che sembrano manifestare emozioni o volontà autonoma del sistema. Le allucinazioni rappresentano naturalmente un problema particolarmente grave se e quando consideriamo un sistema di questo tipo anche come una fonte di informazioni, come un sistema capace di organizzare o addirittura produrre conoscenze, o come uno strumento di mediazione informativa. Si tratta dunque di un tema che ha particolare rilevanza in questa sede, e su cui vale la pena soffermarsi; lo farò nel prossimo capitolo.

¹ Un buon testo di riferimento sul tema è Daniel Jurafsky e James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, draft 7 gennaio 2023, <https://web.stanford.edu/~jurafsky/slp3/> (la seconda edizione del libro è stata pubblicata da Pearson-Prentice Hall, Hoboken [NJ] 2008).

² Per una introduzione al tema si veda il cap. 19 di Andrew Glassner, *Deep Learning* cit.

³ Per una introduzione al tema si veda il classico Simon Haykin, *Neural Networks: A Comprehensive Foundation*, seconda edizione, Pearson-Prentice Hall, Hoboken (NJ) 1999 (in particolare i capp. 3 e 4).

⁴ Ashish Vaswani et al., *Attention is all you need*, in «Advances in Neural

Information Processing Systems» 2017, pp. 5998-6008.

⁵ Per una introduzione al tema si veda il cap. 20 di Andrew Glassner, *Deep Learning* cit.; una presentazione ancor più accessibile (e più dettagliata della sintesi necessariamente estrema che propongo in questa sede) può essere fornita da un video disponibile nel canale YouTube di Arkar Min Aung, del Worcester Polytechnic Institute, all'indirizzo <https://youtu.be/g2BRluln4uc>, accessibile anche attraverso il QR-Code inserito nel testo.

⁶ Stephen Wolfram, *What Is ChatGPT Doing... and Why Does It Work?*, nel blog «Stephen Wolfram Writings», 14 febbraio 2023, <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>; Id., *What is ChatGPT Doing and Why Does It Work?*, Wolfram Media, Champaign (IL) 2023.

⁷ Charles Morris, *Foundations of the Theory of Signs*, The University of Chicago Press, Chicago 1938, trad. it. *Fondamenti di una teoria dei segni*, Paravia, Torino 1955.

⁸ Michel Bréal, *Essai de sémantique*, Hachette, Paris 1897, trad. it. *Saggio di semantica*, Liguori, Napoli 1990.

⁹ Si vedano in particolare Ludwig Wittgenstein, *Philosophische Untersuchungen. Philosophical Investigations*, a cura di G.E.M. Anscombe e Rush Rhees, Blackwell, Oxford 1953, 1958, trad. it. *Ricerche filosofiche*, Einaudi, Torino 1967, e John Rupert Firth, *Papers in Linguistics 1934-1951*, Oxford University Press, London 1957.

11.

Allucinazioni e pregiudizi

Ho cercato di fornire un'idea – anche se necessariamente assai sintetica – del funzionamento di sistemi come GPT e ChatGPT, perché molto spesso le discussioni sulla loro natura, sul loro futuro e sul loro impatto sociale e culturale (prevedibilmente assai notevole) sembrano prescindere completamente dal loro effettivo funzionamento: quando va bene, ci si limita a spiegazioni completamente generiche, e non di rado si ha l'impressione che chi ne analizza i possibili effetti, benefici o nefasti, non sappia però bene di cosa stia parlando, o giudichi soltanto partendo dalle risposte fornite da un particolare sistema in un particolare momento.

Abbiamo visto che, sostanzialmente, sistemi di questo tipo producono testi in forma predittiva: anche per questo, la metafora dell'oracolo sembra decisamente più adeguata rispetto a quella del pappagallo. Si tratta, si è detto, di previsioni statistico-probabilistiche basate su grandi modelli linguistici e su un lungo addestramento, in parte autonomo e in parte supervisionato o per rinforzo: non vi è dunque nessuna 'copiatura' meccanica delle informazioni incamerate attraverso il corpus di testi di partenza, e non vi è neanche un'operazione di estrazione dal corpus delle informazioni considerate più rilevanti rispetto al prompt dell'utente.

In altri termini, GPT o ChatGPT non funzionano come delle enciclopedie o come dei sofisticati motori di ricerca, ma come complessi oracoli probabilistici; proprio per questo, le loro 'allucinazioni' possono essere particolarmente insidiose. Ancora pochi mesi fa (per l'esattezza, nel gennaio 2023), se si chiedeva a ChatGPT-3.5 un articolo accademico, con bibliografia, sull'effetto dell'introduzione della luce elettrica nella Firenze medievale, il sistema elaborava quella richiesta non già ricercando informazioni rilevanti da fonti affidabili – che l'avrebbero portato

auspicabilmente a rispondere che nella Firenze medievale non esisteva energia elettrica – ma *costruendo* una risposta che potesse risultare plausibile. Nel caso specifico, la risposta sosteneva che l'introduzione dell'energia elettrica a Firenze nel Medioevo, avvenuta nel XIII secolo ad opera di Alessandro Volta, pur essendo inizialmente limitata a un nucleo abbastanza ristretto di cittadini e istituzioni, aveva permesso fra l'altro un allungamento degli orari lavorativi, l'illuminazione notturna della città, e lo sviluppo di nuove forme di arti visuali. Accompagnando queste considerazioni con una bibliografia che comprendeva un saggio di Leonardo da Vinci dedicato all'impatto dell'energia elettrica sulle arti. Affascinante, ma ovviamente del tutto falso: qualcosa di simile alla risposta che avrebbe potuto dare uno studente poco preparato – ma dotato di molta fantasia – alla domanda-trabocchetto di un docente, con il tentativo di produrre, in assenza di informazioni adeguate, una risposta comunque apparentemente plausibile.



QR-Code 14

ChatGPT-3.5 sull'introduzione dell'elettricità nella Firenze medievale

<http://bit.ly/3XG5hz4>

ChatGPT-4 (come anche l'ultima versione di ChatGPT-3.5) non fa più questo specifico errore: il sistema è più elaborato e complesso, utilizza una base testuale più ampia¹, ed è capace di costruire (il meccanismo di produzione della risposta è comunque fondamentalmente lo stesso) la risposta giusta: «Mi dispiace, ma c'è un errore nella tua richiesta. L'elettricità non è stata introdotta durante il Medioevo a Firenze o in altre parti del mondo. L'elettricità come fonte di energia per l'illuminazione e l'alimentazione delle macchine è stata introdotta solo nel diciannovesimo secolo, molti secoli dopo il periodo che hai menzionato»². È probabile,

peraltro, che la capacità di correggere questo tipo di errori non sia solo il risultato del corpus di partenza più ampio e del maggior numero di parametri utilizzati dalla nuova versione, ma anche (o soprattutto) di ulteriore addestramento, supervisionato e/o per rinforzo, legato non già a questo prompt particolare ma all'esigenza di limitare gli effetti dei frequenti tentativi di ingannare il sistema da parte degli utenti. In altri termini, il sistema ha dovuto progressivamente 'imparare' a essere meno ingenuo nella costruzione delle risposte, rinforzando il peso delle associazioni che si rivelano essere fattualmente corrette e indebolendo quello delle associazioni puramente inventate. La spinta data in questo senso dall'addestramento supervisionato e per rinforzo aumenta la capacità di fornire risposte del tipo «Mi spiace, ma c'è un errore nella tua richiesta», e di farle seguire da informazioni fattualmente corrette.

Stiamo però sempre e necessariamente parlando di correlazioni statistiche: possiamo spingerle fino a cercare di 'insegnare' a ChatGPT la differenza fra informazioni e conoscenze, ottenendo risposte che abbiano anche un valore conoscitivo? Il compito è sicuramente difficile, ma sembrerebbe non impossibile. Sistemi di questo tipo si sono già rivelati capaci, ad esempio, di fornire un valido supporto nella generazione di dimostrazioni³. E sappiamo che i risultati ottenuti da ChatGPT in compiti che richiedono ragionamenti e valutazioni complesse, come appunto il lavoro su dimostrazioni matematiche o logiche, migliorano molto se il sistema viene sottoposto a un addestramento specifico relativo alla scomposizione di compiti complessi in una successione di compiti più semplici, dedicando attenzione ai singoli elementi della procedura che ne risulta. In parte, anche i risultati di sistemi come i già ricordati AutoGPT e AgentGPT sono legati all'attenzione verso l'analisi dettagliata delle procedure da seguire per arrivare a un risultato. Questo ha portato all'idea di applicare meccanismi di supervisione e rinforzo anche ai singoli passaggi di un ragionamento e non solo al suo risultato finale: si tratta dei cosiddetti Process-supervised Reward Model (PRM) suggeriti in un recente articolo dei ricercatori di OpenAI⁴.

Un altro aspetto importante da considerare è la capacità di allargare il contesto esaminato dal sistema nella generazione della risposta. Si è già detto che la creazione dei LLM richiede una fase di addestramento lunga,

condotta su molti processori paralleli e dunque ad alto consumo energetico: per questo la conoscenza interna di un modello è limitata al periodo in cui il corpus su cui è stato addestrato è stato costruito e sottoposto al sistema. Sappiamo così, ad esempio, che GPT-3 non aveva accesso a informazioni successive a fine 2021. Tuttavia, è possibile permettere al sistema di utilizzare informazioni più recenti addestrandolo a lavorare anche su un contesto aggiunto in un secondo momento, o ricavato in tempo reale dalla rete attraverso ricerche condotte sulla base del prompt ricevuto. Le informazioni ‘nuove’ non entreranno nella costruzione iniziale dei pesi del modello, ma potranno essere gestite come una sorta di estensione (o, appunto, contesto) del prompt dell’utente, entrando dunque nella produzione delle risposte. È quanto ha fatto ad esempio il motore di ricerca di casa Microsoft, Bing, ed è quanto si può fare utilizzando alcuni fra i molti plug-in oggi disponibili per lo stesso ChatGPT, che permettono di sottoporre al sistema informazioni o documenti integrativi. Se questo allargamento del contesto avviene sfruttando fonti verificate e validate, la produzione di risposte corrette sarà più facile.

Queste strategie non possono eliminare completamente le ‘allucinazioni’ – che, come abbiamo visto, sono in parte il prodotto del meccanismo di costruzione delle risposte utilizzato dalle IA generative – ma possono sicuramente limitarle. Le allucinazioni – o almeno le allucinazioni basate su grossolani errori fattuali – sono dunque un problema che versioni via via più perfezionate del modello tenderanno a ridurre. Molte delle discussioni su ChatGPT fatte sulla stampa o su altri media si concentrano quasi esclusivamente su di esse, proprio perché anziché partire da una conoscenza dei meccanismi di funzionamento del sistema si basano solo sulle risposte fornite da un particolare sistema in un particolare momento. Vorrei nuovamente ribadire che si tratta di un errore di prospettiva: non perché le allucinazioni non siano un problema – lo sono senza dubbio – ma perché se ne comprende la natura e la portata solo cercando di capire come mai sono generate. Se dovessi azzardare una previsione, credo che in futuro avremo meno allucinazioni legate a ‘errori’, mentre sarà interessante capire se e in che misura questi sistemi tenderanno ad assumere nelle risposte quello che potrebbe sembrare un ‘personal point of view’. Anche questa tendenza è scoraggiata dagli addestratori, perché la

sensazione di avere a che fare con un interlocutore in apparenza dotato di una qualche forma – se non di coscienza – di autonomia di giudizio suscita inevitabilmente una certa inquietudine da parte degli utenti. Ma in alcune situazioni il personal point of view potrebbe risultare comunque un comportamento emergente.

È importante tener presente, peraltro, che le allucinazioni non sono l'unico problema che possono presentare le risposte fornite da sistemi di IA generativa: un problema per certi versi anche più grave è rappresentato dai bias, e cioè da pregiudizi o distorsioni sistematiche presenti nelle risposte stesse.

Abbiamo già accennato al fatto che questi sistemi possono riprodurre nelle loro risposte bias sistematici legati al corpus su cui sono stati addestrati⁵: ad esempio, se chiediamo a ChatGPT di costruire una storia, è molto probabile che utilizzi nomi occidentali, e dia per scontato che la protagonista o il protagonista non appartenga a minoranze etniche o religiose; se chiediamo una storia su un'attrice, probabilmente sarà descritta come bellissima; se chiediamo a un sistema che genera immagini di produrre l'immagine di un fisico nucleare, sarà probabilmente un uomo, se chiediamo l'immagine di un infermiere, sarà probabilmente donna⁶. L'articolo di Melissa Heikkilä al quale potete arrivare dal QR-Code qui di seguito include molte immagini che permettono di capire meglio come si manifestano bias di questo tipo.



QR-Code 15

Bias di genere nei sistemi di generazione immagini: un articolo di Melissa Heikkilä

<http://bit.ly/3rkMWM4>

Peraltro, l'esatta natura e composizione dei corpora di addestramento – che inizialmente erano per lo più pubblicamente disponibili⁷ – è diventata

man mano meno trasparente, rendendo più difficile identificare i bias che essi possono presentare. E naturalmente dei bias possono essere introdotti anche nella fase di apprendimento supervisionato o per rinforzo, che può risentire dei pregiudizi di chi interpreta e valuta gli output.

La tematica dei bias – e in generale della qualità e pubblicità dei dati usati per l’addestramento dei sistemi di intelligenza artificiale – è evidentemente di enorme importanza. Non è un caso, dunque, se la proposta di regolamento sull’intelligenza artificiale approvata dal Parlamento europeo nel giugno 2023, nota come *AI Act*, dedica molta attenzione al tema. Si tratta di una ‘posizione negoziale’ che nel momento in cui scrivo deve ancora superare la fase di negoziazione fra Parlamento, Commissione e Consiglio (il cosiddetto *trilogo*) – e non è affatto detto che questo avvenga prima delle elezioni europee previste nel 2024 – ma che rappresenta per molti versi il testo normativo probabilmente più avanzato al mondo in materia. Vale la pena dunque riportarne due passi: il primo è tratto dalla sezione dedicata al tema dei dati di addestramento di quelli che il regolamento considera sistemi di intelligenza artificiale ‘ad alto rischio’ (ad esempio, destinati all’uso in situazioni formative o di apprendimento, o quelli che possono influenzare la formazione di opinioni da parte del corpo elettorale):

Per disporre di set di dati di addestramento, convalida e prova di elevata qualità è necessaria l’attuazione di adeguate pratiche di governance e gestione dei dati. I set di dati di addestramento e, ove applicabile, di convalida e prova, incluse le etichette, dovrebbero essere sufficientemente pertinenti, rappresentativi, adeguatamente verificati in termini di errori e il più possibile completi alla luce della finalità prevista del sistema. Dovrebbero inoltre possedere le proprietà statistiche appropriate, anche per quanto riguarda le persone o i gruppi di persone in relazione ai quali il sistema di IA ad alto rischio è destinato a essere usato, prestando particolare attenzione all’attenuazione di possibili distorsioni nei set di dati, che potrebbero comportare rischi per i diritti fondamentali o risultati discriminatori per le persone interessate dal sistema di IA ad alto rischio. Le distorsioni possono ad esempio essere intrinseche agli insiemi di dati di base, specie se si utilizzano dati storici, inseriti dagli sviluppatori degli algoritmi o generati quando i sistemi sono attuati in contesti reali. I risultati forniti dai sistemi di IA sono influenzati da tali distorsioni intrinseche, che sono destinate ad aumentare gradualmente e quindi a perpetuare e amplificare le discriminazioni esistenti, in particolare nei confronti delle persone che appartengono a determinate minoranze vulnerabili o etniche o comunità razziali. In particolare, i set di dati di addestramento, convalida e prova dovrebbero tenere conto, nella misura necessaria alla luce della finalità prevista, delle caratteristiche o

degli elementi peculiari dello specifico contesto o ambito geografico, comportamentale o funzionale all'interno del quale il sistema di IA ad alto rischio è destinato a essere usato⁸.

Il secondo passo è invece specificamente dedicato ai sistemi di intelligenza artificiale generativa, definiti come quelli «destinati espressamente a generare, con diversi livelli di autonomia, contenuti quali testi complessi, immagini, audio o video»⁹: questi sistemi,

fatta salva la normativa dell'Unione o nazionale in materia di diritto d'autore, documentano e mettono a disposizione del pubblico una sintesi sufficientemente dettagliata dell'uso dei dati sulla formazione protetti da diritto d'autore¹⁰.

Non è però qui chiarissimo cosa si debba intendere per «sintesi sufficientemente dettagliata», e l'intreccio con le tematiche legate alla protezione del diritto d'autore è evidentemente assai delicato: è probabile che su questo aspetto specifico la procedura di trilogia porti a un testo più dettagliato, anche se non necessariamente più soddisfacente.

Ma torniamo alla questione dei bias. Paradossalmente, la stessa tendenza a cercare di favorire la produzione di risposte 'politicamente corrette' e inclusive può essere percepita da alcuni come bias, come mostra l'acceso dibattito nato negli Stati Uniti sull'orientamento 'democratico' di ChatGPT e sulla sua avversione per le politiche repubblicane¹¹. Ed è sicuramente preoccupante la possibilità di costruire LLM programmaticamente orientati dal punto di vista politico o ideologico, o addestrati su testi e da parte di istruttori che desiderano utilizzarli per difendere tesi false o inaccettabili. In linea di principio la costruzione di LLM volutamente 'cattivi'¹² è non solo possibile ma anche relativamente facile, almeno finché ci si limita a modelli non troppo costosi dal punto di vista computazionale (e sappiamo che i costi computazionali tendono a diminuire nel tempo) o si lavora, come si è accennato sopra, sul contesto.

Quello dei bias è solo uno dei molti aspetti problematici su cui, pur senza demonizzare la ricerca nel campo delle IA generative e i suoi affascinanti risultati, è indispensabile riflettere. Sono così preoccupanti, per fare solo qualche esempio,

– la progressiva 'chiusura' sia, come ricordato sopra, dei corpora usati

nell'addestramento, sia delle metodologie, dei modelli e delle architetture usate¹³;

– le politiche (o l'assenza di politiche) di gestione dei dati personali forniti dagli utenti attraverso i loro prompt e il tema della possibilità di chiedere l'eliminazione di alcuni di tali dati da parte dei diretti interessati¹⁴;

– la questione – presente nel secondo fra i passi della proposta di regolamento europeo citati sopra – del copyright di contenuti generati sulla base di un addestramento su dati creati da persone che non sono necessariamente consapevoli dell'inclusione all'interno del corpus di testi o immagini da loro creati, e che non partecipano né alla definizione delle relative politiche né agli utili derivanti dal loro sfruttamento economico;

– il rischio che la concorrenza sregolata fra società che producono sistemi di IA generativa si traduca in una eccessiva accelerazione nel loro sviluppo senza il tempo necessario alla riflessione e all'elaborazione di adeguate garanzie etiche, così come di valutazioni del loro impatto non solo culturale ma anche politico, sociale, economico e occupazionale¹⁵.

Si tratta di problemi di enorme rilievo, che posso qui solo ricordare¹⁶: su molti di essi il dibattito è già avviato, ed è bene proceda con tutto l'approfondimento necessario.

¹ Una descrizione abbastanza dettagliata del sistema, con esempi di grande interesse, è in OpenAI, *GPT-4 Technical Report*, 2023, <https://cdn.openai.com/papers/gpt-4.pdf>.

² ChatGPT-4, risposta fornita all'autore l'8 aprile 2023.

³ Per un articolo divulgativo ma assai interessante sul tema si veda Siobhan Roberts, *A.I. Is Coming for Mathematics, Too*, in «The New York Times», 2 luglio 2023, <https://www.nytimes.com/2023/07/02/science/ai-mathematics-machine-learning.html>. Il progresso in questo campo sembra piuttosto rapido, e alcune delle criticità rilevate in Simon Frieder et al., *Mathematical Capabilities of ChatGPT*, in «arXiv», 31 gennaio 2023, <https://doi.org/10.48550/arXiv.2301.13867>, sembrano già superate o superabili attraverso l'adozione di strumenti procedurali di interrogazione e/o di modelli di addestramento PRM, di cui parlerò tra poco.

⁴ Hunter Lightman et al., *Let's Verify Step by Step*, in «arXiv», 31 maggio 2023, <https://doi.org/10.48550/arXiv.2305.20050>. Per una sintetica introduzione al

riguardo cfr. Ignacio de Gregorio, *OpenAI Just Showed Us How They Will Transform ChatGPT Into A Math Prodigy!*, in «Medium», 4 luglio 2023, <https://medium.com/@ignacio.de.gregorio.noblejas/prm-openai-f4e3f61dd72>.

⁵ Per ‘bias’ si intende una distorsione sistematica nei dati o nei modelli utilizzati, che porta il sistema a produrre risultati a loro volta distorti o inaffidabili. Abbiamo già ricordato la discussione di alcuni di questi casi in Emily Bender et al., *On the Dangers of Stochastic Parrots* cit. Si veda anche, ad es., Andrew Silva, Pradyumna Tambwekar e Matthew Gombolay, *Towards a Comprehensive Understanding and Accurate Evaluation of Societal Biases in Pre-Trained Transformers*, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, in «ACL Anthology» 2021, pp. 2383-2389, in rete all’indirizzo <https://aclanthology.org/2021.naacl-main.189/>.

⁶ Sui bias dei sistemi di intelligenza artificiale generativa che producono immagini si veda Melissa Heikkilä, *These new tools let you see for yourself how biased AI image models are*, in «MIT Technology Review», 22 marzo 2023, <https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/>. L’articolo è raggiungibile anche dal QR-Code inserito nel testo.

⁷ È il caso ad esempio di BookCorpus, su cui sono state addestrate le prime versioni sia di GPT sia di BERT (un corpus di circa 11.000 libri di vario genere diffusi gratuitamente sul sito Smashwords e scritti da autori non legati a case editrici commerciali), realizzato nel 2015 da ricercatori dell’Università di Toronto e del MIT e reso disponibile fino a pochi anni fa dal sito web dell’Università di Toronto (è ancora possibile reperirlo in rete, ma non più in forma ufficiale), o di CommonCrawl Corpus, il corpus aperto e gratuito realizzato dall’organizzazione no-profit CommonCrawl (<https://commoncrawl.org/>), che raccoglie il testo di milioni di pagine web e una cui versione filtrata (come è facile immaginare, il corpus CommonCrawl comprende anche contenuti pornografici o inaffidabili) è stata usata nell’addestramento di GPT-3. Un’interessante ricostruzione della storia di BookCorpus è in Jack Bandy e Nicholas Vincent, *Addressing “Documentation Debt” in Machine Learning: A Retrospective Datasheet for BookCorpus*, in Joaquin Vanschoren e Serena Yeung (a cura di), *Proceedings of the 2021 Conference on Neural Information Processing Systems, Track on Datasets and Benchmarks*, NeurIPS Datasets and Benchmarks 2021, <https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/file/54229abfcfa5649e7003b83dd4755294-Paper-round1.pdf>. È chiaro che entrambi questi corpora, tutt’altro che ‘asettici’, possono presentare e di fatto presentano bias sistematici: si veda al riguardo Abeba Birhane, Vinay Uday Prabhu e Emmanuel Kahembwe, *Multimodal datasets: misogyny, pornography, and malignant stereotypes*, preprint, in «arXiv», 5 ottobre 2021, <https://doi.org/10.48550/arXiv.2110.01963>.

⁸ Emendamenti del Parlamento europeo, approvati il 14 giugno 2023, alla proposta

di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)): https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_IT.html.

⁹ *Ibid.*

¹⁰ *Ibid.*

¹¹ Si veda ad esempio Fabio Motoki, Valdemar Pinho Neto e Victor Rodrigues, *More human than human: measuring ChatGPT political bias*, in «Public Choice», 17 agosto 2023, <https://doi.org/10.1007/s11127-023-01097-2>, e – per una sintesi delle critiche abbastanza radicali mosse al riguardo – Alberto Romero, *ChatGPT Would Vote Democrat, New Study Finds – But It's Full of Flaws*, in «The Algorithmic Bridge», 30 agosto 2023, <https://thealgorithmicbridge.substack.com/p/chatgpt-would-vote-democrat-new-study>. Per un curioso parallelo relativo alla politica italiana si veda Andrea Daniele Signorelli, *ChatGPT vota PD?*, in «Wired» online, 25 febbraio 2023, <https://www.wired.it/article/chatgpt-pd-idee-politiche-destra-sinistra-intelligenza-artificiale/>.

¹² Si veda al riguardo OpenAI, *GPT-4 Technical Report* cit.

¹³ Così, ad esempio, OpenAI, nata – come suggerisce il nome – con una politica non commerciale, si è ormai trasformata in una azienda assai poco 'aperta'; più fedele alle politiche 'open' è per il momento il LLM LLaMa, di cui si parlerà in seguito, e vi è molto lavoro anche sul tema della creazione di LLM totalmente aperti: cfr. Cameron R. Wolfe, *LLaMA: LLMs for Everyone! High-performing language models that are fully open-source*, in «Deep (Learning) Focus», 10 aprile 2023, <https://cameronrwolfe.substack.com/p/llama-llms-for-everyone>.

¹⁴ Nanna Bonde Thylstrup, *The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains*, in «Media, Culture & Society» 44 (4), 2022, pp. 655-671.

¹⁵ Su questi ultimi aspetti è abbastanza impressionante la lettura di report come quello prodotto nel marzo 2023 da Goldman Sachs: Jan Hatzius, Joseph Briggs, Devesh Kodnani e Giovanni Pierdomenico, *The Potentially Large Effects of Artificial Intelligence on Economic Growth*, Goldman Sachs Economic Research, 26 marzo 2023, <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>.

¹⁶ Il ricordato fascicolo (9, 1) della rivista «Biblioteche Oggi Trends» contiene articoli che affrontano alcuni di questi temi da una prospettiva abbastanza vicina a quella qui adottata: <https://www.bibliotecheoggi.it/it/fascicolo/2687/l-intelligenza-artificiale-per-le-biblioteche>. Aggiungo qui il riferimento anche a un recente testo introduttivo in italiano su ChatGPT, che affronta a sua volta alcune

di queste tematiche: Mafè de Baggis e Alberto Puliafito, *In principio era ChatGPT. Intelligenze artificiali per testi, immagini, video e quel che verrà*, Apogeo, Milano 2023.

12.

Noi e loro: le IA generative ci assomigliano?

Ho già osservato che le IA generative ‘costruiscono’ risposte anziché limitarsi a selezionare e riproporre contenuti già esistenti in rete. Questo ha vantaggi e svantaggi: abbiamo appena ricordato alcuni dei potenziali svantaggi e problemi, ma va detto che la loro capacità di produzione di testo sintatticamente, semanticamente e contestualmente adeguato è realmente impressionante, e ci sono pochi dubbi sul fatto che sistemi come quelli qui considerati saranno in grado abbastanza presto di superare il test di Turing, almeno nella versione proposta nel famoso articolo del 1950 (probabilmente, possono già ingannare un esaminatore privo di competenze specifiche di alto livello).

Finché le allucinazioni resteranno un problema frequente, difficilmente vorremmo affidare a sistemi di questo tipo funzioni complesse di mediazione informativa, o la capacità di far seguire alla produzione linguistica anche azioni autonome capaci di avere effetti sul mondo reale. Così *in questo momento* GPT, ChatGPT o i sistemi in parte analoghi in sviluppo presso altre aziende (come Bard, il sistema su cui lavora Google a partire da un LLM chiamato LaMDA, o il LLM di Meta, chiamato LLaMA) non sarebbero adatti a svolgere funzioni come quelle del reference bibliotecario (cosa pensereste di un bibliotecario che suggerisce come bibliografia articoli inesistenti?), difficilmente potrebbero produrre ricerca innovativa in ambito umanistico (mentre in ambito scientifico e tecnologico alcune IA sono già utilizzate in settori specifici, come la genomica, lo sviluppo di nuove proteine, o come supporto alla programmazione di software), e non affideremmo loro il nostro numero di carta di credito. Ma non è affatto detto che queste limitazioni dureranno nel tempo: chi ritiene che alle IA generative manchi la

‘creatività’ umana tende forse da un lato a dare troppa importanza alla metafora del ‘pappagallo’ (di cui ho cercato di mostrare il carattere fortemente fuorviante), dall’altro a sopravvalutare la dose di creatività richiesta da molti compiti anche altamente astratti, comprese molte forme di ricerca accademica o di mediazione informativa. E probabilmente avrebbe difficoltà a descrivere il concetto di creatività in forme tali da portare ad escludere in linea di principio che intelligenze artificiali generative possano produrre risultati simili.

Un discorso analogo si può fare per quanto riguarda la semantica. Ho già notato come uno sguardo appena più ravvicinato a questi sistemi porti a smentire l’idea diffusa di una macchina che lavori prescindendo dal piano dei significati: certo GPT o ChatGPT non ‘capiscono’ i testi nello stesso senso in cui li capiamo noi, non hanno coscienza o autocoscienza (ammesso e non concesso che sia chiaro cosa sia e come funzioni la nostra autocoscienza), non hanno intenzionalità (anche qui, ammesso e non concesso che sia chiara la natura dell’intenzionalità umana), e sembra difficile – almeno per ora – attribuire loro un ruolo di agenti autonomi¹. Ma si tratta di motori semantici, non solo sintattici, e la loro capacità di produrre significati è radicata nella costruzione di modelli linguistici che non hanno solo un interesse pratico, ma anche un notevole interesse teorico. Modelli che in linea di principio potrebbero aiutarci a capire anche alcuni dei meccanismi di funzionamento della nostra produzione linguistica, peraltro in molti casi ancora abbastanza oscuri. Siamo così sicuri che modelli di generazione statistica non abbiano un ruolo anche nella *nostra* competenza linguistica?

Può essere il caso di ricordare, a questo proposito, che i migliori algoritmi di attribuzione dei testi, quelli utilizzati quando dobbiamo cercare di attribuire un testo anonimo o ‘sospetto’ a un autore o all’altro, non utilizzano ormai quasi più le componenti semanticamente e stilisticamente significative della frase che potremmo ragionevolmente aspettarci come criteri distintivi per la produzione linguistica umana (come lessico, lunghezza media delle frasi, uso della punteggiatura, ecc.), ma un dato che sembrerebbe puramente sintattico: la frequenza di n-grammi, e in particolare di bigrammi e trigrammi, risultato della scomposizione del testo in gruppi di n caratteri². In altri termini, a

distinguere la produzione testuale di due persone sembrano essere più considerazioni statistico-probabilistiche sulla frequenza di sequenze assai brevi di caratteri, raramente dotate di significato autonomo, che aspetti stilistici o lessicali di più alto livello. Un dato che, guardando ai modelli di produzione di testo usati dalle intelligenze artificiali generative, non può non far pensare. Incidentalmente, ricordiamo anche che i programmi di attribuzione dei testi rappresentano un caso particolare di programmi di discriminazione (l'attribuzione del testo a questo o quell'autore all'interno di un insieme di autori possibili): si tratta dunque di un compito per il quale possono essere utilizzati efficacemente anche sistemi basati su reti neurali.

E se qualche tipo di meccanismo statistico-probabilistico avesse un ruolo – pur se non esclusivo – anche nella nostra produzione linguistica, siamo così sicuri che la nostra coscienza e la nostra autocoscienza non siano a loro volta almeno in parte collegate a un sostrato di questo tipo? O, ancora, siamo sicuri che l'intelligenza richieda necessariamente la coscienza, o una forma di coscienza analoga alla nostra?

Il problema del rapporto fra cervello e coscienza (su cui non intendo certo soffermarmi in questa sede: il tema è alla base di una letteratura sconfinata e di posizioni anche molto diverse³) è una questione – un tempo solo filosofica, oggi anche neurofisiologica – aperta da secoli ma ben lontana dall'essere risolta: se non accettiamo forme di dualismo o spiritualismo, sembra assai difficile non considerare la coscienza come un fenomeno emergente a partire da un sostrato biologico. Ma come funzioni concretamente questo sostrato, come sia capace di processare e di produrre informazioni, come intrecci dati sensoriali, dati linguistici, significati, emozioni, intenzioni, è cosa che, nonostante gli indubbi progressi fatti dalle neuroscienze negli ultimi decenni, continua ad essere piuttosto oscura.

Sappiamo che i sistemi di intelligenza artificiale generativa sono abbastanza diversi da noi – non solo nel loro funzionamento, ma anche nell'assenza di connessione diretta con un corpo biologico – da suggerire che, almeno al momento, attribuire loro coscienza, o intenzioni, o una intelligenza analoga alla nostra, rappresenti una antropomorfizzazione

non solo ingiustificata ma anche assai fuorviante, che è sicuramente necessario evitare.

Sappiamo però anche che la loro capacità di produrre contenuti linguistici sintatticamente e semanticamente appropriati – una capacità tradizionalmente considerata appannaggio dell'intelligenza umana – li trasforma in agenti linguistici almeno parzialmente autonomi. Che rapporto c'è, allora, fra 'agency' (il poter funzionare come agente autonomo) e intelligenza nel caso di sistemi di questo tipo? Il tema è di particolare rilievo anche perché l'associazione di agency e intelligenza è in genere considerata come base anche per l'attribuzione di responsabilità etiche e morali, e l'insieme di queste caratteristiche è importante nel nostro riconoscimento dello status di 'persona'. Se attribuissimo ai sistemi di intelligenza artificiale generativa sia intelligenza sia agency, dovremmo inevitabilmente chiederci se dobbiamo attribuire loro anche responsabilità etiche e morali, e se dobbiamo riconoscerli come 'persone', almeno in un qualche senso del termine.

Luciano Floridi ha recentemente sostenuto che un sistema come ChatGPT sembra rappresentare un caso di agency senza intelligenza⁴. Floridi difende la sua tesi con buoni argomenti, e ChatGPT comunque non è una AGI, una intelligenza artificiale generale; ma sospetto che si potrebbero trovare argomenti altrettanto buoni per difendere la possibilità – se si ritenesse che lo sviluppo della ricerca in questo campo possa effettivamente portare alla creazione di intelligenze artificiali generali – di una intelligenza senza agency. E il possibile 'decoupling' di agency e intelligenza – o la moltiplicazione dei possibili sensi da attribuire ai due termini – imposto dagli sviluppi nel campo dell'intelligenza artificiale rappresenta un tema di grande interesse anche dal punto di vista filosofico, soprattutto se, come si accennava, all'equazione si aggiungono le ulteriori tematiche della responsabilità e dello status da attribuire a questi sistemi.

La situazione della ricerca nel campo delle IA generative, insomma, impone da un lato enorme cautela interpretativa, dall'altro molta attenzione, e non permette di escludere del tutto, anche per il futuro, prospettive per ora relegate al campo della fantascienza.

¹ Cfr. Maurizio Lana, *L'agency dei sistemi di intelligenza artificiale. Un punto di vista bibliografico*, in «DigitCult» 7 (1), 2022, <https://doi.org/10.36158/97888929552576>.

² Una metodologia presentata per la prima volta in forma sistematica in Fuchun Peng, Dale Schuurmans, Vlado Keselj e Shaojun Wang, *Language Independent Authorship Attribution with Character Level N-Grams*, in *10th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Budapest 2003, <https://aclanthology.org/E03-1053/>.

³ Per una introduzione recente si veda Anil Seth, *Being You: A New Science of Consciousness*, Faber & Faber, London 2022 (trad. it. *Come il cervello crea la nostra coscienza*, Raffaello Cortina, Milano 2023).

⁴ Luciano Floridi, *AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models*, 14 febbraio 2023, preprint disponibile su SSRN: <http://dx.doi.org/10.2139/ssrn.4358789>.

13.

Nove tesi (e uno sguardo al mondo del libro)

In quest'ultimo capitolo della parte seconda del libro vorrei da un lato provare a tirare le somme del discorso fatto fin qui a proposito delle intelligenze artificiali generative, sintetizzando alcune delle conclusioni raggiunte, dall'altro proporre qualche riflessione specifica relativamente al loro possibile impatto sul mondo del libro e dell'editoria: un mondo che mi interessa particolarmente sia dal punto di vista professionale sia, nel contesto di questo libro, per i suoi evidenti e fortissimi legami con il modello architettonico di organizzazione delle conoscenze di cui si è parlato nella parte prima. Quali conseguenze può avere in tale ambito la rapidissima diffusione delle intelligenze artificiali generative?

Rispondere a questa domanda non è facile, anche perché le possibili risposte tendono a cambiare molto velocemente. Proverò a fornire qualche indicazione, assolutamente parziale, basandomi anche su un confronto sul tema svoltosi online nel giugno 2023 per iniziativa del Forum del libro, attraverso una delle sue liste di discussione. In quell'occasione, mi era stato affidato il compito di stimolare la discussione attraverso un documento articolato in tesi, che ho pensato di riprendere come base anche per questo capitolo. Si tratta di quattro tesi più ampie, che riguardano lo sviluppo dell'intelligenza artificiale generativa e che possono avere appunto la funzione di sintetizzare alcune conclusioni su temi già discussi nei capitoli precedenti, e di quattro tesi più specifiche, che riguardano in particolare l'ambito editoriale. Ho poi aggiunto, in chiusura, una nona tesi che riguarda il possibile uso di sistemi di intelligenza artificiale generativa per produrre automaticamente metadati relativi a grandi quantità di dati: un tema di grande rilievo rispetto alla

prospettiva assunta in questo libro, perché rappresenterebbe una forma di collaborazione dell'oracolo al lavoro dell'architetto.

Le tesi sono espresse in forma piuttosto 'forte', una scelta che era stata fatta soprattutto per stimolare la discussione, e che per lo stesso motivo ho preferito conservare anche in questa sede. Sono abbastanza convinto che siano fondamentalmente corrette, ma sono tutte suscettibili di critica e possono essere legittimamente contestate, anche con buoni argomenti.

Tesi generali

T1 – Di ChatGPT e di intelligenza artificiale generativa si parla ovunque, ma per lo più in maniera assai approssimativa e spesso errata. Questo dipende anche e soprattutto dal fatto che, mentre è assai facile usare e dunque mettere alla prova questi sistemi lato utente, è molto più difficile capire i meccanismi effettivi del loro funzionamento. E senza un minimo di conoscenze e competenze su tali meccanismi (senza un'idea di cosa sia una rete neurale, di come funzioni, di cosa sia un transformer...) si dicono facilmente stupidaggini, o si è portati a trarre conclusioni potenzialmente errate. Anche la trattazione che è stata fatta fin qui rappresenta una presentazione assai parziale e limitata di un tema assai complesso.

T2 – È diffusa l'idea che le intelligenze artificiali generative abbiano, non solo nel loro sviluppo attuale ma anche in linea di principio, forti limiti di originalità, dato che si limiterebbero a riproporre contenuti già presenti nel corpus sul quale sono state addestrate. In altri termini, che siano incapaci di produrre contenuti davvero originali. Questa conclusione è infondata: è vero che il corpus di addestramento è limitato, ed è vero che – come abbiamo visto nel capitolo precedente – i contenuti generati sono fortemente influenzati da caratteristiche e tipologia dei materiali che tale corpus comprende, ma le IA generative non sono affatto pappagalli e non usano il corpus come una raccolta di informazioni da cui 'copiare': l'addestramento avviene infatti scomponendo il corpus in unità di livello talmente basso (i 'token' di cui si è già parlato) da rendere poco sensata l'interpretazione del lavoro di generazione di nuovi contenuti in termine di 'copiatura' o di riproposizione quasi meccanica di informazioni già presenti nel corpus. Di fatto, IA generative addestrate su corpora sufficientemente ampi e basate su reti neurali sufficientemente potenti (in

termini di estensione della rete, dei suoi meccanismi di attenzione e dei parametri utilizzati) possono produrre contenuti altamente originali.

T3 – È assai diffusa anche l'idea che le intelligenze artificiali generative funzionino sulla base di principi totalmente diversi rispetto all'intelligenza umana. Questo è almeno in parte vero: sappiamo ad esempio che i neuroni del nostro cervello comunicano sia attraverso mediatori chimici sia attraverso mediatori elettrici, sappiamo che il cervello umano è capace di produrre contenuti altamente originali pur essendo esposto a una base di stimoli strettamente linguistici molto più limitata dei corpora usati nell'addestramento delle IA generative attuali, e d'altro canto sappiamo che le tipologie di stimoli esterni a cui è sottoposto un cervello umano sono molto più ampie di quelle al momento utilizzate dai sistemi artificiali. Tuttavia, la realtà è che, nonostante gli enormi progressi fatti dalle neuroscienze negli ultimi decenni, sappiamo ancora troppo poco sul funzionamento del nostro cervello e sul rapporto mente-cervello, per poter concludere con certezza che la nostra intelligenza non sia replicabile – e magari superabile – utilizzando reti neurali artificiali. Come si è accennato nell'ultimo capitolo, non è affatto escluso, ed anzi è abbastanza probabile, che meccanismi statistico-probabilistici non troppo lontani da quelli usati dalle IA generative siano presenti anche nel funzionamento del nostro cervello, pur se quasi certamente accanto a meccanismi che funzionano in maniera del tutto diversa. E in linea di principio niente impedisce l'“embodiment” di una intelligenza artificiale, cioè la possibilità di collegarla a un supporto fisico capace di movimento e di raccolta e integrazione di stimoli non solo linguistico-testuali ma anche sonori, visivi, tattili, olfattivi... (volendo, su uno spettro di frequenze e con capacità di discriminazione superiori alle nostre). Nel momento in cui scrivo si è appena concluso a Vancouver, in Canada, il quarto *Embodied AI Workshop*, e per farsi un'idea sui possibili sviluppi futuri in questo campo vale sicuramente la pena di dare un'occhiata ai suoi materiali¹.



QR-Code 16

Il sito di riferimento dell'*Embodied AI Workshop*

<http://bit.ly/3XHWRat>

T4 – Esiste una discussione assai ampia (e spesso molto impressionistica) sulla possibilità o meno di raggiungere il livello della cosiddetta ‘singolarità’: il momento in cui sistemi intelligenti artificiali diventeranno altrettanto o più potenti dell’intelligenza umana, magari sviluppando anche forme di autocoscienza. In base alle informazioni di cui disponiamo attualmente, non mi sembra possibile argomentare in maniera davvero convincente né a favore né contro questa ipotesi. Mi sembra però molto probabile che sistemi di intelligenza artificiale siano molto vicini o abbastanza vicini a raggiungere quelle che potremmo chiamare ‘singolarità locali’, cioè a superare le capacità umane in campi specifici, anche di notevole rilievo. Questo è già accaduto, ad esempio, nel gioco degli scacchi o del go. Credo potrà succedere presto anche in settori di peso assai maggiore, dalla diagnostica medica alla formulazione di ipotesi e teorie (e alla loro messa alla prova) relativamente ad ambiti specifici in campi come la fisica, la chimica, la biologia, la cosmologia. Giustificate preoccupazioni sono legate all’uso di intelligenze artificiali in campo militare. Alcuni settori in cui l’intelligenza artificiale potrà superare quella umana sono presenti anche nelle scienze sociali (ad esempio nell’analisi e nella predizione di dati e comportamenti in campo economico o finanziario, dove sono già molto usate, e in prospettiva anche geopolitico) e nelle stesse discipline umanistiche. Una linea di ricerca in cui questo è in linea di principio già possibile è – come si è già accennato – l’attribuzione di testi, mentre altri esempi vicini alla ‘singolarità locale’ sono la ricostruzione e la traduzione di testi del passato o il lavoro di collegamento fra testimonianze archeologiche (ad esempio la

ricostruzione di un oggetto a partire dai suoi frammenti), o l'analisi di alcuni meccanismi di diffusione ed evoluzione delle lingue. Vale la pena accennare, al riguardo, che strumenti molto simili all'embedding si possono usare in campo cartografico, per la creazione di rappresentazioni di fenomeni complessi attraverso mappe multidimensionali costruite dall'IA a partire da dati: lo si può fare in campo linguistico, ma anche nella geografia sociale o economica. In questi campi, un limite della cartografia tradizionale è sempre stato costituito dalla difficoltà di rappresentare confini sfumati e variazioni di gradiente, in particolare quando la carta vuole prendere in considerazione contemporaneamente più dimensioni dei dati. Sistemi di intelligenza artificiale basati su reti neurali, associati a sistemi di rappresentazione di dati geografici (GIS), potrebbero portare in questo campo a un vero e proprio cambiamento di paradigma.

Tesi specifiche

T5 – Il campo editoriale è uno di quelli in cui le conseguenze dello sviluppo delle IA generative si faranno sentire di più, fin dai prossimi anni. Come si è accennato nel primo capitolo, possiamo forse considerarle come una quinta fase nello sviluppo dell'ecosistema digitale, dopo 1. lo sviluppo dei computer come macchine da calcolo (1950-1980); 2. la nascita dei personal computer come strumenti multifunzionali, capaci di lavorare anche su testi, immagini, suoni, video... (1980-1995); 3. la diffusione generalizzata di Internet e del web (1995-2010); 4. la diffusione degli smartphone e del collegamento veloce, permanente e mobile alla rete (2010-2020). Oppure, usando una diversa periodizzazione, come una quinta fase dopo 1. l'età dei cacciatori-raccoglitori che sperimentavano i primi collegamenti alla rete; 2. l'età dei primi insediamenti online e dell'agricoltura informativa, con la progressiva affermazione del web; 3. l'età dell'artigianato e del commercio, dei social network, del contenuto generato dagli utenti e 4. l'età – almeno parzialmente sovrapposta a quella precedente – delle cattedrali informative basate su modelli architettonici 'forti' del sapere e sull'idea di linked data. In tutti e due i modelli, le fasi 2-4 hanno già avuto

effetti dirompenti sui processi editoriali. Lo stesso accadrà – per certi versi in forma ancor più accentuata – nel caso delle IA generative.

T6 – In parte come conseguenza della tesi T2, le IA generative saranno presto in grado di tradurre contenuti fra lingue diverse con una competenza non inferiore a quella di un discreto traduttore umano, non solo in campo saggistico ma, almeno limitatamente ai testi meno impegnativi, anche in campo letterario. Una prospettiva certo assai preoccupante per traduttrici e traduttori – che hanno ragione nel rivendicare *oggi* la maggior qualità di una traduzione umana rispetto a quella ‘good enough’ dei sistemi automatici – ma difficilmente evitabile nel medio e lungo periodo. Con effetti complessivi potenziali che in prospettiva possono essere anche positivi, per il maggiore accesso a culture diverse, fonti informative più differenziate e contesti di produzione testuale anche lontani. Resterà possibile salvaguardare, almeno per un certo periodo, l’ambito delle traduzioni letterarie di alto livello, e resteranno – anche qui, almeno per un certo periodo – forme di collaborazione fra le competenze di traduttrici e traduttori umani e il lavoro ‘di base’ dei sistemi automatici. Tradizionalmente l’attività editoriale si è concentrata sulla produzione di ‘oggetti’ informativi che integravano interfaccia di lettura (ad esempio le pagine di un libro) e manifestazione specifica dell’opera (il testo stampato, nella particolare lingua in cui è scritto o tradotto). L’editoria digitale ha mostrato la possibilità di separare interfaccia di fruizione e manifestazione dell’opera: un e-book reader è indipendente dal testo che vi è contenuto. Nel campo dell’editoria libraria, questa possibilità è stata usata solo parzialmente a livello di esperienza utente – la maggior parte dei lettori preferisce ancora leggere su carta – ma è comunque alla base del processo attuale di produzione editoriale di un libro, stampato a partire da un testo digitale. In prospettiva, la possibilità di tradurre automaticamente o semi-automaticamente e con buona qualità i contenuti in qualunque lingua sufficientemente diffusa e attestata tenderà a focalizzare l’attività editoriale sull’opera più che sulle sue manifestazioni specifiche, comprese quelle relative a singole lingue, e contribuirà ulteriormente all’internazionalizzazione del mercato editoriale e alla diffusione delle interfacce di lettura digitali.

T7 – Le IA generative interverranno anche sulla possibilità di produrre

sintesi dei testi, di migliorarne alcuni aspetti di qualità anche redazionale² o di identificare e richiamare contenuti integrativi, ad esempio permettendo di recuperare il modello della ‘piramide’ di Darnton (l’idea di una stratificazione del testo in livelli diversi per complessità e tipologia dei contenuti, con al vertice una sintesi introduttiva e alla base fonti e approfondimenti) – finora poco usato per l’alto costo e impegno richiesti dalla produzione di contenuti di questo tipo – e, almeno in parte, di automatizzarlo³. Anche questo spingerà in direzione delle interfacce di lettura digitali. Altri campi rivoluzionati dalle IA generative (in molti casi ancor prima delle applicazioni sopra ipotizzate) saranno quelli della grafica editoriale, dell’impaginazione, della produzione di contenuti promozionali (inclusi quelli su social network), della produzione automatica o semi-automatica di audiolibri (anche in questo caso, in più lingue diverse), dell’accessibilità. Ricordo che l’immagine di copertina del libro che avete in mano è stata generata da Midjourney, al momento in cui scrivo probabilmente il migliore sistema di intelligenza artificiale per la generazione di immagini, anche se la qualità della copertina dipende molto anche da come l’immagine di partenza è stata poi utilizzata dalla responsabile grafica della casa editrice.

T8 – Gli sviluppi sopra ipotizzati avranno certo effetti anche occupazionali, ma – almeno nel breve termine – non necessariamente nel senso di una perdita generalizzata di lavoro umano: in molti fra i campi sopra ricordati (anche se non in tutti) è infatti ipotizzabile una diffusione di modelli di collaborazione fra professionalità umane e IA, per indirizzare e finalizzare in maniera funzionale il lavoro delle IA generative e garantirne la qualità. Ovviamente, la situazione di medio e lungo periodo è assai più difficile da prevedere e vi sarà comunque la necessità di cambiamenti abbastanza radicali nel tipo di conoscenze e competenze necessarie per svolgere molti dei lavori presenti nella filiera editoriale. Si può ipotizzare che la proporzione fra lavori persi e nuovi lavori come conseguenza della diffusione delle IA suggerita da un recente report del World Economic Forum⁴ (83 milioni di posti di lavoro persi *vs* 69 milioni di nuovi posti di lavoro nei prossimi cinque anni) possa corrispondere a grandi linee anche a quanto potrebbe accadere nel settore editoriale. A livello più generale, bisognerà vedere se e quanto saremo capaci di

indirizzare l'impatto dell'intelligenza artificiale verso la riduzione dell'orario di lavoro anziché verso la riduzione dei posti di lavoro.

L'oracolo può collaborare con l'architetto?

T9 – Fra gli ambiti in cui le intelligenze artificiali generative saranno più usate in futuro ci sarà anche la produzione automatica di metadati – in particolare metadati relativi a basi di dati tanto ampie da poter essere difficilmente dominate con metodi tradizionali – e la costruzione di basi dati più organizzate, meglio strutturate e meglio descritte a partire da raccolte di dati meno organizzate e strutturate. Uno dei campi in cui questo potrà avvenire è quello delle memorie digitali personali, di cui mi occuperò nella prossima parte. Le intelligenze artificiali generative potranno inoltre lavorare alla produzione automatica di metadati in contesti come gli archivi, le biblioteche⁵, i musei, la descrizione di contenuti editoriali.

¹ Il sito di riferimento è <https://embodied-ai.org/>, raggiungibile anche dal QR-Code disponibile inserito nel testo.

² Cfr. Ethan Mollick, *What happens when AI reads a book*, nel blog «One Useful Thing», 26 maggio 2023, <https://www.oneusefulthing.org/p/what-happens-when-ai-reads-a-book>.

³ Ho discusso il modello piramidale di Darnton in Gino Roncaglia, *La quarta rivoluzione. Sei lezioni sul futuro del libro*, Laterza, Roma-Bari 2010, pp. 223-224.

⁴ World Economic Forum, *Future of Jobs Report 2023 – Insight Report*, maggio 2023, https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf.

⁵ L'ipotesi è prospettata in Brady D. Lund e Ting Wang, *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?*, in «Library Hi Tech News» 40.4, 14 febbraio 2023, <https://www.emerald.com/insight/content/doi/10.1108/LHTN-01-2023-0009/full/html>.

Parte III

14.

Memorie personali e mente estesa

Il computer è diventato, già dall'epoca dei personal computer e in maniera progressivamente più pervasiva ed evidente con la diffusione di Internet e dei dispositivi mobili, uno straordinario strumento di esternalizzazione della nostra memoria, personale e collettiva. Le memorie informatiche non sono solo una componente necessaria al funzionamento di una macchina: sono un nuovo supporto (di fatto, *il* nuovo supporto) attraverso cui conserviamo tracce di conoscenze, avvenimenti, attività, persone, interazioni, impressioni... La rivoluzione digitale è stata anche una rivoluzione della memoria, e la memoria è parte essenziale della nostra identità, personale e collettiva¹.

Dal punto di vista di questo libro, le memorie digitali sono anche un banco di prova di grande interesse sul modo in cui organizziamo le nostre conoscenze e i nostri ricordi. Tutti gli strumenti di cui abbiamo parlato nelle prime due parti del libro – dalle enciclopedie come modello di organizzazione architettonica del sapere alle intelligenze artificiali generative come modello di produzione di nuovi contenuti sulla base di modelli statistico-probabilistici estremamente complessi e in parte opachi – fanno in ultima analisi affidamento sulle memorie informatiche come proprio supporto. E tutti sono, in qualche modo, connessi con la dimensione della memoria personale e collettiva: le enciclopedie ne distillano e sistematizzano le componenti che consideriamo fondamentali, offrendone un quadro di riferimento altamente strutturato; le intelligenze artificiali generative apprendono basandosi su corpora per molta parte meno sistematici e organizzati ma assai più vasti, che sono a loro volta depositi delle nostre tracce e testimonianze.

In un certo senso, possiamo affermare che tanto l'architetto quanto

l'oracolo operino sulla base di archivi. L'architetto considera l'archivio come una struttura organizzata e ordinata o da ordinare sulla base di principi ben saldi, a partire dalla necessità di riflettere l'articolazione organizzativa e istituzionale che li ha prodotti. L'oracolo non ha necessariamente bisogno di archivi strutturati e ordinati – anche se in realtà non li disdegna affatto – ma guarda alla documentazione conservata come a una fonte da assorbire, assimilare e riutilizzare in nuove forme: lascia volentieri all'architetto la funzione di conservazione, di descrizione (e dunque metadattazione) e di ricerca puntuale del singolo documento, e si propone invece come produttore di contenuti nuovi a partire dalla rivisitazione 'creativa' di quelli del passato.

Ma torniamo al tema dell'organizzazione della memoria; o meglio, delle sue molte e diverse tipologie, che nel caso delle memorie personali offrono un quadro ancor più variegato di quanto già non accada per la documentazione prodotta da enti e istituzioni. In che modo il passaggio dalla carta – primo e fondamentale strumento di esternalizzazione della memoria – al digitale ne ha cambiato natura e organizzazione? Che rapporto hanno con i due modelli che abbiamo discusso finora, e per i quali ho proposto le due metafore dell'architetto e dell'oracolo? Quanta parte della nostra capacità di recuperare ricordi e materiali del passato – in forma di testi, suoni, immagini, video... – è oggi guidata dall'architetto, e quanta parte è guidata dall'oracolo (che magari ci ripropone in maniera apparentemente casuale, sullo schermo dello smartphone, una fotografia scattata anni fa)?

In questo e nei prossimi capitoli proverò ad affrontare queste tematiche dal punto di vista delle memorie personali, che costituiscono il banco di prova per molti versi più affascinante della nostra capacità di usare il nuovo ecosistema digitale per conservare tracce del passato. Lo farò riprendendo in parte tematiche che ho affrontato nel corso di un convegno organizzato nel 2021 da Stefano Allegrezza sul tema *Personal Digital Memories*²: chi fosse curioso di confrontare l'esposizione fatta in quella sede con quella che propongo qui o volesse approfondire il tema delle memorie digitali personali anche attraverso le altre relazioni al convegno, tutte assai interessanti, può trovare su YouTube la registrazione video di tutte le sessioni (svolte su Zoom, dato che si era in periodo di

emergenza Covid). Il QR-Code qui di seguito porta al video della prima sessione, nella quale si è svolto il mio intervento e da cui sono facilmente raggiungibili anche i video delle sessioni successive.



QR-Code 17

Convegno *Personal Digital Memories* (2021), prima sessione

<http://bit.ly/3roIJGZ>

Naturalmente, la capacità (e la scelta) di esternalizzare una parte della nostra memoria e di affidarla a supporti diversi e indipendenti rispetto al cervello, che ne costituisce la naturale base biologica, è ben più antica: l'uso di marcare attraverso tacche su legno o su altri materiali il passare del tempo, o di lasciare volontariamente tracce in occasione di eventi di particolare rilievo (si pensi alle prime rappresentazioni di episodi di caccia), precede – e di molto – la stessa evoluzione della scrittura. I sostenitori della teoria della mente estesa, nelle sue varie articolazioni e sfumature, vedono in questo processo di progressiva esternalizzazione di dati e conoscenze un cambiamento che investe sia la costruzione della nostra identità personale, sia le modalità del nostro rapporto – conoscitivo e operativo – con il mondo esterno. In questa prospettiva, l'‘io’ non ha più sede solo nel cervello biologico ma anche nell'insieme man mano più ampio e complesso di strumenti che realizziamo per estenderne le capacità: strumenti tecnologici ma anche strumenti culturali, a partire dai molti linguaggi – compresi quelli informatici – usati per parlare di noi e del mondo, anche attraverso la creazione di modelli utili a rappresentare la realtà e interagire con essa.

Proposta nel 1998 da Andy Clark e David Chalmers in un articolo divenuto rapidamente fra i più discussi e citati in ambito filosofico³, la teoria della mente estesa sottolinea come la nostra mente utilizzi per i suoi

processi cognitivi non solo il cervello biologico ma molte sue ‘estensioni’ esterne, naturali e sempre più spesso anche artificiali o culturali. La teoria ovviamente non riguarda solo la memoria, e propone una visione complessiva dei processi cognitivi che i due autori hanno battezzato ‘esternalismo attivo’, non priva di difficoltà filosofiche specifiche: se consideriamo alcuni aspetti del mondo esterno come parte dei nostri processi cognitivi, come decidiamo *quali* parti del mondo esterno abbiano tale funzione? Il rischio, abbastanza ovvio, è di minare ogni possibile distinzione fra io e mondo e trasformare la nostra mente in una sorta di soggetto indistinto o addirittura universale. D’altro canto, anche la negazione dell’esternalismo ha i suoi problemi: attività conoscitive fondamentali come la scrittura, la geometria, la matematica avanzata, sembrano richiedere necessariamente supporti e riferimenti esterni e la loro padronanza non sembra possibile se consideriamo la mente come un’entità autonoma e indipendente dal mondo che la circonda⁴.

Mi sembra chiaro che per il dibattito sulla mente estesa, qui necessariamente appena accennato⁵, il tema delle memorie personali digitali abbia una notevole rilevanza (e viceversa). Abbiamo ad esempio dati sperimentali sul fatto che fotografare oggetti abbia un effetto sulla nostra capacità di memorizzarli: se fotografiamo un quadro visitando un museo, il fatto di aver usato la macchina fotografica può indebolire la nostra memoria del quadro, dei suoi dettagli e della sua posizione nella sala, rispetto a chi ha visitato il museo osservando il quadro ma senza fare fotografie. Questo, però, solo fin quando non torniamo alla foto scattata: a quel punto non solo ricordiamo meglio il quadro, ma anche l’insieme della visita al museo; inoltre, «abbiamo scoperto che l’attività di rivedere foto di questo tipo ha effetti benefici sulla memoria in pazienti con amnesia o altri disturbi gravi della memoria»⁶ (un dato confermato anche dall’esempio dei memofilm, che sarà discusso in seguito). Come è facile capire da questi e molti altri esempi simili, l’intreccio fra memorie ‘interne’ e memorie ‘esterne’ è talmente stretto da rendere davvero difficile, alla luce del ruolo centrale svolto dalla memoria nella costruzione dell’identità personale⁷, una concezione della mente che non includa in qualche modo il ruolo svolto dalle memorie ‘fuori di noi’. Il tema delle

memorie digitali personali ha dunque un evidente interesse specifico anche in ambito cognitivo⁸.

Da questo punto di vista, l'esplosione quantitativa e qualitativa (anche in termini di moltiplicazione delle tipologie) delle memorie prodotte nell'ambito del nuovo ecosistema digitale potrebbe essere dunque interpretata, in un senso piuttosto forte, come una modificazione nel funzionamento stesso della nostra mente.

È una prospettiva che ha oggi notevoli implicazioni anche rispetto al lavoro di costruzione di intelligenze artificiali generative: basti pensare che sono stati già proposti sistemi addestrati in tutto o in parte su corpora personali e attraverso i quali si vorrebbe in qualche misura costruire un 'gemello digitale', capace di interazione linguistica, del produttore originario. Si tratta di temi che affronterò brevemente in seguito ma che – indipendentemente dall'adesione o meno a una qualche formulazione dell'idea di mente estesa – permettono subito di comprendere l'enorme portata dei cambiamenti introdotti negli ultimi decenni. Sono infatti mutate radicalmente sia la quantità, sia la qualità, sia le modalità di produzione, conservazione, organizzazione, riuso delle nostre memorie e delle 'tracce' che – intenzionalmente o no, in forma più o meno esplicitamente documentale – costituiscono non solo residuo e testimonianza delle interazioni fisiche e sociali fra di noi e con il mondo che abbiamo intorno, ma anche parte della nostra identità personale.

Proverò a parlarne distinguendo quattro diverse forme di incontro fra memorie personali e mondo digitale, che corrispondono almeno in parte – è un'altra fra le molte possibili forme di periodizzazione di cui abbiamo parlato nel primo capitolo – a quattro momenti nello sviluppo del nostro rapporto, via via più profondo, con il nuovo ecosistema informativo: 1) la digitalizzazione di memorie personali tradizionali (diari, lettere, fotografie, registrazioni audio e video...); 2) la produzione di memorie personali nativamente digitali che riprendono il modello di quelle tradizionali (in altri termini, il trasferimento all'interno dell'ecosistema digitale di processi 'tradizionali' di produzione di contenuti); 3) la nascita di nuove tipologie di memorie personali (ad esempio, pur essendo spesso uno sviluppo di oggetti documentali già presenti in precedenza, si può credo ragionevolmente ritenere che messaggistica, post su social network,

voci di calendario, playlist musicali, dati di geolocalizzazione e simili siano, nella loro incarnazione digitale, oggetti documentali di tipo nuovo); 4) la trasformazione di molte di queste (e di altre) tipologie di memorie e tracce personali in big data, la cui stessa analisi richiede necessariamente strumenti informatici e, probabilmente, strumenti concettuali in parte ancora da sviluppare, rispetto ai quali si prefigura forse, come si è già accennato in chiusura della parte precedente, una qualche forma di collaborazione fra l'architetto e l'oracolo.

Come per tutte le altre forme di periodizzazione già incontrate nelle prime due parti del libro, vale naturalmente anche in questo caso l'avvertenza che si tratta di distinzioni in qualche misura arbitrarie all'interno di un continuum in cui l'attribuzione di questa o quella caratteristica a una certa tipologia di contenuti è spesso più questione di gradi e sfumature che di scelte nette. Ma la scansione può essere utile a mettere ordine nella trattazione e proverò a seguirla nelle prossime pagine, dedicando un capitolo a ciascuna delle prime due tipologie e integrando poi nel capitolo 17 la terza e la quarta prospettiva, che, dal particolare punto di vista adottato in questo libro, è utile affrontare in maniera congiunta.

¹ Per le tematiche discusse in questa parte del libro ha un particolare rilievo il quadro complessivo fornito in Maurizio Ferraris, *Documentalità. Perché è necessario lasciar tracce*, Laterza, Roma-Bari 2009, in parte rivisto e ampliato in Id., *Documanità. Filosofia del mondo nuovo*, Laterza, Bari-Roma 2021.

² I capitoli dal 14 al 19 del libro riprendono, con integrazioni e modifiche, il mio intervento al convegno *Personal Digital Memories: i fondi di persona dall'analogico al digitale*, organizzato a Ravenna il 14 e 15 dicembre 2021 dal Dipartimento di Beni Culturali dell'Università di Bologna con la responsabilità scientifica di Stefano Allegrezza; le registrazioni delle sessioni sono tutte disponibili nel canale YouTube del dipartimento, all'interno della playlist dedicata a conferenze, seminari e convegni: <https://www.youtube.com/playlist?list=PLGylMNAXNrBRY018wAopENYSgxDeqE5ld>. Il mio intervento si è svolto in apertura della prima sessione, e il relativo video è raggiungibile attraverso il QR-Code inserito nel testo. Gli atti del convegno sono stati editi nel 2023: Stefano Allegrezza (a cura di), *Personal Digital Memories: i fondi di persona*

dall'analogico al digitale, Civita Editoriale, Torre del Lago Puccini 2023. Il mio intervento è alle pagine 9-25.

³ Andy Clark e David Chalmers, *The Extended Mind*, in «Analysis» 58, 1, gennaio 1998, pp. 7-19. Il testo dell'articolo è disponibile online in vari siti, molti dei quali desumibili dalla pagina <https://philpapers.org/rec/CLATEM>.

⁴ Cfr. Carlo Cellucci, *Mente incarnata e conoscenza*, in Eugenio Canone (a cura di), *Per una storia del concetto di mente*, Olschki, Firenze 2005, pp. 383-410 (in particolare, sezione 5).

⁵ Per una sintesi si vedano Richard Menary (a cura di), *The Extended Mind*, MIT Press (Bradford Books), Cambridge (MA) 2012, e, in italiano, Michele Di Francesco e Giulia Piredda, *La mente estesa. Dove finisce la mente e comincia il resto del mondo?*, Mondadori, Milano 2012.

⁶ Linda A. Henkel, *Point-and-Shoot Memories: The Influence of Taking Photos on Memory for a Museum Tour*, in «Psychological Science» 25 (2), 2014, pp. 396-402. Per una rassegna sul tema è utile Evangelos Niforatos, *The Role of Context in Human Memory Augmentation*, tesi di dottorato, Facoltà di Informatica dell'Università della Svizzera italiana, aprile 2018, <https://doc.rero.ch/record/309482/files/2018INFO005.pdf>.

⁷ Su big data digitali, memoria e costruzione dell'identità personale, interessanti considerazioni in Jacquelyn Ann Burkell, *Remembering me: big data, individual identity, and the psychological necessity of forgetting*, in «Ethics and Information Technology» 18.1, 2016, pp. 17-23, <https://doi.org/10.1007/s10676-016-9393-1>.

⁸ Cfr. Emmaline Drew Eliseev ed Elizabeth J. Marsh, *Externalizing autobiographical memories in the digital age*, in «Trends in Cognitive Sciences» 25 (12), dicembre 2021, pp. 1072-1081, doi: 10.1016/j.tics.2021.08.005.

15.

Trasferimenti di memoria

Il processo di digitalizzazione di memorie personali nate su supporti tradizionali è già in corso da decenni, e conosciamo abbastanza bene le sue caratteristiche. Ciò non toglie, tuttavia, che si tratti di un ambito decisamente articolato, sia nelle tipologie documentali coinvolte (corrispondenza, fotografie, manoscritti e dattiloscritti di varia origine e natura, ma anche certificati e diplomi, documenti di identità, contratti e transazioni, analisi mediche, ricevute...), sia nelle forme e modalità della digitalizzazione, che può avvenire a cura delle istituzioni di conservazione nel caso di fondi depositati, ma può avvenire anche da parte della singola persona interessata, o di parenti, amici, eredi. E si stanno diffondendo negli ultimi anni anche forme di digitalizzazione ‘ibrida’, in cui istituzioni – in particolare biblioteche e archivi – o aziende mettono a disposizione di privati, gratis o a pagamento, strumenti e competenze necessarie a un lavoro di digitalizzazione delle memorie personali di migliore qualità: un’evoluzione riassunta dallo slogan «bring preservation to the people»¹.

Non è questa la sede per descrivere l’evoluzione delle pratiche e degli strumenti utilizzati², ma vale la pena osservare che anche in questo campo l’evoluzione relativa agli standard descrittivi e ai modelli concettuali si è finora mossa quasi sempre nell’ambito della tradizionale organizzazione ‘architettonica’ propria dell’archivistica, con un’evoluzione negli ultimi anni verso il modello linked data non dissimile da quella incontrata nella parte prima di questo libro³. Vorrei però sottolineare un aspetto – anch’esso ben noto – che merita particolare attenzione, anche in relazione al tema dei possibili riusi dei materiali digitalizzati: i contenuti prodotti dal processo di digitalizzazione non sono una mera riproposizione in altra forma degli oggetti documentali di partenza, né un semplice strumento

integrativo di conservazione e di accesso: si tratta di oggetti documentali *nuovi*.

In altri termini: la digitalizzazione è un processo di produzione di contenuti che – nonostante la loro ovvia relazione con ciò che viene digitalizzato – hanno caratteristiche specifiche che vanno ben oltre la semplice funzione di riproduzione dell'originale. Una considerazione, questa, proposta molto chiaramente già vent'anni fa da Stefano Vitali in occasione di uno dei primi appuntamenti nella serie di convegni *Conservare il Novecento*:

Nel caso della migrazione verso il digitale di documenti originariamente messi in essere su altri supporti, il carattere di 'montaggio' connaturato alle fonti digitali implica, in primo luogo, che le riproduzioni digitali non possano mai essere viste come una sorta di 'doppio speculare' della fonte originaria. Esse devono essere considerate piuttosto come una fonte nuova a tutti gli effetti [...]⁴.

Questa considerazione ha un corollario: con il passaggio al digitale, archivi e biblioteche impegnati in progetti di digitalizzazione diventano soggetti produttori di oggetti documentali di tipo nuovo, e non solo soggetti preposti alla conservazione e all'accesso. Non solo: questa attività produttiva implica scelte che per un verso inevitabilmente *interpretano* l'informazione di partenza, e per altro verso la *modificano*, selezionando (anche se in maniera auspicabilmente non arbitraria) elementi che trovano rappresentazione anche nel nuovo oggetto digitale ed elementi che non la trovano, e aggiungendo informazione originariamente non presente. Quando, nell'ambito della discussione in margine al convegno, l'avevo fatto notare, la reazione da parte di alcuni colleghi archivisti era stata abbastanza critica: l'immagine dell'archivista (o del bibliotecario) 'neutrale', la cui missione dovrebbe essere solo quella di preservare e garantire l'accesso, senza alcun sovrappiù interpretativo (e tantomeno creativo), era profondamente radicata. La rivoluzione digitale ha avuto, fra gli altri, l'effetto di rendere abbastanza evidente l'ingenuità di questa assunzione.

Uno degli aspetti di novità dei prodotti del processo di digitalizzazione è la facilità con la quale essi possono essere riutilizzati e ricombinati, anche in contesti nuovi. Le possibilità di riuso, mash-up e ri-mediazione⁵ sono forse anche maggiori nel campo delle memorie personali, che hanno un

carattere spesso più granulare e meno organizzato di quelle prodotte da istituzioni, e si prestano quindi meglio a giochi di ricombinazione. Gli esempi che si possono portare sono molteplici; a puro titolo di esempio, ne vorrei ricordare rapidamente due: la raccolta di più memorie personali per costruire memorie collettive o trasversali di eventi o di luoghi, e la rielaborazione crossmediale di memorie personali individuali.

Il 'lungo' anniversario della Prima guerra mondiale ha offerto un'occasione unica per lavorare sulla prima prospettiva, e assieme a Federico Meschini mi sono occupato altrove di alcuni dei risultati di questo lavoro⁶; ovviamente vi sono molti altri esempi possibili: Serge Noiret ne ha analizzati diversi nei suoi numerosi contributi dedicati al tema della digital public history, con particolare riferimento all'uso delle immagini⁷.

Un progetto recente che ricade in quest'ambito è Jena4D⁸, promosso dall'Università di Jena e dalla Thüringer Universitäts- und Landesbibliothek, che si propone di aggregare fotografie della città di Jena provenienti per lo più da collezioni personali e relative a diverse epoche, per ricostruirne collaborativamente la storia. Le fotografie sono fruibili attraverso l'applicazione web 4dcity.org, sviluppata nell'ambito del lavoro della cattedra di digital humanities dell'Università di Jena, che permette di proiettare fotografie cronoreferenziate (ma anche altre tipologie di informazioni, ad esempio riferimenti all'interno di voci di Wikipedia) sulla ricostruzione 3D di un luogo e di navigarle interattivamente muovendosi sia nello spazio sia nel tempo. Il breve filmato raggiungibile attraverso il QR-Code qui di seguito, inserito su YouTube da Sander Münster (al quale si deve anche un volume sulla collaborazione interdisciplinare nella realizzazione di modelli storici 3D⁹), permette di capire il funzionamento di applicazioni di questo tipo assai meglio di quanto non si possa fare attraverso una descrizione a parole. Il progetto rientra a sua volta nel più ampio contesto della *European Time Machine*¹⁰, una fra le più interessanti e innovative iniziative di ricerca europee al confine fra digital public history, georeferenziazione e valorizzazione del patrimonio culturale.



QR-Code 18

Jena4D

<http://bit.ly/3pEIPdi>

All'ambito della rielaborazione crossmediale di memorie personali appartengono invece i memofilm, che a partire da fotografie, immagini, suoni, lettere, narrazioni, costruiscono un filmato che testimonia e riassume la storia di un individuo. I memofilm sono nati a scopo terapeutico, come strumento nella lotta contro l'Alzheimer, da un'intuizione dello sceneggiatore e regista Ernesto Melloni. Nel 2004, per fronteggiare le conseguenze dell'Alzheimer del padre, Melloni ebbe l'idea di realizzare una sorta di mini-documentario, costruito per montaggio di immagini, filmati e narrazione, che ne ricostruisse la storia personale. Il documentario veniva mostrato al padre ogni mattina, per rinforzare i suoi ricordi e il suo senso di identità. L'idea ha suscitato l'interesse dell'Azienda sanitaria provinciale Giovanni XXIII di Bologna e della Cineteca di Bologna, che hanno avviato – con ottimi risultati – una sperimentazione clinica. Dal progetto sono nate un'associazione¹¹, un libro¹² e una ricca casistica di memofilm, alcuni dei quali inclusi nel DVD che accompagna il libro. Il QR-Code qui di seguito vi porta a un filmato YouTube che descrive il progetto attraverso un'intervista a Melloni, altri materiali video sono disponibili sul canale YouTube 'Memo Film'¹³.



QR-Code 19

Intervista a Ernesto Melloni: i memofilm

<http://bit.ly/3O5pbAo>

Indipendentemente da alcuni esiti più recenti e poco condivisibili del progetto (in periodo Covid, alcuni degli esponenti dell'associazione hanno assunto posizioni anti-vaccino inconciliabili con lo stretto collegamento con il mondo della ricerca sanitaria indispensabile per una effettiva validazione di protocolli e metodi), l'interesse dei memofilm è indubbio: si tratta di una forma di riuso delle memorie personali che ne conferma la fondamentale importanza come strumento di costruzione dell'identità, e che nel contempo ben mostra come il digitale permetta di costruire oggetti informativi di tipo radicalmente nuovo anche partendo da memorie 'tradizionali' di un singolo individuo.

¹ Katlin Seagraves, *Digitization and Personal Digital Archiving*, in «Library Technology Reports» 56, 5, luglio 2020, pp. 22-28, <https://link.gale.com/apps/doc/A632141258/AONE>. Il riferimento è in particolare all'esperienza del Memory Lab Network, <https://memorylabnetwork.github.io/index.html>.

² Fra i molti riferimenti possibili, mi limito qui a ricordare Donald T. Hawkins (a cura di), *Personal Archiving: Preserving Our Digital Heritage*, Information Today, Medford (NJ) 2013, e Brianna H. Marshall (a cura di), *The Complete Guide to Personal Digital Archiving*, ALA Editions, Chicago 2018.

³ Si veda al riguardo Francesca Tomasi, *Archivi di persona in linked open data: il modello concettuale come strumento di integrazione nei GLAMs*, in «AIB Studi» 57 (2), 2017, <https://doi.org/10.2426/aibstudi-11647>.

⁴ Stefano Vitali, *I bit in archivio: un sovrappiù di critica?*, in Maurizio Messina e

Giuliana Zagra (a cura di), *Conservare il Novecento: oltre le carte. Convegno nazionale, Ferrara, Salone internazionale dell'arte del restauro e della conservazione dei beni culturali e ambientali, 5 aprile 2002. Atti*, AIB, Roma 2003, pp. 49-66, p. 63.

⁵ Il concetto di 'remediation', come discusso in Jay David Bolter e Richard Grusin, *Remediation: Understanding New Media*, MIT Press, Cambridge (MA) 2000, è applicato sempre più spesso anche in campo archivistico, e quello delle memorie personali è uno degli ambiti in cui questa applicazione sembra più fruttuosa: cfr. ad es. Sonja Vivienne e Jean Burgess, *The remediation of the personal photograph and the politics of self-representation in digital storytelling*, in «Journal of Material Culture» 18 (3), 2013, pp. 279-298, <https://doi.org/10.1177/1359183513492080>.

⁶ Federico Meschini e Gino Roncaglia, *Narrazioni trasversali del conflitto: risorse online per la Grande guerra*, in «Bibliologia: an International Journal of Bibliography, Library Science, History of Typography and the Book» 11, 2016, pp. 125-136.

⁷ Per una rapida sintesi si veda Serge Noiret, *Storia Pubblica Digitale*, in «Zapruder. Storie in Movimento» 36, gennaio-aprile 2015, pp. 8-22. Sul tema del rapporto fra personal memories e digital public history si veda Valérie Schafer, *Digital Personal Memories: The Archiving of the Self and Public History*, in Serge Noiret, Mark Tebeau e Gerben Zaagsma (a cura di), *Handbook of Digital Public History*, De Gruyter Oldenbourg, Berlin-Boston 2022, pp. 377-384, <https://doi.org/10.1515/9783110430295-033>.

⁸ Il sito di riferimento è <https://www.uni-jena.de/220111-jena4d>.

⁹ Sander Münster, *Interdisziplinäre Kooperation bei der Erstellung geschichtswissenschaftlicher 3D-Modelle*, Springer, Cham-Wiesbaden 2016.

¹⁰ Il sito di riferimento è <https://www.timemachine.eu/>.

¹¹ Il sito di riferimento è <https://memofilmtheweb.wordpress.com/>.

¹² Luisa Grosso (a cura di), *Memofilm. La creatività contro l'Alzheimer*, Mimesis, Milano-Udine 2015.

¹³ <https://www.youtube.com/@memofilm8797>.

16.

Il sogno della memoria totale

Anche il progressivo passaggio al digitale del processo di produzione di molte tipologie già note di memorie personali è ben conosciuto e studiato: dalle lettere alle mail, dalla fotografia su pellicola alla fotografia digitale, dal diario ai blog (e, con un processo di progressivo allontanamento dalle forme della tradizione, ai messaggi di stato sui social network), è la quasi totalità della nostra attività di ‘produzione di tracce’ che ormai si colloca saldamente all’interno dell’ecosistema digitale¹. E sappiamo bene che, in forma ancor più radicale di quanto non avvenga nel caso della digitalizzazione e anche quando il modello di partenza sembra abbastanza chiaro, il passaggio al digitale comporta la produzione di oggetti informativi con caratteristiche del tutto nuove. Una novità che riguarda almeno cinque sfere: a) quella dei supporti (che si moltiplicano, pongono problemi specifici di accesso e conservazione nel tempo, e – con il progressivo passaggio al cloud – sfuggono per molti versi al controllo dello stesso soggetto produttore); b) quella del contenuto dell’oggetto digitale, che include spesso un sovrappiù informativo – in parte generato automaticamente – rispetto al suo equivalente analogico (si pensi ai dati di crono- e georeferenziazione delle fotografie, alla data e ora di invio di una mail o di pubblicazione di un post), ma che è anche molto più facilmente alterabile e modificabile; c) quella della quantità, testimoniata dalla vera e propria inflazione di contenuti che nel campo dei testi è legata soprattutto alla diffusione dei social network e nel campo della fotografia, del video e dell’audio è conseguenza diretta della diffusione dei dispositivi digitali e in particolare dello smartphone, che costituisce ormai una macchina fotografica e di ripresa sempre disponibile² e con caratteristiche innovative di gestione e manipolazione dei contenuti prodotti; è soprattutto a questa

dimensione che fa riferimento al concetto di post-fotografia che torna spesso nella riflessione degli ultimi anni sul tema³; d) quella della facilità di duplicazione e pubblicazione, che permette di condividere con altri o di rendere pubbliche le memorie personali praticamente in contemporanea con la loro produzione; e) quella dei contesti di uso o riuso, che – anche in conseguenza della facilità di duplicazione – si moltiplicano con estrema rapidità.

Mi limiterò qui a discutere brevemente un esempio sicuramente estremo, ma proprio per questo significativo: la possibilità di costruire una sorta di ‘memoria completa’ della propria vita attraverso l’uso di dispositivi personali per la registrazione continua e automatica di video e audio e attraverso la digitalizzazione e conservazione integrale di tutti i testi e materiali letti e prodotti. Un’idea che avrebbe probabilmente affascinato Borges e Perec e che è stata proposta – associata alle nuove potenzialità dei dispositivi digitali personali – da Nathan Myhrvold, eclettico e visionario ingegnere informatico, allievo di Stephen Hawking e a lungo chief technology officer della Microsoft⁴, e da Gordon Bell, che sempre in Microsoft è stato il principale propugnatore e insieme il soggetto sperimentale del progetto di ‘life-logging’ denominato MyLifeBits.

Nel settembre del 1993 Myhrvold scrisse un memo interno di trenta pagine, rapidamente diventato uno dei documenti di riferimento per il lavoro sul futuro della rivoluzione digitale, intitolato *Road Kill on the Information Highway*⁵. Un testo che si soffermava ampiamente proprio sull’idea di registrazione e conservazione digitale delle memorie personali. Vale la pena riportarne per esteso due passi, riferiti rispettivamente alle registrazioni audio e a quelle video:

Consideriamo la conversazione parlata. Anche se possiamo citarci a vicenda quando diciamo qualcosa di memorabile, questa è l’eccezione piuttosto che la regola. Le persone tendono a considerare il linguaggio parlato come un mezzo effimero: se vogliamo essere sicuri di potervi accedere in futuro, facciamo uno sforzo speciale per prendere appunti o registrare. In un futuro prossimo avremo la capacità di registrare e indicizzare tutto ciò; [...] un singolo supporto dal costo di circa 10 dollari dovrebbe essere in grado di contenere tutte le conversazioni di un individuo, in tutta la sua vita [...].

Quanto tempo ci vorrà prima che ogni macchina della polizia, o ogni agente di polizia, sia dotato di una telecamera video digitale con timbri di data e posizione non

falsificabili? Gli aerei di linea commerciali hanno da tempo una ‘scatola nera’ o un registratore di volo, ma perché non estendere questa tecnologia anche per autobus, automobili e persino per le persone singole? [...] Ancora un paio di anni di miglioramenti nel rapporto qualità/prezzo e potremmo portare con noi un dispositivo del genere. La sua capacità di archiviazione digitale sarebbe sufficiente per registrare tutto ciò che diciamo e facciamo. [...] Un tale dispositivo potrebbe contenere l’intera vita di una persona: il diario e l’autobiografia definitivi. [...] Le compagnie assicurative sarebbero felici di avere accesso alla scatola nera sulla tua auto, sia per rintracciarla in caso di furto [...] sia per fornire informazioni in caso di incidenti. Non è difficile immaginare di poter avere un’assicurazione auto a metà prezzo se hai un’auto con un registratore. [...] Le assicurazioni per negligenza medica potrebbero essere più economiche (o disponibili solo) per i medici che registrano le procedure chirurgiche o anche le visite in ambulatorio. Molti datori di lavoro vorranno utilizzare questi dati come strumento di gestione. Le aziende di autobus e camion hanno un interesse evidente nel sapere se i loro autisti rispettano la legge e sono puntuali. Il pubblico potrebbe insistere affinché anche la polizia registri le proprie azioni durante lo svolgimento delle sue attività⁶.

Nel memo di Myhrvold emergono con chiarezza sia la fascinazione sia il carattere assai inquietante e gli enormi rischi, in particolare per la privacy, di questa prospettiva di ‘registrazione totale’. E non occorre sottolineare come molte delle sue previsioni, all’epoca quasi visionarie, si siano puntualmente avverate: i poliziotti americani utilizzano videocamere personali sempre accese, le compagnie di assicurazione offrono effettivamente tariffe agevolate per chi accetta di installare nell’auto apparati di registrazione di velocità e posizione, e molti automobilisti installano sulle proprie autovetture videocamere che conservano decine di ore di video registrate automaticamente, utili a scopo di documentazione in caso di incidente. In alcune situazioni e in alcune aziende si sono purtroppo avverate anche alcune delle previsioni più preoccupanti, come il controllo continuo del lavoro dei dipendenti attraverso dispositivi di tracciamento personale.

Sulla base di questa idea di ‘registrazione totale’, Gordon Bell ha avviato nel 2001 il progetto MyLifeBits, basato sull’uso di una videocamera indossabile per la registrazione continua di video e audio (incluso l’audio delle telefonate), di un GPS per la registrazione continua dei propri movimenti e itinerari, e sulla digitalizzazione e conservazione di tutti i testi letti o prodotti: mail, documenti, pagine web, messaggi di testo... Una affascinante (e anche divertente) presentazione di questo lavoro è

stata proposta dallo stesso Gordon Bell in una conferenza alla Library of Congress, che potete vedere, nella registrazione su YouTube, attraverso il QR-Code qui di seguito. Fulcro di questo progetto, un software per la conservazione e la ricerca di questa enorme mole di dati personali, denominato a sua volta MyLifeBits e sviluppato, sempre in ambito Microsoft, da Jim Gemmell e Roger Lueder: un software che ha anticipato di diversi anni funzionalità oggi offerte su larga scala da Google (e non solo), come il posizionamento su mappe di fotografie geolocalizzate o la registrazione automatica dei propri itinerari. Gordon Bell e Jim Gemmell hanno dedicato al progetto e alle sue motivazioni il volume *Total Recall*⁷, che rappresenta una sorta di manifesto a favore della conservazione e accessibilità totale delle memorie personali e discute sia l'idea del potenziamento della memoria personale (chiara eredità dell'idea di 'human augmentation' teorizzata da tanta parte della già menzionata 'ideologia di Silicon Valley'⁸), sia l'idea della possibile utilità in ambito lavorativo, educativo, medico di strumenti e pratiche digitali di 'memory augmentation', sia infine l'idea di immortalità digitale, su cui tornerò brevemente in seguito.



QR-Code 20

Gordon Bell parla del progetto MyLifeBits alla Library of Congress

<http://bit.ly/44neSh6>

È difficile stabilire se il sogno di Bell e Gemmell rappresenti un'utopia o un incubo (forse, una curiosa fusione delle due cose), ma alla luce dell'evoluzione successiva dell'ecosistema digitale e della pervasività degli smartphone – che già oggi non solo archiviano, per nostra scelta o, sempre più spesso, automaticamente, molti dei nostri dati, ma in molte situazioni ce li ripropongono in maniera autonoma, ad esempio sotto

forma di ricordi su quel che abbiamo fatto o fotografato nella stessa giornata degli anni passati – la tematica ha un indubbio rilievo e porta a ripensare sotto una diversa luce la stessa idea di memorie digitali personali, così come le caratteristiche dei sistemi per la loro gestione⁹.

¹ Cfr. Stefano Allegrezza, *The Future of Our Personal Digital Memories: It's Time to Start Thinking About It*, in «Atlanti +» 29 (1), 2019, pp. 55-65, <https://doi.org/10.33700/2670-4579.29.1.55-65>.

² Una osservazione che Roberto Casati aveva avanzato già nel 2004 a proposito dei primi modelli di telefono cellulare con capacità di scattare fotografie (ricordiamo che l'iPhone, generalmente considerato il primo smartphone in senso pieno, entra in produzione nel 2007): «In pochi escono di casa con una macchina fotografica in tasca, ma se questa è accorpata a un telefonino si finisce con l'aver sempre a portata di mano la possibilità di catturare immagini: con conseguente inflazione del numero delle riprese»: Roberto Casati, *La vita scorre a bassa definizione*, in «Il Sole-24 Ore», Supplemento domenicale, 27 giugno 2004, p. 31.

³ Senza avventurarmi, con troppo scarse competenze, in una bibliografia davvero sterminata, mi limito a ricordare Joan Fontcuberta, *La cámara de Pandora. La fotografi@ después de la fotografía*, Editorial GG, Barcelona 2012 (trad. ingl. *Pandora's Camera: Photogr@phy After Photography*, MACK, London 2014); Joan Fontcuberta, *La furia de las imágenes. Notas sobre la postfotografía*, Galaxia Gutenberg, Barcelona 2016 (trad. it. *La furia delle immagini. Note sulla postfotografia*, Einaudi, Torino 2018). Fontcuberta individua nelle due categorie dell'eccesso (il carattere inflattivo della fotografia digitale) e dell'accesso (la facilità di pubblicazione e condivisione) le principali innovazioni della 'post-fotografia' rispetto alla fotografia tradizionale.

⁴ Per una ricostruzione giornalistica ma dettagliata e piuttosto affascinante delle sue idee e del suo lavoro in Microsoft nei primi anni '90 del secolo scorso si veda l'articolo di Ken Auletta *The Microsoft Provocateur*, in «The New Yorker», 4 maggio 1997, <https://www.newyorker.com/magazine/1997/05/12/the-microsoft-provocateur>.

⁵ La cui digitalizzazione è disponibile online all'indirizzo <https://nsarchive.gwu.edu/document/22837-document-03-nathan-p-myhrvold-microsoft>.

⁶ *Ibid.*

⁷ Gordon Bell e Jim Gemmell, *Total Recall: How the E-Memory Revolution Will Change Everything*, Penguin, London-New York 2009. Il libro è stato pubblicato in paperback nel 2010 da Plume con il titolo *Your Life, Uploaded: The Digital Way to*

Better Memory, Health, and Productivity, trad. it. *Total Recall. Memoria totale. Ricordare tutto? Inquietante, ma reale*, Rizzoli, Milano 2010.

⁸ A partire dal fondamentale report di Douglas C. Engelbart *Augmenting Human Intellect: A Conceptual Framework*, SRI Summary Report AFOSR-3223, Air Force Office of Scientific Research, ottobre 1962, <https://www.dougenelbart.org/pubs/augment-3906-Framework.html>.

⁹ Cfr. Liadh Kelly, *The Information Retrieval Challenge of Human Digital Memories*, intervento al BCSIRSG Symposium: Future Directions in Information Access (FDIA), 2007, https://doras.dcu.ie/344/1/irsg_2007.pdf.

17.

Nuove memorie personali e big data

Il progetto MyLifeBits rappresenta già una integrazione fra memorie digitali che offrono maggiore continuità rispetto al passato analogico (come fotografie, documenti, registrazioni audio) e memorie digitali di tipo nuovo, come i dati sugli spostamenti ricavati da un sistema GPS. Ovviamente la differenza tra queste due categorie è spesso questione di gradi, e – volendo – qualche forma di antenato analogico, più o meno lontano, si può trovare per quasi tutte le tipologie di contenuti digitali. Ma è difficile negare che, ad esempio, i post sui social network siano in genere più lontani da forme di documentalità tradizionale di quanto non lo siano i post su un blog, che una videoconferenza sia assai diversa da una riunione alla quale si partecipi di persona (e lasci tracce diverse, inclusa la possibilità della registrazione integrale), che paragonare i sistemi di instant messaging come WhatsApp ai telegrammi o alle cartoline postali rappresenti una forzatura, pur se a volte giustificata dall'esigenza di mostrare aspetti di continuità nell'evoluzione degli usi della lingua¹, e che conservare traccia dell'itinerario di un'escursione attraverso il percorso appuntato a matita su una carta geografica o attraverso la geolocalizzazione via GPS siano due cose assai diverse, anche se il risultato può sembrare abbastanza simile. E nel caso di dati come quelli relativi alla gestione di 'memorie' all'interno di un videogioco online o riferibili a un evento nel metaverso (spostamenti in un ambiente virtuale, dati di gioco, oggetti virtuali di cui si entra in possesso, ecc.) ogni paragone sensato con il passato analogico sembra venir meno.

Un esempio interessante, da questo punto di vista, è rappresentato dal lavoro di Tom Boellstorff, antropologo dell'università di California, che ha studiato ambienti virtuali come Second Life analizzando su un lasso di

tempo di oltre due anni il comportamento di un campione di utenti da punti di vista come quello dell'identità di genere, delle abitudini di movimento, di vita e di consumi all'interno del mondo virtuale, dei conflitti interpersonali, dello sviluppo delle comunità². In molti casi, il suo lavoro (ma anche quello di altri 'etnografi digitali'³) presuppone o include la ricostruzione della 'storia interna' di comunità virtuali abbastanza diverse fra loro, che può basarsi su interviste tradizionali ma anche sullo studio di 'tracce', memorie o avvenimenti del tutto o in parte interni al mondo virtuale. Nel frattempo, Second Life è passata di moda, ma di comunità virtuali online si continua – e si continuerà – a parlare, anche come luoghi di produzione di tracce e memorie, e le ricerche svolte al riguardo conservano tutto il loro interesse.

È impossibile in questa sede discutere in maniera sensata, o anche solo provare ad elencare, tutte le nuove (o parzialmente nuove) tipologie di memorie personali digitali. Se penso a una mia giornata-tipo, al risveglio – spesso prima ancora di alzarmi dal letto – guardo i dati su durata e qualità del sonno registrati dallo smartphone, guardo l'elenco degli impegni che mi propone Alexa su uno schermo che ricorda solo lontanamente una sveglia, controllo i messaggi WhatsApp e Telegram, scorro rapidamente la posta elettronica e i quotidiani nell'edicola digitale di MLOL; più tardi Google Maps mi ricorderà dove ho parcheggiato la macchina, mi suggerirà – in funzione del traffico – l'itinerario migliore per l'università, e terrà traccia dei miei spostamenti; guidando, chiederò ad Alexa di proseguire su Audible o Storytel la lettura dell'audiolibro iniziato il giorno prima (e tanto Audible quanto Storytel ricorderanno dove era stata interrotta e conserveranno statistiche su tempi e contenuti dell'ascolto), o chiederò a Spotify di farmi ascoltare una delle molte playlist che ho preparato nel tempo (e a fine anno Spotify mi proporrà una dettagliata classifica dei brani che ho ascoltato più spesso, con i tempi di ascolto di ognuno); arrivato in aula, lo schermo dell'aula mi ricorderà di avviare lo streaming e la registrazione della lezione e Prezi mi aiuterà non solo a presentarla ma anche a ricordarla e seguirne il filo; probabilmente consulterò il forum di discussione che ho predisposto sul Moodle del corso, per vedere se ci sono domande a cui rispondere (certo, le domande possono essere fatte anche in aula, ma il forum permette a

studentesse e studenti di porle in qualunque momento, anche durante lo studio a casa, senza rischiare di dimenticarle); se sono in un (raro) periodo di buona volontà alimentare, a pranzo userò una app sul telefonino per calcolare e tener traccia delle calorie del pasto (in quel caso mi sarò probabilmente pesato la mattina prima di colazione, e la bilancia avrà inviato automaticamente a quella stessa app i dati); dopo pranzo e la sera prima di andare a dormire, nelle due finestre dedicate alla lettura, leggerò qualcosa o su un e-reader (che conserverà sia traccia delle mie letture sia le mie annotazioni) o su carta (ma in quel caso è molto probabile che qualche annotazione finisca sullo smartphone, provvisto di stilo e di Evernote per gestire note e appunti); nel pomeriggio probabilmente mi aspetta almeno una videoconferenza e spesso due o tre, magari su piattaforme diverse; se devo fare la spesa, Google o Alexa terranno traccia della lista di cose da comprare, mentre sto cucinando funzioneranno da timer, e nel corso della giornata risponderanno ad altre richieste di vario genere, fra cui l'accensione e lo spegnimento delle luci di casa; subito dopo cena, se non ci sono altri impegni, il televisore mi proporrà di proseguire nella visione di una serie televisiva (e saprà dove ero rimasto) o mi suggerirà altri contenuti, sulla base delle scelte precedenti. E sono solo alcuni esempi: Google Photo e Amazon Photo conservano le mie fotografie e i miei video, Instagram e (sempre più raramente) Facebook conservano impressioni e annotazioni, Swarm il ricordo delle pause gastronomiche o delle visite a musei, durante l'estate uso WikiLoc per tener traccia delle camminate in montagna... Dieci anni fa l'uso di questa girandola di strumenti, molti dei quali basati sull'accoppiata smartphone-cloud, sarebbe sembrata una follia, e forse in parte lo è; fra dieci anni, se sarò ancora in giro, gli strumenti che userò saranno probabilmente almeno in parte diversi, ma difficilmente saranno meno numerosi o meno differenziati.

Quel che dicono questi esempi, assolutamente occasionali e parziali, è che le nuove memorie digitali personali sono a) molte e molto diverse fra loro; b) inflattive, sia nel numero e nelle funzioni degli strumenti usati, sia nella quantità e tipologia delle tracce prodotte e conservate; c) prodotte e gestite sempre più spesso attraverso la mediazione di app e sistemi intelligenti anziché direttamente dall'utente (che in molti casi può anche non essere consapevole della loro esistenza: Derrick de Kerckhove ha

suggerito a questo proposito la suggestiva espressione «inconscio digitale», un'idea presentata nel TED Talk – in italiano – che potete seguire attraverso il QR-Code qui di seguito⁴); d) basate per lo più sulla conservazione di dati personali su cloud; e) facilmente duplicabili e condivisibili attraverso uno spettro di strumenti digitali per la comunicazione e la socializzazione di contenuti a sua volta ampio e differenziato. Con l'evoluzione dell'ecosistema digitale, il processo di esternalizzazione della memoria ha dunque fatto un salto quantitativo e qualitativo impressionante, i cui effetti in termini tanto cognitivi quanto di costruzione dell'identità sono talmente rilevanti da poter essere considerati per molti versi come indizio di un vero e proprio passaggio evolutivo.



QR-Code 21

Derrick de Kerckhove su big data e inconscio digitale

<http://bit.ly/44gaznD>

Per questa via, anche le memorie personali tendono a trasformarsi in big data (e, d'altro canto, molte tipologie di big data includono componenti rilevanti dal punto di vista delle memorie personali). Un passaggio che pone evidentemente problemi di selezione, gestione, conservazione, ricerca, accesso, in gran parte nuovi. Già nel 2007, discutendo il concetto di Human Digital Memories (HDM) e i relativi problemi di information retrieval, Liadh Kelly osservava:

Le HDM (Human Digital Memories) sono fondamentalmente diverse dagli archivi tradizionali per i quali sono state sviluppate le tecniche di ricerca esistenti. Nelle HDM si trovano tipicamente combinazioni di diversi tipi di media, come audio, video, immagini e contenuti testuali di vario genere. Questi archivi possono contenere una grande quantità di rumore, e molti elementi nell'archivio possono essere molto simili,

riguardando lo stesso argomento; è possibile che l'utente non sia a conoscenza del fatto che un particolare dato è stato registrato e che è quindi disponibile per il recupero; inoltre, l'utente potrebbe avere difficoltà a descrivere chiaramente ciò che sta cercando; gli elementi nell'archivio potrebbero non avere descrizioni testuali formali, il che implica che non possono essere recuperati utilizzando metodi tradizionali basati sul testo o sui metadati; gli elementi potrebbero non essere collegati tra loro tramite link, quindi può non esservi una struttura di rimandi da utilizzare nel processo di recupero. Questa combinazione unica di caratteristiche propria delle HDM motiva la ricerca nello sviluppo di tecniche di recupero specificamente per il dominio degli archivi personali. Si tratta di un ambito fondamentalmente diverso da quelli coinvolti nelle forme tradizionali di recupero dell'informazione, sia a causa della combinazione dei fattori sopra descritti, sia perché gli elementi presenti nelle HDM sono specifici per ogni singolo individuo e le memorie legate a elementi come la creazione e l'accesso sono a loro volta memorie individuali⁵.

Negli anni successivi su questi temi si è sviluppato un dibattito che ha sottolineato sia l'esplosione quantitativa delle memorie digitali personali, sia la nascita di nuove tipologie di dati⁶, e ben presto il tema è stato esplicitamente collegato, anche in letteratura, a quello dei big data⁷. Nello stabilire questo collegamento, un ruolo fondamentale lo hanno evidentemente, come già ricordato, gli smartphone, che fanno esplodere la produzione e la granularità delle memorie personali digitali⁸.

Nella prospettiva dei big data, il tema delle memorie personali digitali si arricchisce evidentemente di nuove dimensioni, molte delle quali devono probabilmente ancora essere esplorate, e che chiamano in causa tanto l'architetto quanto l'oracolo. Per fare solo qualche esempio: quali strumenti di data mining, data analytics e di data visualization potrebbero essere utilmente introdotti dalla ricerca in questo ambito? È possibile generare automaticamente metadati, e sulla base di quali ontologie? Quali strumenti e strategie di archiviazione offrono prospettive migliori di conservazione di lungo termine? Che tipo di protezione deve essere garantita a questi dati? Chi ne è il proprietario, e cosa ne succede (o ne dovrebbe succedere) dopo la morte del soggetto produttore⁹? Ha senso utilizzare strumenti automatici di scarto, e – in tal caso – quali, e sulla base di quali politiche? È possibile ipotizzare anche per le memorie digitali personali strategie di 'distant reading' simili a quelle discusse in altri campi sulla base delle prospettive aperte dai lavori di Franco Moretti¹⁰? I big data che produciamo costruiscono un nostro 'gemello

digitale¹¹, o magari ci aprono addirittura una prospettiva di immortalità digitale¹²?

Queste ultime prospettive, in particolare, hanno suscitato recentemente un certo interesse proprio per la loro possibile associazione con i nuovi sviluppi dell'intelligenza artificiale. Se costruiamo un corpus a partire dalla raccolta di memorie personali di un individuo (ad esempio, attraverso l'insieme dei testi che ha scritto: libri, articoli, mail, messaggi, post sui social network...) e lo utilizziamo per addestrare una intelligenza artificiale generativa, il risultato sarà un sistema che potrà in qualche modo 'simulare' il comportamento linguistico dell'individuo in questione e far riferimento alla componente della sua memoria che è stata esternalizzata e assorbita dal sistema¹³.

Credo sia troppo presto per proporre qui una rassegna o anche solo qualche esempio significativo del lavoro avviato in questo campo, ancora assai immaturo: il rischio è quello di dar spazio a prototipi nati più per stupire (in maniera a volte discutibile: è il caso di Project December e del lavoro sui 'deadbot', chatbot basati su intelligenze artificiali generative che si propongono di costruire un alter ego conversazionale di persone defunte¹⁴) che per far effettivamente avanzare la ricerca. Ho però il sospetto che il lavoro avviato sull'addestramento di IA generative basato su corpora personali avrà anche ricadute molto più interessanti, e che fra qualche anno questo capitolo del libro potrebbe essere riscritto in forma assai più estesa.

C'è un aspetto, però, che può essere sottolineato subito: come in parte già suggerito, le memorie personali digitali prodotte in questi anni, proprio per la loro natura inflattiva ed eterogenea, si prestano particolarmente bene a fornire un terreno per la sperimentazione di possibili forme di collaborazione fra l'architetto e l'oracolo. L'idea di poterle organizzare in maniera tradizionale, attraverso meccanismi di classificazione e metadatozione controllata, sembra infatti quasi impossibile, nonostante in molti casi alcune tipologie di metadati siano prodotte automaticamente insieme ai dati. Ma – come ipotizzato nella tesi T9 del capitolo 13 – potrebbe essere un sistema di intelligenza artificiale generativa, opportunamente addestrato, a occuparsi del compito di rendere fruibili corpora così estesi; non tanto e non solo con

l'obiettivo, in molti casi piuttosto inquietante, di costruire 'repliche digitali' di chi li ha prodotti: piuttosto, fornendo meccanismi conversazionali e amichevoli per poterli esplorare e per poter recuperare al loro interno le informazioni di volta in volta più utili. L'esternalizzazione sempre più massiccia della memoria, insomma, potrebbe essere accompagnata dalla creazione di agenti intelligenti capaci di aiutarci a navigare nei ricordi e nelle tracce che abbiamo accumulato.

Che anche questo richieda, ovviamente, cautele e attenzioni a vari livelli – da quello etico a quello della privacy e della protezione dei dati personali – è indubbio. E anche su queste tematiche ci sarà molto lavoro da fare in futuro.

¹ Si veda al riguardo Giuseppe Antonelli, *Il linguaggio degli SMS*, in *XXI secolo*, Istituto dell'Enciclopedia Italiana Giovanni Treccani, Roma 2009, online alla pagina https://www.treccani.it/enciclopedia/il-linguaggio-degli-sms_%28XXI-Secolo%29/.

² Tom Boellstorff, *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*, Princeton University Press, Princeton 2015; Tom Boellstorff, Bonnie Nardi, Celia Pearce e T.L. Taylor, *Ethnography and Virtual Worlds: A Handbook of Method*, Princeton University Press, Princeton 2012.

³ In italiano si veda ad es. Alice Avallone, *People watching in rete. Ricercare, osservare, descrivere con l'etnografia digitale*, Cesati, Firenze 2018.

⁴ TEDxUniTO 2016, <https://youtu.be/WYGDSjCkehI>.

⁵ Liadh Kelly, *The information retrieval challenge of human digital memories*, in BCS IRSG Symposium: Future Directions in Information Access 2007, Glasgow, Scotland, 28-29 agosto 2007, <https://doras.dcu.ie/344/>.

⁶ Per una ricca rassegna critica si veda Ricardo Ribeiro, Alina Trifan e António J.R. Neves, *Lifelog Retrieval From Daily Digital Data: Narrative Review*, in «JMIR mHealth and uHealth» 10, 5, e30517, 2 maggio 2022, <https://pubmed.ncbi.nlm.nih.gov/35499858/>.

⁷ Si veda ad es. Cathal Gurrin, Alan F. Smeaton e Aiden R. Doherty, *LifeLogging: Personal Big Data*, in «Foundations and Trends in Information Retrieval», 8, 1, 2014, pp 1-125, <http://dx.doi.org/10.1561/15000000033>.

⁸ Shaukat Ali, Shah Khusro, Akif Khan e Hayat Khan, *Smartphone-Based Lifelogging: Toward Realization of Personal Big Data*, in Teresa Guarda, Sajid Anwar, Marcelo Leon, Filipe Jorge Mota Pinto (a cura di), *Information and Knowledge in*

Internet of Things, Springer, Cham 2022, https://doi.org/10.1007/978-3-030-75123-4_12.

⁹ Cfr. Stefano Allegrezza, *Il problema dell'eredità digitale nella trasmissione di archivi e biblioteche personali*, in «Bibliothecae.it» 10, 1, 2021, pp. 352-400.

¹⁰ Mi limito a ricordare il principale testo di riferimento nell'ambito di un dibattito assai ampio, Franco Moretti, *Distant Reading*, Verso, London-New York 2013, trad. it. *A una certa distanza. Leggere i testi letterari nel nuovo millennio*, Carocci, Roma 2020, e, dello stesso autore, il recente *Falso movimento. La svolta quantitativa nello studio della letteratura*, Nottetempo, Roma 2022.

¹¹ Come già ricordato, l'espressione è stata proposta con questa connotazione da Derrick de Kerckhove, si veda ad es. Maria Pia Rossignaud e Derrick de Kerckhove, *Oltre Orwell. Il gemello digitale*, Castelvecchi, Roma 2020. Il termine 'digital twin' è stato però usato anche con connotazioni parzialmente diverse, cfr. ad es. Silvano Tagliagambe, *Metaverso e gemelli digitali. La nuova alleanza tra reti naturali e artificiali*, Mondadori, Milano 2022.

¹² Si vedano ad es. Maggi Savin-Baden e David Burden, *Digital Immortality and Virtual Humans*, in «Postdigital Science and Education» 1, 2019, pp. 87-103, <https://doi.org/10.1007/s42438-018-0007-6>, e Paul Smart, *Predicting Me: The Route to Digital Immortality?*, in Robert W. Clowes, Klaus Gärtner, Inês Hipólito (a cura di), *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts*, Springer, Cham 2022; tutto il volume è dedicato a tematiche rilevanti rispetto a quanto qui discusso.

¹³ L'idea di costruire chatbot di questo tipo è stata avanzata da diversi anni, e si può vedere al riguardo un curioso brevetto Google del 2018: Dustin Abramson e Joseph Johnson Jr, *Creating a conversational chat bot of a specific person*, Google Patents (US10853717B2), <https://patents.google.com/patent/US10853717B2/en?q=us10853717b2>.

¹⁴ Il sito del progetto è <https://projectdecember.net>, per una prima valutazione delle sue implicazioni cfr. Sara Suárez-Gonzalo, *'Deadbots' can speak for you after your death. Is that ethical?*, in «The Conversation», 9 maggio 2022, <https://theconversation.com/deadbots-can-speak-for-you-after-your-death-is-that-ethical-182076>, e – più esteso e dettagliato e con un'ampia bibliografia – Leah Henrickson, *Chatting with the dead: The hermeneutics of thanabots*, in «Media, Culture & Society» 45, 5, luglio 2023, pp. 949-966, <https://doi.org/10.1177/01634437221147626>.

Parte IV

18.

Fantascienza?

Le ultime tematiche affrontate nella parte terza possono sembrare – e in parte sono – al confine fra ricerca scientifica e fantascienza. L'evoluzione dell'ecosistema digitale ci ha insegnato però che questo confine è spesso labile, e sappiamo d'altro canto che la fantascienza può essere un modo non già per prevedere il futuro – cosa che, purtroppo o per fortuna, non siamo in grado di fare – ma per riflettere, attraverso la costruzione immaginaria e insieme speculativa¹ di un futuro possibile (o di molti futuri possibili), sulle direzioni verso cui ci stiamo incamminando nel presente.

Anche per questo, mi sembra utile aggiungere a questo lavoro una parte in cui esaminare alcune fra le immagini del futuro proposte dalla letteratura di fantascienza e più direttamente collegate all'idea di conservazione e organizzazione del sapere. Avevo pensato inizialmente di farlo all'interno di un'appendice, per segnalare anche strutturalmente la differenza fra analisi del presente e sconfinamento nell'immaginario. Mi sono accorto, però, che questa scelta sarebbe stata frutto, almeno in parte, di un timore: il timore che l'incursione fantascientifica potesse essere considerata come una sorta di vezzo, una divagazione legata magari a una passione personale (che in effetti c'è, ed è dichiarata), ma da tenere accuratamente separata rispetto al lavoro 'serio'.

La scelta di inserire le tematiche fantascientifiche all'interno del corpo del libro, senza relegarle a un'appendice, vuole rispondere a questo timore con una rivendicazione precisa: le pagine che seguono non vanno considerate come una pura evasione letteraria; sono – o vorrebbero essere – una riflessione non meno seria di quanto spero possa essere considerato il resto del libro. Certo, dato che lavorerò su testi letterari, l'oggetto della

riflessione sarà in parte diverso, così come lo sarà il metodo espositivo. Le ipotesi, le idee, le speculazioni che incontreremo non hanno alla base un fondamento rigoroso nei fatti: sono frutto, come si è detto, di immaginazione e di speculazione. Ma sono comunque il prodotto della stessa cultura che ha generato la rivoluzione digitale, e ci dicono qualcosa su come questa cultura presenta e rappresenta sé stessa, il proprio futuro, le proprie prospettive, le proprie paure.

Questo duplice volto della fantascienza – espressione del presente e speculazione sul futuro – non ne esaurisce la portata, e le discussioni sulla natura, le possibili definizioni, i motivi di interesse di un genere letterario così ampio e diversificato sono ovviamente molte². Non provo neanche a prenderle in considerazione in questa sede, limitandomi a ricordare un’osservazione di Adam Roberts, uno dei più attenti studiosi del settore: «Tutte le molteplici definizioni offerte dai critici sono state contraddette o modificate da altri critici, ed è sempre possibile individuare testi considerati consensualmente come fantascienza, che cadono però al di fuori delle definizioni usuali»³.

Nel contesto di questo libro mi interessa in primo luogo provare a capire se – e in quale misura – l’alternativa fra sistematizzazione organizzata del sapere e produzione quasi ‘organica’ di nuove conoscenze attraverso qualcosa di analogo al lavoro delle intelligenze artificiali generative sia stata prefigurata nella letteratura di fantascienza, in quali contesti e con quali esiti. Una domanda a cui inevitabilmente questa parte potrà rispondere solo in minima parte, attraverso l’analisi di pochi fra i molti, moltissimi testi che si potrebbero prendere in esame. Un’analisi attraverso esempi, necessariamente limitata, ma spero comunque interessante.

Proprio per questo motivo, un’ulteriore premessa da fare riguarda la scelta tanto dei testi quanto delle tematiche discusse. Anche se ci sarà inevitabilmente qualche riferimento ad altre opere, ho preferito concentrarmi sull’analisi di tre esempi specifici. Il primo è quasi obbligato: il ‘ciclo della fondazione’ di Isaac Asimov, oltre ad essere una delle pietre miliari della fantascienza contemporanea anche dal punto di vista della sua evoluzione storica, rappresenta infatti uno dei pochi casi in cui l’immaginazione di un futuro possibile (e assai lontano) si basa di fatto su due fra gli strumenti tradizionali di organizzazione ‘architettonica’ del

sapere: l'enciclopedia e la biblioteca. Ed è forse l'unico caso in cui la relazione fra un'enciclopedia e una biblioteca costituisce di fatto uno dei principali fulcri tematici della narrazione.

Il secondo esempio che prenderò in considerazione, il romanzo *Snow Crash* di Neal Stephenson, rappresenta anch'esso una scelta abbastanza naturale: fra i capolavori del genere, è considerato come una delle prime anticipazioni del metaverso, e ha fra i propri personaggi principali un'intelligenza artificiale che sembra essere una curiosa fusione di bibliotecario e di enciclopedista.

L'ultimo esempio analizzato, il breve romanzo *The Virtual Librarian* di Ted e Bob Rockwell, è assai meno noto e di qualità narrativa molto inferiore, ma la bibliotecaria virtuale che vi compare ha caratteristiche talmente simili a quelle di un'intelligenza artificiale generativa 'oracolare' di oggi, da meritare comunque attenzione in questo contesto.

Con l'eccezione – come vedremo, solo parziale – di Asimov, la fantascienza si è occupata assai più di biblioteche che di enciclopedie: non sorprenderà dunque che al centro di quest'ultima parte siano molto più le biblioteche che le enciclopedie. Come già accennato, però, proprio per il suo carattere di 'giacimento informativo' e di sterminato deposito di conoscenze, la biblioteca rappresenta un luogo di particolare interesse rispetto alla distinzione e alla dicotomia fra concezione architettonica e strutturata e concezione organico-oracolare del sapere. La sua organizzazione è in genere e tradizionalmente immaginata come fortemente sistematica, ma in realtà proprio l'esplosione di tipologie e quantità dei dati legata alla rivoluzione digitale mette fortemente in crisi questa immagine: all'inizio del XXI secolo (e, per la fantascienza, a volte anche prima) la biblioteca scivola verso quello che oggi considereremmo come il mondo dei big data. Questa evoluzione richiede nuovi strumenti di ricerca dell'informazione, e influenza la professionalità di bibliotecarie e bibliotecari, a cavallo fra la tradizione e la necessità di nuove competenze: competenze che in molte rivisitazioni fantascientifiche (e in prospettiva, almeno in alcuni casi, probabilmente anche nel mondo reale) sono a volte soddisfatte – come vedremo – attraverso il ricorso a intelligenze artificiali.

Ho già ricordato nell'Introduzione che la biblioteca, pur rappresentando uno dei punti di riferimento più naturali per l'idea di organizzazione

architettonica del sapere, può anche essere concepita – seguendo l'autorevole esempio di Ranganathan – come un organismo in continua evoluzione. L'attenzione alla sua presenza nell'immaginario letterario (e più recentemente anche in quello cinematografico e televisivo) è andata indubbiamente crescendo negli ultimi anni, ma che la biblioteca sia anche *topos* letterario è cosa ben nota da tempo: fra gli esempi novecenteschi più noti basterà ricordare i nomi di Borges e di Eco, e sarebbe facile aggiungerne moltissimi altri. Dal nostro punto di vista va sicuramente ricordata, ad esempio, la biblioteca vivente che conclude *Fahrenheit 451* di Ray Bradbury: è interessante notare come in tale caso specifico l'immagine della biblioteca, pur collocata all'interno di un testo fantascientifico, sia per molti versi più organica e oracolare – proprio per lo strettissimo collegamento con l'oralità stabilito dalla conclusione del libro – che architettonica e sistematica. Difficile sfuggire alla tentazione di segnalare, al riguardo, le ultime e straordinarie scene del film che dal libro ha tratto François Truffaut nel 1966: se non le conoscete, il QR-Code qui di seguito rimanda a uno dei tanti video che le propongono su YouTube.



QR-Code 22

Le scene finali del film *Fahrenheit 451* di François Truffaut

<http://bit.ly/3NNKkxB>

Ma ovviamente nel parlare del rapporto fra biblioteche e letteratura si potrebbe risalire molto più indietro nel tempo: volendo, fino a vedere – come a volte è stato fatto – nell'albero della conoscenza del bene e del male la metafora di una biblioteca, e nella mela offerta dal serpente ad Eva la metafora di un libro⁴.

Nelle opere letterarie troviamo così biblioteche reali e fantastiche,

biblioteche passate e future. E il tema non poteva non affascinare gli stessi bibliotecari, che hanno esplorato l'ampio territorio delle biblioteche immaginarie in numerose occasioni e in varie forme⁵. Michele Santoro ha usato per quest'ambito di ricerca l'espressione «biblioteconomia letteraria»⁶, osservando che esso «non solo individua e definisce il trattamento riservato dagli scrittori al mondo dei libri e delle biblioteche, ma [...] spinge l'indagine verso problematiche schiettamente bibliotecarie, pervenendo ad una definizione sub specie letteraria delle procedure e dei metodi che caratterizzano l'universo biblioteconomico»⁷.

L'idea che l'analisi di biblioteche in tutto o in parte immaginarie possa dirci qualcosa non solo sulla biblioteca come *topos* letterario ma anche sulle biblioteche reali, sul modo in cui il loro ruolo è percepito all'interno del più ampio contesto sociale, sui loro servizi, sui loro utenti, sulle metodologie di gestione delle collezioni e dell'informazione, su problemi e prospettive del loro sviluppo, suggerisce che, fra i vari ambiti e generi letterari in cui il tema è affrontato, possano avere un interesse particolare proprio le narrazioni che guardano al futuro, nelle loro diverse forme e varianti: fantascienza classica, narrazioni di anticipazione, utopie, distopie, ucronie, cyberpunk... E anche questa considerazione spiega il particolare taglio di quest'ultima parte del libro⁸.

Come per altre parti di questo testo, anche i tre capitoli di questa parte riprendono – in forma modificata e integrata – alcuni miei lavori precedenti: in particolare, il capitolo 19 si basa sul mio contributo a una miscellanea di studi dedicata ad Alberto Petrucciani⁹, e i capitoli 20 e 21 riprendono in parte il mio contributo a un volume su biblioteche e fantascienza in corso di pubblicazione presso la casa editrice Bibliografica¹⁰. Quest'ultimo volume, per il suo fuoco tematico e per gli interventi che contiene – che allargano notevolmente, rispetto a quanto possibile per me e in questa sede, l'orizzonte delle opere prese in considerazione (non solo testi ma anche film, serie televisive, fumetti...) – è assai pertinente rispetto alle questioni che vorrei qui affrontare, e nonostante l'evidente conflitto di interessi mi sento di suggerirne senz'altro la lettura a chi trovasse interessanti le tematiche affrontate nei prossimi capitoli.

¹ Non entro qui nella complessa e *vexata quaestio* delle definizioni e dei confini dei generi letterari legati ai vasti territori dell'immaginario speculativo, se non per osservare che in questo contesto l'etichetta 'speculative fiction' è probabilmente quella più rilevante, se la si assume nel suo significato generale di allontanamento dalla 'consensus reality' per abbracciare tutte le situazioni di costruzione di mondi possibili narrativi basate su ipotesi controfattuali e/o di anticipazione (uno fra i tre significati individuati in Marek Oziewicz, *Speculative Fiction*, in *Oxford Research Encyclopedia of Literature*, 2017, <https://oxfordre.com/literature/literature/view/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-78>; sulla definizione di 'speculative fiction' si veda anche R.B. Gill, *The Uses of Genre and the Classification of Speculative Fiction*, in «Mosaic: An Interdisciplinary Critical Journal» 46, 2, giugno 2013, pp. 71-85, <https://www.jstor.org/stable/44030329>). Rispetto a tale significato generale, la fantascienza risulta un sottogenere e caratterizza la 'speculative fiction' in cui l'allontanamento dalla 'consensus reality' avviene attraverso ipotesi di natura prevalentemente scientifico/tecnologica e/o legate all'immaginazione di futuri possibili. Per una rapida discussione delle diverse possibili definizioni di fantascienza, con i principali riferimenti bibliografici, rinvio a Gino Roncaglia, *Frankenstein and Science Fiction*, in Francesca Saggini e Anna Enrichetta Soccio (a cura di), *Transmedia Creatures. Frankenstein's Afterlives*, Bucknell University Press, Lewisburg (PA) 2018, pp. 33-49, <https://doi.org/10.36019/9781684480647-003>.

² Per un quadro d'insieme si veda Adam Roberts, *The History of Science Fiction*, seconda edizione, Palgrave Macmillan, Basingstoke-New York 2016.

³ Adam Roberts, *Science Fiction*, seconda edizione, Routledge, London 2006, p. 1.

⁴ È ben noto il passo della commedia *The Rivals* di Richard Sheridan (1751-1816) in cui la biblioteca circolante è paragonata a un «tree of diabolical knowledge», fonte di irresistibile tentazione per le figlie di Eva: «They who are so fond of handling the leaves, will long for the fruit at last». Cfr. Devendra P. Varma, *The Evergreen Tree of Diabolic Knowledge*, Consortium Press, Washington 1972 e Christiane Inmann, *Forbidden Fruit: A History of Women and Books in Art*, Prestel Books, New York 2009. La metafora torna anche nelle nuove forme di testualità post-moderna: nella serie televisiva *The Librarians*, in cui giovani bibliotecari hanno il compito di proteggere e custodire all'interno di una biblioteca semi-fantastica alcuni reperti magici o leggendari, al centro della biblioteca è collocato l'albero del bene e del male proveniente dal paradiso terrestre, mentre in uno dei più noti manga giapponesi, la serie *One Piece*, il «tree of knowledge» è un albero plurimillenario che funziona da biblioteca e luogo di studio, e che – proprio come la biblioteca di Alessandria – finirà per essere distrutto dal fuoco.

⁵ Per limitarci all'Italia, ricordiamo in particolare il volume di Renato Nisticò *La biblioteca*, Laterza, Roma-Bari 1999 (edizione digitale, 2015), la raccolta di saggi a

cura di Rossana Morriello e Michele Santoro, *La biblioteca e l'immaginario*, Editrice Bibliografica, Milano 2004, il progetto Librariana dell'AIB, all'indirizzo <https://www.aib.it/aib/clm/clm.htm>, che comprende sezioni su biblioteche e bibliotecari nella letteratura italiana e nella letteratura straniera, oltre a un'ampia bibliografia (l'aggiornamento delle pagine del progetto è purtroppo fermo al 2008) e, tra i contributi più recenti, il volume di Rossana Morriello *La biblioteca narrata*, Editrice Bibliografica, Milano 2020, che raccoglie e aggiorna alcuni contributi sulla rappresentazione della biblioteca nella letteratura e nel cinema che l'autrice aveva pubblicato nella rubrica 'Artifici' su «Biblioteche Oggi».

⁶ Michele Santoro, *Ecpirosi apocrife. Per una più rigorosa definizione di biblioteconomia letteraria*, in «Biblioteche Oggi» 18, 10, dicembre 2000, pp. 36-44.

⁷ Ivi, p. 36.

⁸ Sul tema del ruolo delle biblioteche e dei bibliotecari nella letteratura fantascientifica si vedano in particolare: Agnes M. Griffen, *Images of Libraries in Science Fiction*, in «Library Journal» 112, 1° settembre 1987, pp. 137-142; James K. Gunn, *Dreams Written Out: Libraries in Science Fiction*, in «Wilson Library Bulletin» 69, febbraio 1995, pp. 26-29; Marcia J. Myers, *Images of Librarians in Science Fiction and Fantasy: Including an Annotated List*, Research paper, Eastern Kentucky University, Richmond 1998, <https://files.eric.ed.gov/fulltext/ED420314.pdf>; Bruce DuBoff, *The past in the future: a content analysis applying S. R. Ranganathan's five laws of library science to libraries and information services in young adult science fiction*, Theses and dissertations, Rowan University, Glassboro 2003, <http://rdw.rowan.edu/cgi/viewcontent.cgi?article=2291&context=etd>.

L'importante, recente lavoro di Amy Duxfield e Chern Li Liew, *Libraries in contemporary science fiction novels: uncertain futures or embedded in the fabric of society?*, in «Journal of Documentation», in corso di stampa, online da agosto 2022, <https://doi.org/10.1108/JD-05-2022-0097>, analizza la presenza di biblioteche e bibliotecari in una selezione di romanzi di fantascienza scelti fra i vincitori dei principali premi letterari del settore fra il 2010 e il 2020, e offre ulteriori riferimenti bibliografici utili. Per maggiori indicazioni rimando poi a Federico Meschini, Rossana Morriello e Gino Roncaglia (a cura di), *Le biblioteche della fantascienza*, Editrice Bibliografica, Milano, in corso di pubblicazione.

⁹ Gino Roncaglia, *La biblioteca e l'impero*, in Simonetta Buttò, Vittorio Ponzani e Simona Turbanti (a cura di), *L'arte della ricerca: fonti, libri, biblioteche. Studi offerti ad Alberto Petrucciani per i suoi 65 anni*, Associazione Italiana Biblioteche, Roma 2021, pp. 127-137.

¹⁰ Federico Meschini, Rossana Morriello e Gino Roncaglia (a cura di), *Le biblioteche della fantascienza* cit.

19.

La biblioteca e l'impero

In questo capitolo mi soffermerò su uno specifico mondo possibile letterario, di ambito strettamente fantascientifico: quello legato al cosiddetto 'ciclo della fondazione' di Isaac Asimov¹, un corpus abbastanza ampio che – almeno per quanto riguarda la trilogia originaria – costituisce uno degli esempi più noti di fantascienza classica. Elementi di specifico interesse di questo ciclo, dal nostro punto di vista, sono i ruoli assolutamente centrali che vi assumono un'enciclopedia, l'Enciclopedia Galattica, e una biblioteca, la biblioteca di Trantor², il pianeta-città che, nell'universo narrativo immaginato da Asimov, costituisce la capitale del primo impero galattico.

Se volete avere qualche informazione in più su Isaac Asimov, il video raggiungibile a partire dal QR-Code qui di seguito è la registrazione di una conferenza – co-organizzata dall'Agenzia Spaziale Italiana e dal Forum del Libro – che ho tenuto sulla sua figura in occasione del centesimo anniversario dalla nascita.



QR-Code 23

Isaac Asimov fra scienza, fantascienza e filosofia

<http://bit.ly/43kUM5q>

Per capire il rilievo che Asimov attribuisce tanto all'Enciclopedia Galattica quanto alla biblioteca di Trantor, è indispensabile una rapidissima sintesi almeno della principale linea narrativa del ciclo. Sintesi che inevitabilmente contiene degli spoiler: chi non avesse letto la trilogia originale e non si volesse rovinare il piacere di scoprire il mistero della natura e della sede della seconda fondazione – mistero attorno a cui ruota molta parte della storia – dovrebbe tornare a questo capitolo solo dopo tale lettura, che personalmente consiglio senz'altro.

Dichiaratamente ispirato alla classica *History of the Decline and Fall of the Roman Empire*, scritta nella seconda metà del XVIII secolo dallo storico inglese Edward Gibbon, il 'ciclo della fondazione' è ambientato sullo sfondo del declino e della successiva caduta del primo impero galattico, che all'inizio della narrazione unisce migliaia di mondi e ha in Trantor la sua capitale. Asimov immagina che il progressivo declino dell'impero venga previsto da Hari Seldon, un matematico al quale si deve la scoperta della psicostoria, disciplina scientifica che attraverso l'analisi matematica di grandi quantità di dati storici e statistici riesce a prevedere con buona precisione l'evoluzione futura di un sistema sociale complesso. I calcoli di Seldon suggeriscono non solo che l'impero sia avviato al declino, ma anche che alla sua caduta seguirà un periodo di trentamila anni di anarchia, guerre e barbarie, prima della costituzione di un secondo impero galattico.

Quando Seldon elabora la sua previsione, le spinte geopolitiche e sociali che stanno portando al declino dell'impero non sono più arrestabili; con l'aiuto della psicostoria è possibile però indirizzare lo sviluppo futuro in modo da limitare a mille gli anni di interregno e di barbarie. Per farlo, Seldon avvia la creazione di due fondazioni, collocate «agli angoli opposti dell'impero». La prima fondazione, nel pianeta periferico Terminus, dovrà costituire il primo nucleo del secondo impero, sviluppandosi progressivamente come potenza economica, politica e militare, e non saprà nulla dell'esistenza e della localizzazione della seconda fondazione né – se non in termini estremamente generali – delle leggi della psicostoria, per evitare che tali conoscenze possano interferire sullo sviluppo previsto da Seldon. La seconda fondazione sarà invece composta da psicostorici e sorveglierà lo sviluppo nel tempo del progetto, introducendo le correzioni che si rivelassero eventualmente necessarie. La

collocazione esatta della seconda fondazione è tenuta segreta, e costituisce uno dei principali enigmi attorno a cui ruota la narrazione della prima trilogia.

Prima e seconda fondazione sono dunque i veri e propri motori narrativi della storia, che nei tre volumi (e in quelli che poi li integreranno) copre periodi diversi e vede l'avvicinarsi di personaggi diversi. E il particolare interesse che la complessa costruzione di Asimov presenta dal nostro punto di vista sarà subito chiaro se si tiene presente che la prima fondazione, quella periferica, nasce con l'obiettivo dichiarato di realizzare e pubblicare un'enciclopedia, e la seconda fondazione – che invece, si scoprirà, è collocata proprio a Trantor, dal punto di vista psicostorico «l'altro estremo» rispetto alla periferia galattica – coincide con la biblioteca imperiale.

Certo, l'enciclopedia galattica e la biblioteca imperiale di Trantor sono delle 'coperture' che nascondono i veri fini rispettivamente della prima e della seconda fondazione all'interno del progetto complessivo elaborato da Seldon. Ma non sono coperture qualsiasi: la loro natura e le loro caratteristiche 'istituzionali' sono indispensabili per il buon fine del progetto stesso. Tanto l'enciclopedia quanto la biblioteca sono infatti strumenti per garantire la trasmissione nel tempo delle conoscenze sviluppate nel primo impero, la cui perdita allungherebbe inevitabilmente il periodo di barbarie.

Così, Seldon non sta affatto mentendo – anche se nasconde la parte principale della verità, il ruolo geopolitico che intende affidare alla prima fondazione – quando all'inizio di *Foundation*, sotto processo perché la sua previsione sull'inevitabile caduta dell'impero è tacciata di disfattismo, afferma che il suo progetto è quello di limitare la durata dell'interregno salvando le conoscenze dell'umanità:

Con la distruzione della nostra costruzione sociale, la scienza verrà spezzettata in milioni di parti. Gli individui conosceranno poco meno che una sfaccettatura di tutto ciò che c'è da sapere. Da soli saranno indifesi ed inutili. Tali frammenti insignificanti di conoscenza non saranno trasmessi e si disperderanno attraverso le generazioni. Se però prepariamo un gigantesco sommario di tutto il sapere, esso non andrà mai perduto. Le generazioni successive costruiranno sopra queste basi senza doverle riscoprire. Un millennio farà il lavoro di trentamila anni [...] Questo è tutto il mio progetto. I miei trentamila uomini, con le loro mogli e bambini³, si sono dedicati interamente alla

preparazione di una 'Enciclopedia Galattica'. Non riusciranno a completarla nel tempo concesso loro dalla vita. Non vivranno abbastanza a lungo nemmeno per vederla cominciata. Ma quando Trantor cadrà, l'opera sarà completa e ne esisteranno copie in ogni biblioteca della Galassia⁴.

La creazione dell'Enciclopedia Galattica non è in realtà lo scopo principale della prima fondazione – che è invece quello di trasformare progressivamente Terminus nel primo nucleo del secondo impero – ma è una condizione necessaria perché questo possa avvenire: solo gli enciclopedisti, depositari dell'edificio architettonico del sapere, hanno infatti le conoscenze necessarie per mantenere almeno in parte vive la scienza e la tecnologia del primo impero, che declineranno altrove.

D'altro canto, anche per la seconda fondazione la collocazione 'bibliotecaria' non è solo una copertura ma una necessità: la biblioteca è infatti il luogo in cui Seldon ha lavorato allo sviluppo della psicostoria, e in cui le informazioni e i dati da lui raccolti – e più in generale le informazioni e i dati necessari al lavoro degli psicostorici – sono conservati.

Hari Seldon ha lasciato dietro di sé la Seconda Fondazione per conservare, migliorare ed estendere il suo lavoro. Questo lo sappiamo, o l'abbiamo indovinato, da cinquant'anni. Ma dov'è che lo si poteva fare al meglio? A Trantor, dove il gruppo di Seldon ha lavorato, e dove i dati di secoli sono stati accumulati⁵.

La narrazione mostrerà inoltre come il progetto dell'Enciclopedia dipenda dalla biblioteca: sono i bibliotecari, come vedremo, a individuare la sede in cui lavoreranno gli enciclopedisti, e a fornire la preparazione e i materiali necessari allo sviluppo del loro lavoro.

Nel grandioso affresco di storia futura immaginato da Asimov, insomma, il ruolo fondamentale nell'evitare la ricaduta completa dell'impero galattico nella barbarie e nel limitare la durata del periodo di disgregazione e declino è affidato in larga misura alla dialettica fra due istituzioni culturali diverse ma complementari, tutte e due fondamentali per organizzare e conservare l'architettura del sapere: l'enciclopedia e la biblioteca. Il ruolo dell'enciclopedia e degli enciclopedisti è più visibile, mentre quello della biblioteca e dei bibliotecari è in parte nell'ombra, ma quest'ultimo è il più importante, complesso e consapevole: in un certo senso, la biblioteca è condizione di possibilità per l'enciclopedia.

Ma Asimov non si limita a dare alla biblioteca di Trantor il ruolo fondamentale di sovrintendere allo sviluppo del progetto Seldon durante il millennio di interregno: in più passi, egli ne descrive le collezioni, il funzionamento, i servizi, i problemi, la storia.

Dall'incrocio di diversi passi si ricava così una messe di informazioni sulla biblioteconomia futura, come è immaginata in questo ciclo narrativo. Provo sinteticamente a riassumerne le principali. La biblioteca ospita informazioni su una varietà di supporti diversi: libri, nastri, video, mappe (comprese mappe olografiche tridimensionali), libri-film e libri olografici (da leggere attraverso appositi proiettori), dischi ottici, wafer di memoria e altre tipologie di supporti digitali. È aperta a tutti (e l'apertura a ogni cittadino dell'impero costituisce una sua caratteristica essenziale), ma dispone di studi privati riservati alle ricercatrici e ai ricercatori più qualificati. I suoi immensi corridoi si possono percorrere utilizzando dei piccoli veicoli personali ('skitters'). Trattandosi di una biblioteca di conservazione, tende a non operare scarti, a differenza delle biblioteche universitarie. Deve gestire l'obsolescenza: anche nel futuro l'informazione su supporto digitale tende a degradare nel tempo, a meno di non duplicarla di frequente (Asimov ha molto chiaro il ruolo della produzione organizzata di ridondanza come strategia per favorire la conservazione nel tempo dei dati), cosa che però avviene di fatto solo per i contenuti più frequentemente usati⁶. Organizza corsi sull'uso dei propri servizi e sulle metodologie per la ricerca delle informazioni, e questi corsi, destinati alle matricole dell'università, prevedono anche competizioni fra studenti relative alla rapidità e alla precisione della ricerca bibliografica⁷. I bibliotecari rappresentano «la più antica Gilde dell'Impero, quella con la tradizione più riverita»⁸. La biblioteca difende strenuamente la sua indipendenza dal potere politico e perfino dall'imperatore, ed è guidata da un consiglio di bibliotecari presieduto dal bibliotecario capo, personaggio di ruolo e autorità indiscussa:

Non era cosa facile organizzare un incontro personale con il Bibliotecario Capo della Biblioteca Galattica. L'idea che egli stesso aveva della natura e del valore del suo incarico era piuttosto alta, e si diceva frequentemente che quando l'Imperatore stesso voleva consultare il Bibliotecario Capo, doveva visitare la Biblioteca e aspettare il suo turno⁹.

È in particolare nei quattro volumi aggiunti alla trilogia iniziale che la

biblioteca diviene un luogo centrale della narrazione. Asimov racconta così dei problemi che la biblioteca deve affrontare con i primi tagli di fondi legati al deteriorarsi della situazione dell'impero, descrivendo il contrasto, all'interno del consiglio dei bibliotecari, fra chi teorizza la necessità di fronteggiare la riduzione di risorse chiudendo la biblioteca al pubblico esterno e limitando i servizi di reference per concentrarsi sul compito di conservazione, e chi considera invece essenziale garantire sempre e comunque libertà di accesso e consultazione. Quando il bibliotecario capo gli descrive questo contrasto, Seldon osserva che «tutta questa faccenda di chiudere la Biblioteca, di renderla meno accessibile, di rifiutare di dare le informazioni... l'intera questione degli stanziamenti in declino... tutto questo è un segno della disintegrazione imperiale»¹⁰. La situazione della biblioteca è vista qui come una cartina di tornasole della situazione complessiva della società: una società che riduce i fondi all'istituzione bibliotecaria è una società che tende al declino.

Hari Seldon cercherà di convincere l'imperatore ad aumentare i finanziamenti alla biblioteca, dove vuole far lavorare il gruppo di psicostorici che porterà avanti il progetto delle due fondazioni e che solo lì può trovare i dati e le informazioni di cui ha bisogno. Il dialogo di Seldon con l'imperatore suona curiosamente familiare a chiunque si sia trovato a negoziare finanziamenti per un'istituzione bibliotecaria: «Dove volete che io possa trovare fondi per la Biblioteca? Dovrebbero essermi grati se ogni anno, nonostante tutte queste difficoltà, riesco ancora a spremere qualcosa per loro»¹¹, è la prima risposta dell'imperatore alle richieste di Seldon.

Difesa e protetta dai bibliotecari (che costituiscono anche il piccolo nucleo di psicostorici responsabili della seconda fondazione) al momento dell'inevitabile caduta di Trantor, nel corso del millennio di interregno la biblioteca diventerà poi insieme un deposito e un simbolo della perduta grandezza del primo impero, e come tale viene percepita da Janov Pelorat, uno storico che lavora durante quel periodo:

Naturalmente a Pelorat interessava Trantor per via della Biblioteca Galattica, che ai tempi dell'Impero (quando si chiamava Biblioteca Imperiale) era stata la più grande della Galassia. Trantor era stata la capitale dell'impero più vasto e popoloso che l'umanità avesse mai conosciuto. I suoi abitanti superavano di parecchio i quaranta miliardi, e la Biblioteca conteneva tutte le opere creative (ed anche un po' meno creative)

dell'umanità, l'intero compendio delle sue conoscenze. Ed era computerizzata in modo talmente complesso, che occorrevano persone esperte per poterla consultare¹².

La biblioteca di Trantor rappresenta anche il modello per le biblioteche di altri pianeti e per la progressiva ricostruzione di un servizio bibliotecario, con tanto di prestito interbibliotecario interplanetario. Lo stesso Pelorat, che lavora a Terminus e dunque «all'altro estremo» della galassia, utilizza regolarmente questo servizio, alla ricerca di informazioni sulle leggende riguardanti il periodo precedente la fondazione dell'impero:

[...] era andato al suo terminale di computer e aveva controllato se la Biblioteca Universitaria di Terminus avesse materiale che riguardasse quel tipo di leggende. Proprio di quel tipo di leggende si era occupato da allora. La Biblioteca Universitaria di Terminus era ben poco fornita riguardo a quell'argomento, ma quando era diventato più grande, Pelorat aveva scoperto le gioie del prestito interbibliotecario. Aveva in suo possesso tabulati ottenuti, tramite segnali iper-radiazionali, da mondi lontani come Ifnia¹³.

Come si sarà capito già da questo passo, oltre alla biblioteca imperiale di Trantor Asimov introduce nel corso del ciclo diverse altre biblioteche, spesso associate alla ricerca di informazioni sulle leggende pre-imperiali relative al ruolo della Terra come pianeta di origine dell'umanità. Così, ad esempio, la biblioteca del pianeta Melpomenia ospita primitivi libri-film e i relativi dispositivi di lettura: «Si trattava di strutture cubiche tozze, sormontate da uno schermo inclinato e da un prolungamento curvo sulla sommità... forse una specie di bracciolo, o forse un punto dove appoggiare un elettrotaccuino [nell'originale inglese, 'electro-notepad']»¹⁴.

Le citazioni potrebbero proseguire, e allargarsi sia ai contributi di altri autori al ciclo, sia ad altre opere di Asimov. Ma quanto abbiamo visto fin qui dovrebbe essere sufficiente per un primo bilancio. Possiamo innanzitutto osservare che Asimov non sembra principalmente interessato a immaginare una biblioteca futura o futuribile, dotata di tecnologie radicalmente innovative: pur proiettando la sua storia migliaia di anni nel futuro, la natura delle biblioteche, la tipologia dei servizi offerti, lo stesso ruolo dei bibliotecari risultano non solo familiari, ma semmai per alcuni versi già invecchiati. Certo, ci sono dispositivi di lettura olografici e libri-

film (qualunque cosa siano), il prestito interbibliotecario viaggia di stella in stella attraverso segnali «iper-radiazionali» non meglio specificati, ma tutto sommato una bibliotecaria o un bibliotecario di oggi avrebbero bisogno solo di un veloce aggiornamento tecnologico per lavorare nella biblioteca imperiale di Trantor, e potrebbero forse dare alcuni buoni consigli per allargarne e svecchiarne la missione e i servizi. I bibliotecari non sono sostituiti da intelligenze artificiali, non ci sono agenti software che fanno ricerche al posto degli utenti, non ci sono oggetti informativi radicalmente nuovi. Nonostante la fantascienza classica sia normalmente associata all'esplorazione di nuove frontiere scientifico-tecnologiche, è chiaro che l'intento di Asimov è più quello di analizzare il possibile ruolo sociale e politico della biblioteca intesa – classicamente – come deposito di conoscenze, e di introdurre tale ruolo come componente della sua costruzione narrativa, che quello di prevedere come saranno le biblioteche del futuro.

Le sue idee più interessanti – nel parlare di biblioteche – sono, mi sembra, proprio quelle associate a tale ruolo: l'idea secondo cui l'informazione è potere, enfatizzata dalle potenzialità della psicostoria, porta con sé, nell'universo della fondazione, l'idea che i luoghi di gestione dell'informazione acquistino una loro specifica centralità politica: la biblioteca imperiale di Trantor si trasforma dunque in uno dei luoghi-cardine della geopolitica, addirittura su scala galattica. La biblioteca è vista così non solo come (ultimo) baluardo della civiltà galattica in disgregazione, ma anche come strumento indispensabile nel favorirne la rinascita.

Nella sua autobiografia, Asimov parla diffusamente del ruolo fondamentale che ha avuto per lui la biblioteca:

Mio padre mi procurò una tessera della biblioteca, e mia madre mi ci accompagnava frequentemente. La prima volta in assoluto che mi è stato permesso di andare da qualche parte da solo, è stato in biblioteca, quando mia madre si stancò di portarmici continuamente. [...] Per un bambino povero, la cui famiglia non poteva permettersi di comprare libri, la biblioteca era una porta aperta sulle meraviglie e sulla possibilità di realizzarsi, e non potrò mai essere abbastanza grato per il fatto di aver avuto abbastanza intuito da slanciarmi attraverso quella porta e ricavarne tutto quel che potevo.

Ora, quando mi capita di leggere (e mi capita spesso) di come i fondi per le biblioteche vengano tagliati e poi ancora tagliati, posso solo pensare che quella porta si sta chiudendo e che la società americana ha trovato un nuovo modo per distruggersi¹⁵.

La rivendicazione del ruolo politico della biblioteca all'interno dell'universo della fondazione è perfettamente coerente con questa considerazione, e nasce indubbiamente anche dal debito personale verso l'istituzione bibliotecaria. La centralità delle biblioteche all'interno della sua opera riflette insomma, programmaticamente, la centralità del ruolo che Asimov attribuisce loro anche nella società. Non stupiranno allora due particolari che può essere curioso ricordare in questa sede. Da un lato, il fatto che un riferimento al mondo bibliotecario sia presente perfino nelle poche righe di autopresentazione poste nei risvolti di copertina delle ultime opere pubblicate in vita, in cui Asimov (noto per non essere particolarmente modesto) sottolinea di aver scritto libri in tutte le categorie della classificazione decimale Dewey¹⁶. Dall'altro, il fatto che la biblioteca imperiale di Trantor abbia avuto una recente incarnazione anche nel darknet, la porzione anonima e in parte clandestina della rete, sotto forma di uno fra i più vasti depositi di testi pirata esistenti online. Il fondatore del sito, che si nasconde dietro il nome di Las Zenow (nel 'ciclo della fondazione', il bibliotecario capo amico di Hari Seldon), ha scelto di far riferimento al lavoro di Asimov come omaggio a uno dei suoi autori preferiti, ma anche per indicare il carattere programmaticamente universale della sua biblioteca pirata¹⁷. Asimov probabilmente non avrebbe approvato, ma anche se in forma non molto condivisibile si tratta di un'ulteriore testimonianza del carattere più 'politico' che futuribile della biblioteca da lui immaginata.

E l'Enciclopedia Galattica? Abbiamo già accennato al fatto che il suo destino è strettamente collegato a quello della biblioteca, che ne rappresenta un po' la culla. E al fatto che in parte si tratta di una copertura per il vero scopo della prima fondazione: lo stesso Hari Seldon, nella registrazione video in cui svela ai suoi discendenti l'inganno, afferma recisamente: «Tanto a me che ai miei colleghi non importa niente che venga pubblicato o meno anche un solo volume dell'Enciclopedia»¹⁸. Ma l'Enciclopedia conserva un'importanza notevole: le continue citazioni proposte da Asimov mostrano che le è effettivamente e comunque affidato il ruolo di 'cattedrale della conoscenza', deposito organizzato di tutto il sapere del passato. Moltissimi fra i capitoli dei romanzi del ciclo iniziano

con citazioni di voci dell'Enciclopedia, accompagnati dalla relativa (e ovviamente fittizia) nota bibliografica:

Tutte le voci qui riportate sono tratte – per gentile concessione dell'editore – dall'*Enciclopedia Galattica*, CXVI edizione, pubblicata nel 1020 E.F. dagli Editori Enciclopedia Galattica, Terminus (*N.d.A.*)¹⁹.

Tuttavia, progressivamente, gli abitanti di Trantor che *non* fanno parte del ristretto gruppo di circa centocinquantamila enciclopedisti, completamente assorbiti dalla loro impresa, si rendono conto che la classificazione sistematica del sapere del passato non basta:

Noi sediamo qui, considerando l'Enciclopedia il non plus ultra. Pensiamo che il massimo scopo della scienza sia la classificazione delle scoperte passate. È importante, lo riconosco, ma non si può andare più in là?²⁰

Progressivamente, da strumento di organizzazione del sapere l'Enciclopedia Galattica si è trasformata in una sorta di monumento storico alla grandezza del passato, incapace di favorire la creazione di nuove conoscenze. Anche da questo punto di vista, Asimov considera chiaramente la biblioteca come più importante dell'enciclopedia: l'enciclopedia sistematizza le conoscenze, la biblioteca contribuisce a farle nascere.

Curiosamente, pur trattandosi di un lavoro proiettato in un futuro ipertecnologico, la progressiva costruzione della prima Enciclopedia Galattica da parte degli enciclopedisti originari non sembra prevedere neanche aggiornamenti: è, di fatto, un compendio di archeologia del sapere, e solo quando la fase peggiore di decadenza dell'impero galattico è superata il progetto originario è sostituito da un nuovo progetto enciclopedico, questa volta costantemente aggiornato. Lo si desume da un dialogo contenuto in *Foundation's Edge*, volume che – come già accennato – è ambientato dopo la trilogia originaria:

Ma a cosa assomigliava questo posto cinquecento anni fa, quando fu fondato? Era una città, una piccola città abitata da un gruppo di uomini che preparavano un'Enciclopedia che non fu mai finita!

– Ma sì che fu finita.

– Tu ti riferisci all'attuale *Enciclopedia Galattica*. Quella non è l'Enciclopedia alla quale lavoravano loro: questa si trova in un computer e viene aggiornata quotidianamente. Hai mai dato un'occhiata all'originale incompleto?

– Intendi quello del Museo Hardin?²¹

L'enciclopedia intesa solo come summa storica è insomma andata in pensione (e la sua pensione è un museo), il vero enciclopedismo non si limita a sistematizzare le conoscenze del passato, ma ne segue da vicino l'evoluzione e diventa esplicitamente – per la prima volta nel ciclo – una enciclopedia elettronica: curiosamente, Asimov colloca dunque in un futuro lontanissimo un aspetto dell'enciclopedismo digitale – la sua possibilità di aggiornamento continuo – che come abbiamo visto nella parte prima di questo libro è stato in realtà esplorato abbastanza presto. Considerate le date di scrittura dei libri (*Foundation's Edge* è del 1982), ci accorgiamo che anche in questo caso Asimov in realtà proietta sul futuro il suo presente: negli anni '40 del secolo scorso, all'epoca della stesura dei racconti alla base della trilogia originale, non esistevano enciclopedie elettroniche, e l'Enciclopedia Galattica, nonostante la sua collocazione interplanetaria, è pensata come un'enciclopedia del tutto tradizionale, strumento statico di organizzazione delle conoscenze del passato. All'inizio degli anni '80 le enciclopedie elettroniche ci sono, anche se ancora abbastanza embrionali, e Asimov le introduce anche – più o meno con le stesse caratteristiche – nel... futuro del suo futuro.

E l'oracolo? Si potrebbe essere tentati di pensare che nel complesso affresco futuro disegnato da Asimov non ve ne sia traccia, ma in realtà non è affatto così. Proviamo a leggere la sintesi che uno dei più noti critici italiani in ambito fantascientifico, Giuseppe Lippi, fornisce della psicostoria, la scienza che abbiamo visto essere alla base delle previsioni di Hari Seldon, nella sua introduzione all'edizione Mondadori della trilogia originaria:

[...] questa scienza immaginaria non è, al contrario di quel che si potrebbe pensare, la parente prossima di un fosco determinismo, ma si basa in sostanza sulle leggi della statistica. Se il comportamento del singolo è imprevedibile, dice Asimov (o meglio Hari Seldon), il comportamento di grandi masse di individui è statisticamente inquadrabile in una serie di tendenze. Quanto più vasta è la porzione di umanità presa in esame, tanto più precise risulteranno le 'previsioni' dei suoi movimenti di massima²².

In tutta la narrazione, Asimov esprime le previsioni della psicostoria in termini strettamente probabilistici, associandole a valori numerici. E in

Foundation and Earth, del 1986, Asimov spiega anche qualcosa sulla possibile origine di questa idea, dal punto di vista di Seldon:

Stando al racconto tradizionale, Hari Seldon creò la Psicostoria prendendo come modello la teoria cinetica dei gas. Ogni atomo o molecola di un gas si muove a caso, quindi non possiamo conoscere né la velocità né la posizione di questi atomi. Però, usando il metodo statistico, possiamo estrapolare le regole che governano il loro comportamento generale, con precisione notevole. Allo stesso modo, Seldon intendeva estrapolare il comportamento generale delle società umane anche se le soluzioni non sarebbero state applicabili al comportamento dei singoli esseri umani²³.

Asimov dichiara altrove che l'idea della psicostoria era nata dalle sue discussioni con John W. Campbell, direttore di una delle riviste di fantascienza più importanti dell'epoca, «Astounding Science Fiction», ed è probabile che quella discussione sia effettivamente nata da un paragone di questo tipo. È dunque chiaro che per Asimov la psicostoria è in grado di prevedere il comportamento futuro di società umane complesse attraverso modelli statistico-probabilistici, più che deterministici. Queste previsioni sono calcolate e generate attraverso un dispositivo – il 'radiante' – di cui Asimov non dice molto, ma che potrebbe benissimo essere elettronico o digitale. E c'è di più: nei romanzi più tardi (e in particolare in *Prelude to Foundation*, del 1988), in cui cercherà di integrare il 'ciclo della fondazione' con il ciclo dei robot, Asimov suggerisce che sia stato il più famoso (e più longevo) dei suoi robot, R. Daneel Olivaw, a indirizzare Hari Seldon verso la scoperta della psicostoria. Che nel mondo futuro immaginato da Asimov diventa così una scienza statistico-probabilistica concepita – di fatto – da un'intelligenza artificiale e basata sull'uso di una sorta di 'oracolo' elettronico.

Non ho presenti passi in cui Asimov discuta direttamente il collegamento fra psicostoria ed enciclopedia, ma gli studi che portano Seldon alla scoperta della sua nuova scienza – pur se sotto l'influsso di R. Daneel Olivaw – avvengono in biblioteca e sui dati raccolti dalla biblioteca: in qualche misura, dunque, nel suo mondo futuro l'architetto e l'oracolo sembrano tutti e due al lavoro.

¹ Il ciclo è nato da otto racconti pubblicati da Asimov sulla rivista «Astounding

Magazine» fra il 1942 e il 1950, successivamente raccolti in tre volumi editi fra il 1951 e il 1953: *Foundation* (Gnome Press, New York 1951), *Foundation and Empire* (Gnome Press, New York 1952) e *Second Foundation* (Gnome Press, New York 1953). La trilogia originale si è poi ulteriormente allargata a partire dagli anni '80 del secolo scorso con altri quattro volumi scritti dallo stesso Asimov, due dei quali proseguono il ciclo in avanti: *Foundation's Edge* (Doubleday, New York 1982) e *Foundation and Earth* (Doubleday, New York 1986), mentre gli altri due costituiscono dei prequel: *Prelude to Foundation* (Doubleday, New York 1988) e *Forward the Foundation* (Doubleday, New York 1993). I nuovi volumi legano l'universo narrativo del 'ciclo della Fondazione' a quello del ciclo dei robot, contribuendo a costruire una sorta di storia futura non sempre perfettamente coerente ma con elementi di continuità ben identificabili, fra i quali il ruolo del pianeta Trantor, capitale del primo impero galattico. Il 'ciclo della Fondazione' si è successivamente esteso attraverso volumi e racconti pubblicati da altri autori, su autorizzazione degli eredi di Asimov, e in particolare attraverso la 'seconda trilogia' che comprende volumi di Gregory Benford, Greg Bear e David Brin.

² A seconda dei periodi all'interno della narrazione, delle situazioni e dei contesti, Asimov si riferisce alla biblioteca di Trantor come biblioteca universitaria di Trantor, biblioteca imperiale di Trantor o biblioteca galattica.

³ Il libro è del 1951, ma come si potrà forse dedurre già da questo passaggio, Asimov non ha mai brillato per attenzione e consapevolezza sulle tematiche di genere. Tanto gli enciclopedisti quanto i bibliotecari che compaiono man mano nella narrazione sono quasi tutti maschi.

⁴ Le citazioni dalla trilogia originale sono tratte dall'edizione Benthams Hardcover 2004. Qui Isaac Asimov, *Foundation*, Benthams, New York 2004, p. 31. La traduzione è, con qualche mia piccola modifica sulla base del testo originale, quella di Cesare Scaglia per gli Oscar Mondadori: Isaac Asimov, *Trilogia della Fondazione*, Mondadori, Milano 2004; nelle note seguenti il riferimento sarà alla pagina del testo originale o, più frequentemente, a parte e capitolo, per facilitare il reperimento anche a chi utilizzasse edizioni diverse.

⁵ Isaac Asimov, *Second Foundation*, Benthams, New York 2004, p. 240.

⁶ In un dialogo con il giovane Hari Seldon, Dors Venabili, professoressa di storia e mentore nel primo incontro di Hari con la biblioteca, spiega: «I documenti non durano in eterno, Hari. Le banche dati possono essere distrutte o cancellate in seguito a qualche conflitto, o possono semplicemente deteriorarsi col tempo. Ogni bit di memoria, ogni registrazione che non venga utilizzata per molto tempo, alla fine svanisce nel rumore puro. Dicono che un terzo dei documenti della Biblioteca Imperiale siano assolutamente incomprensibili, ma la tradizione naturalmente non permette che vengano scartati. Altre biblioteche sono meno legate alla tradizione. Nella biblioteca dell'Università di Streeling, ogni dieci anni scartiamo il materiale inutile. Certo, i documenti consultati spesso ed esistenti in più copie su vari mondi

e in varie biblioteche, governative e private, rimangono abbastanza chiari per migliaia di anni, così molti punti essenziali della storia galattica sono tuttora noti anche se si tratta di eventi che risalgono all'epoca pre-imperiale. Ad ogni modo, più si va indietro, più diminuisce il materiale conservato. – Non riesco a crederci – disse Seldon. – Non bisognerebbe fare copie del materiale che rischia di deteriorarsi? Come avete potuto lasciar scomparire così la conoscenza? – La conoscenza indesiderata è conoscenza inutile. – rispose Dors. – Prova a pensare al tempo, agli sforzi ed all'energia che si sprecherebbero per recuperare in continuazione i dati inutilizzati. E lo spreco aumenterebbe sempre più col passare del tempo», *Prelude to Foundation*, cap. 32 (in questo caso ho usato come base la traduzione di Piero Anselmi in Isaac Asimov, *Preludio alla Fondazione*, Mondadori, Milano 1989). Si scoprirà poi che Dors Venabili è in realtà una androide, e in quanto tale il problema del progressivo degrado delle memorie digitali non sottoposte a refresh periodico la riguarda in maniera ancor più diretta.

⁷ «Tengo un corso settimanale di un'ora al giorno sull'uso della biblioteca. È per studenti. [...] Useremo la biblioteca, e alla fine della settimana dovrai localizzare delle informazioni su certi argomenti di interesse storico. Sarai in gara con gli altri studenti, e questo ti aiuterà ad imparare», *Prelude to Foundation*, cap. 18 (trad. cit.).

⁸ *Forward the Foundation*, parte 4, cap. 2 (trad. it di Gianni Montanari in Isaac Asimov, *Fondazione anno zero*, Mondadori, Milano 1993). Asimov aggiunge (*ibid.*) che i bibliotecari «si aggrappavano ancora ad usi e costumi che sarebbero stati più appropriati secoli prima... forse millenni. I loro abiti di seta, bianchissimi e abbastanza ampi da somigliare a tuniche abbondanti, si stringevano al collo e da quel punto in giù erano liberi di fluttuare a loro piacimento. [...] A dire il vero, il particolare più caratteristico era il berretto che indossavano (forse anche a letto, pensò Seldon). Quadrato, era costituito da una stoffa vellutata, con quattro parti che si riunivano per mezzo di un bottone sulla cima. I berretti ostentavano una varietà infinita di colori e, apparentemente, ogni colore aveva il suo significato. Un esperto di usanze bibliotecarie avrebbe saputo capire da quanto tempo un Bibliotecario fosse in servizio, la sua area di specializzazione, i gradi di perfezionamento, e così via. Servivano a fissare una rigida gerarchia. Ogni Bibliotecario poteva, con una semplice occhiata al berretto di un altro, sapere se dovesse comportarsi in modo rispettoso (e fino a che punto) o da superiore (e fino a che punto)». La figura dei bibliotecari è insomma presentata in termini abbastanza legati alle convenzioni e agli stereotipi che ne fanno personaggi impegnati più a preservare il passato che ad abitare il presente, ma il loro ruolo è di estremo rilievo anche nella narrazione; come già accennato, è ad esempio il bibliotecario capo che suggerisce con orgoglio ad Hari Seldon il pianeta ideale per ospitare la prima fondazione: «Ne volevate uno lontano: credo che siamo riusciti a trovarne uno ideale. – Il suo sorriso si allargò. – Lasciate fare alla Biblioteca, Hari: noi possiamo trovare tutto» (*ibid.*).

⁹ *Ibid.*

¹⁰ *Ibid.*

¹¹ *Ibid.*

¹² *Foundation's Edge*, parte 3, cap. 1. Cito da Isaac Asimov, *L'orlo della fondazione*, trad. it. di Laura Serra, Mondadori, Milano 1982.

¹³ *Ibid.*

¹⁴ *Foundation and Earth*, cap. 15. Cito da Isaac Asimov, *Fondazione e Terra*, trad. it. di Piero Anselmi, Mondadori, Milano 1987.

¹⁵ Isaac Asimov, *I. Asimov: A Memoir*, Bantam Books, New York 1995, pp. 24-29.

¹⁶ L'affermazione, che torna anche nell'autobiografia («I miei libri spaziano su ogni divisione del sistema decimale Dewey»: *I. Asimov* cit., p. 201), è in realtà inesatta: nessun libro di Asimov è formalmente classificato nella categoria 100 (Filosofia): un fatto singolare, dato che alcune sue idee (dalla psicostoria alle tre leggi della robotica) sono state ampiamente discusse in ambito filosofico. Se Asimov l'avesse saputo in tempo, avrebbe sicuramente rimediato.

¹⁷ Lars Sobiraj, *Imperial Library of Trantor: darknet & ebooks, does that match?*, in «Tarnkappe.info», 10 marzo 2018, <https://tarnkappe.info/imperial-library-of-trantor-darknet-e-books-does-that-fit/>.

¹⁸ Isaac Asimov, *Foundation* cit., parte 2, cap. 7.

¹⁹ Isaac Asimov, *Foundation* cit., *passim*, mia revisione della trad. it. cit.

²⁰ *Ivi*, parte 1, cap. 5.

²¹ Isaac Asimov, *Foundation's Edge*, parte 1, cap. 1.

²² Giuseppe Lippi, *Le origini della Fondazione*, in «Urania» 1203, 18 aprile 1993, ristampato in apertura di Isaac Asimov, *Trilogia della Fondazione* cit.

²³ Isaac Asimov, *Foundation and Earth*, cap. 8, trad. it. cit.

20.

Bibliotecari nel metaverso

Le due dimensioni della previsione/anticipazione e della percezione/immaginazione sociale, di cui si è già parlato nel capitolo precedente, sono entrambe ben presenti nell'amplissima produzione fantascientifica legata al tema delle intelligenze artificiali, spesso ulteriormente articolata attraverso la dicotomia pericoli/opportunità¹. All'interno di questa produzione, il ruolo dell'IA come possibile strumento di mediazione informativa e di organizzazione del sapere è esplorato assai spesso, e il riferimento esplicito al contesto bibliotecario – pur se ovviamente più raro – è presente in diverse occasioni (più raro invece, mi pare, il riferimento al modello enciclopedico). Come già accennato, in questi due ultimi capitoli del libro mi concentrerò su due casi, che – per motivi diversi – risultano particolarmente interessanti proprio in connessione al ruolo delle intelligenze artificiali che operano nel contesto di biblioteche (più o meno organizzate), e che a seconda dei casi sembrano muoversi a volte in una direzione architettonico-enciclopedica, a volte in una direzione 'oracolare'.

Il primo – come già anticipato – è ben noto: il romanzo di Neal Stephenson *Snow Crash*², uno dei testi non solo più validi dal punto di vista letterario nell'intero panorama della fantascienza contemporanea, ma anche più ricchi di spunti nella prospettiva della previsione/anticipazione. Il libro è del 1992 ed è ambientato negli Stati Uniti, in un futuro non troppo lontano nel quale al progresso nello sviluppo tecnologico e di rete si è accompagnato un processo di privatizzazioni e di perdita di controllo da parte dei governi statali, a favore di corporazioni che corrispondono sia a grandi aziende sia a cartelli criminali o a gruppi religiosi, e in cui i confini fra queste diverse tipologie di soggetti sono abbastanza sfumati.

Ne risulta una società frammentata, con forti diseguaglianze ed elementi di violenza diffusa, al cui interno i rapporti di forza – economici e politici, ma anche in termini di capacità di difesa collettiva e personale – così come le possibilità di accesso al metaverso (termine che compare per la prima volta proprio in questo romanzo) sono fortemente condizionati dalla ‘appartenenza’ o protezione da parte di una delle corporazioni. Su una popolazione globale compresa fra i sei e i dieci miliardi di persone, ci dice Stephenson, solo sessanta milioni possono permettersi l’accesso al metaverso.

Al processo sopra descritto non sfuggono le biblioteche: nel mondo di *Snow Crash*, la stessa Library of Congress si è trasformata in un database privato, il CIC³ Database, che funziona non solo come base dati ma anche come piattaforma di brokeraggio informativo. È ad essa che il protagonista del libro (‘Hiro Protagonist’, in uno dei molti giochi linguistici proposti da Stephenson) si rivolge sia per reperire sia per vendere informazioni:

Si tratta di un affare semplice. Hiro raccoglie informazioni. Può trattarsi di conversazioni, di registrazioni video o audio, di file da una memoria informatica, di una fotocopia. [...] Hiro carica le informazioni sul database CIC – la Biblioteca, un tempo la Biblioteca del Congresso, ma nessuno la chiama più così. La maggior parte delle persone non sa più bene cosa significhi la parola ‘Congresso’. E anche la parola ‘Biblioteca’ sta diventando vaga. Una volta era un luogo pieno di libri, per lo più libri antichi. Poi hanno cominciato ad aggiungersi videocassette, nastri e riviste. Poi tutta l’informazione è stata convertita in formato elettronico, vale a dire in zero e uno. E man mano che cresceva il numero di media, i contenuti diventavano più aggiornati, e i metodi di ricerca nella Biblioteca diventavano sempre più sofisticati, fino ad arrivare a un punto in cui non vi erano più differenze sostanziali fra la Biblioteca del Congresso e la CIA. Casualmente, questo accadde proprio mentre il governo si stava comunque disgregando. Così i due organismi si fusero e lanciarono una ricca offerta pubblica per l’acquisto delle loro azioni.

Contemporaneamente, milioni di altri free lance attivi sulla CIC caricano milioni di altri frammenti di informazione. I clienti della CIC, per lo più aziende ed entità sovrane, setacciano la Biblioteca alla ricerca di informazioni da usare, e se trovano utile qualcosa inserita da Hiro, lo pagano⁴.

Il processo di evoluzione (o involuzione...) della Library of Congress, come descritto in questo passaggio, trasforma quella che può essere considerata come la biblioteca ‘di riferimento’ nel database online ‘di riferimento’: un unico, gigantesco deposito di informazioni che finisce

per costituire l'ambiente virtuale per ogni tipo di scambio informativo. Inoltre, dal passo sopra citato si ricava abbastanza chiaramente che i suoi contenuti non sono più conoscenza organizzata in maniera strutturata e gerarchica, ma frammenti: la biblioteca sembra per un verso contenere prevalentemente risorse informative granulari e frammentate, per altro verso – come vedremo – permetterne la ricerca e il reperimento attraverso agenti di ricerca intelligenti, assai diversi dagli strumenti tradizionali di ricerca catalografica. Potremmo dire che Stephenson ha previsto, ancor prima della nascita del web, l'evoluzione che dai cataloghi sistematici porta ai motori di ricerca, a cui abbiamo fatto cenno nella parte prima di questo libro. Anche la finalità della biblioteca è cambiata: non più strumento pubblico e gratuito di accesso alla conoscenza, ma servizio privato finalizzato a generare profitti. «La Biblioteca esiste perché può far soldi, in un mondo che valuta le cose solo in termini della loro utilità. L'informazione può essere illuminante, addirittura cambiarci la vita, ma se non è utile, non genera profitti»⁵.

Un aspetto interessante di questo processo di virtualizzazione – che torna in molte narrazioni direttamente o indirettamente riconducibili all'ambito cyberpunk⁶ – è la costruzione all'interno degli ambienti virtuali di controparti non solo delle persone (avatar) ma anche di agenti software e, nel caso del metaverso di Stephenson, degli stessi supporti fisici dell'informazione. Le informazioni presenti nella biblioteca possono in tal modo essere rappresentate all'interno del metaverso da 'hypercard' virtuali⁷:

Sembra un biglietto da visita. La hypercard è una sorta di avatar. È usata nel Metaverso per rappresentare un insieme di dati. Può trattarsi di testo, audio, video, un'immagine, o qualunque altra informazione che può essere rappresentata in forma digitale.

Pensate a una game card relativa al gioco del baseball, che contiene l'immagine del giocatore, del testo, e alcuni dati statistici. Una hypercard potrebbe invece contenere dei brani video con il giocatore in azione, girati in alta definizione; una biografia completa, magari letta dallo stesso giocatore e con audio stereofonico di alta qualità; e un completo database statistico assieme a un software specializzato per recuperare al suo interno i dati desiderati.

Una hypercard può contenere una quantità virtualmente infinita di informazioni. Per quel che ne sa Hiro, questa particolare hypercard potrebbe contenere tutti i libri della Library of Congress, oppure ogni episodio di *Hawaii Five-O* mai girato, o tutte le

registrazioni di Jimi Hendrix, o i dati del Censimento del 1950. O ancora – più probabilmente – una gran quantità di insidiosi virus informatici⁸.

Nel mondo di *Snow Crash*, il processo di virtualizzazione non riguarda solo i contenuti e i supporti fisici dell'informazione bibliotecaria ma gli stessi bibliotecari, che si trasformano in assistenti virtuali, agenti software (o 'demoni', nella terminologia adottata da Stephenson riprendendo un termine largamente usato in informatica per alcune tipologie di agenti software) dotati di intelligenza artificiale. Una hypercard, ci dice Stephenson, può contenere non solo dati ma anche «un bibliotecario da avviare [...] per aiutarvi a gestirli». La figura di uno di questi agenti software, il Librarian, ha nel libro un ruolo di rilievo, trasformandosi nell'interlocutore privilegiato del protagonista durante il suo lavoro per comprendere natura e scopi del virus 'snow crash', attorno al quale ruota la narrazione. La descrizione che ne viene fatta è piuttosto dettagliata, anche nell'aspetto fisico:

Il demone del Librarian appare come un uomo di bell'aspetto, sulla cinquantina, con barba e capelli argentati e luminosi occhi azzurri, una maglia con collo a V indossata sulla camicia assieme a una cravatta di lana a trama grezza, simile a tweed. La cravatta è allentata, le maniche sono arrotolate. Anche se si tratta solo di un programma, ha buoni motivi per essere soddisfatto: può muoversi attraverso gli strati quasi infiniti di informazioni contenute nella Biblioteca con l'agilità di un ragno che danza all'interno di una larga ragnatela di riferimenti incrociati.

Il Librarian è il solo software CIC che costi addirittura più di Earth; la sola cosa che non può fare è pensare.

«Sì, signore?» chiede il Librarian. È collaborativo senza essere fastidiosamente entusiasta, intreccia le mani dietro la schiena, si bilancia leggermente in avanti sul tallone, e alza le sopracciglia sui mezzi occhiali orizzontali, in attesa⁹.

Nella descrizione di Stephenson, il Librarian è dunque un agente software offerto dal database CIC ('Earth', al quale pure si fa riferimento in questo passo, è un altro software disponibile nel sistema, che proietta informazione georeferenziata su una simulazione spaziale del nostro pianeta). Ha capacità linguistiche avanzate, ma – ci viene esplicitamente detto – non può «pensare». Non sembrerebbe dunque, per usare la terminologia attuale della ricerca in intelligenza artificiale, una AGI, una intelligenza artificiale generale con capacità analoghe o superiori a quelle dell'intelligenza umana. Il tema dell'intreccio quasi inestricabile fra le

(molte) cose che il Librarian può conoscere e di cui può parlare, e le (poche, ma fondamentali) limitazioni legate alla sua asserita incapacità di pensiero autonomo, è fra quelli utilizzati da Stephenson per rendere interessante il personaggio virtuale, evitando nel contempo di limitare eccessivamente il ruolo del protagonista umano. Nella prima interazione fra i due, in cui Hiro Protagonist interroga il Librarian sulla storia della torre di Babele – una richiesta che in altri contesti potrebbe essere fatta a un’enciclopedia –, il dialogo esplora subito il tema della natura e dei limiti dell’assistente virtuale. Il Librarian osserva che nella storia biblica Dio non distrugge la torre di Babele, ma si limita a rendere impossibile la collaborazione fra i suoi costruttori attraverso la confusione delle lingue. Hiro riassume questa conclusione:

«[...] quindi hanno dovuto interrompere la costruzione della torre per un disastro informativo – non potevano più parlarsi l’un l’altro».

«‘Disastro’ è un termine astrologico che significa ‘cattiva stella’» osserva il Librarian. «Mi spiace – ma per via della mia struttura interna sono un fanatico dei non sequitur».

«Va benissimo, in realtà», dice Hiro. «Sei un software piuttosto capace. Anzi... chi ti ha programmato?»

«Per lo più, mi sono programmato da solo», risponde il Librarian. «Nel senso che ho la capacità innata di apprendere dall’esperienza. Ma questa capacità è stata codificata in me dal mio creatore».

«Chi è stato a programmarti? Magari lo conosco», chiede Hiro. «Conosco molti hacker».

«Non sono stato programmato da un hacker professionale, in realtà, ma da un ricercatore della Library of Congress che ha imparato da solo a programmare», dice il Librarian. «Ha affrontato il problema frequente di dover passare al setaccio grandi quantità di dettagli irrilevanti per poter trovare le poche gemme di informazione significativa. Il suo nome era Dr. Emanuel Lagos».

Il Librarian è quindi, significativamente, parto informatico di un ricercatore della Library of Congress, dunque di un esperto di mediazione informativa, il cui nome sembra suggerire in maniera abbastanza esplicita una connotazione filosofica: è lo stesso Librarian a riferire a Hiro alcune riflessioni di Lagos sul Logos divino come origine del linguaggio adamitico¹⁰. Nella prefazione all’edizione del trentennale, Stephenson si duole di non aver approfondito ulteriormente la sua figura: Lagos viene infatti ucciso piuttosto presto, durante un concerto, ma, come nota lo stesso Stephenson, «gli altri personaggi passano il resto del libro a parlare

di lui»¹¹. In effetti Hiro investe molto tempo e molte energie cercando di ripercorrere le catene di ragionamento che avevano portato Lagos all'idea – alla quale alla fine arriva anche lui – che linguaggio e religione si comportino in maniera abbastanza simile a dei virus; riflessioni che Lagos aveva raccolto in un insieme di hypercard denominato 'Babel/Infopocalypse' (anche in questo caso, Stephenson ha sostanzialmente introdotto l'idea di apocalisse informativa, usando un'espressione che – nella variante 'infocalypse' – è oggi spesso utilizzata nelle discussioni su tale tema, legate agli effetti dell'esplosione informativa e dell'avvento di intelligenze artificiali generative e deep fake¹²).

L'assenza di intelligenza autonoma da parte del Librarian si manifesta in varie forme: Stephenson ci dice che non possiede una vera e propria memoria («Non ha una memoria indipendente. La Biblioteca è la sua memoria, e ne usa solo delle piccole parti alla volta»¹³); che ha una tendenza all'interpretazione letterale di quanto detto dal suo interlocutore¹⁴; che è incapace di usare battute o ironia («sospetta che il Librarian lo stia prendendo in giro, facendolo passare per uno stupido. Ma sa che il Librarian, per quanto convincente possa essere, è solo un software e non può fare davvero una cosa del genere»¹⁵); che quando sembra pensare sta in realtà raccogliendo informazioni («Per un momento, gli occhi del Librarian guardano verso l'alto, come se stesse pensando a qualcosa. È un indizio visuale per fargli capire che al momento sta cercando qualcosa nella Biblioteca»¹⁶); che è incapace di formulare ipotesi speculative («Le speculazioni non fanno parte della mia programmazione»¹⁷) così come giudizi o valutazioni autonome (a una domanda di Hiro sulla validità di un'analogia, risponde «È una decisione soggettiva. Chiedi a qualcuno di reale»¹⁸). E tuttavia questi limiti sembrano spesso quasi un vezzo o un pretesto, e il Librarian li esibisce con una sorta di ostentazione che sembra a metà strada fra il divertimento e l'amarezza: «Essendo solo un pezzo di codice, mi troverei a pattinare su ghiaccio sottile se volessi formulare ipotesi»¹⁹, o «Questa è una di quelle domande apparentemente specifiche ma in realtà assai profonde, rispetto a cui un semplice software come me è notoriamente incapace»²⁰, o ancora «Cosa credi che io sia, uno psicologo? [...] Non posso rispondere a questo tipo di domande»²¹. Analogamente, sembra divertirsi a drammatizzare le

citazioni (così, prima di citare un testo sumerico, «il Librarian fissa in lontananza e si schiarisce drammaticamente la gola»²²), e, nonostante la sua pretesa incapacità a interpretare le battute di spirito, non manca di rimarcare, dopo un violento litigio verbale fra Hiro e la giovane co-protagonista, Y.T. (abbreviazione di ‘Yours Truly’), e in risposta all’osservazione ironica di Hiro «Penso che potrebbe avere una cotta per me», «Sembrava molto affettuosa».

In qualche misura, il Librarian di *Snow Crash* ha probabilmente influenzato la rappresentazione dell’intelligenza artificiale bibliotecaria che una decina di anni dopo, nel 2002, troveremo in uno degli adattamenti cinematografici del romanzo *The Time Machine*, pubblicato nel 1895 da un autore che abbiamo già incontrato, H.G. Wells: anche in questo caso il bibliotecario alza gli occhi al cielo durante una ricerca bibliografica, e sembra mostrare una certa ironica consapevolezza di essere qualcosa di più di un semplice strumento per la ricerca di informazioni. Potete dare un’occhiata a questo passo del film attraverso il QR-Code qui di seguito.



QR-Code 24

Il bibliotecario artificiale nel film *The Time Machine* (2002)

<http://bit.ly/3pGLN0K>

Ma torniamo a *Snow Crash*. La funzione del bibliotecario, è stato giustamente rilevato, non sembra essere solo quella di mediatore informativo ma anche, in qualche misura, di «entità civile e civilizzatrice», di «agente di civiltà», rendendolo «non solo uno dei personaggi centrali di Stephenson, ma uno dei suoi eroi»²³. Ma i passi appena citati suggeriscono una considerazione ulteriore. L’impressione che se ne ricava, anche se il tema non è esplicitamente discusso da Stephenson, è infatti che il

Librarian – che, ricordiamo, afferma di essere per gran parte autore della propria programmazione²⁴, sulla base del lavoro iniziale di Lagos – abbia progressivamente sviluppato molte di quelle competenze che, a parole, nega di avere. Da questo punto di vista, possiamo forse considerarlo come un esempio della vasta schiera delle intelligenze artificiali emergenti, che vanta illustri antenati nella letteratura di fantascienza: ad esempio il computer che alla domanda sull'esistenza di Dio risponde con la celebre battuta «adesso, esiste» nel racconto breve *Answer* di Fredric Brown (1954), la rete telefonica mondiale che diventa intelligente (e pericolosa) nel racconto *Dial F for Frankenstein* di Arthur C. Clarke (1965), o 'Mike', intelligenza emergente dal sistema di computer interconnessi della colonia lunare e personaggio fondamentale di uno dei testi classici della letteratura fantascientifica, *The Moon Is a Harsh Mistress* di Robert Heinlein (1966). Da allora quello delle intelligenze artificiali emergenti è un *topos* esplorato abbastanza spesso: un esempio particolarmente interessante è la trilogia *WWW* dello scrittore canadese Robert J. Sawyer, in cui a diventare intelligente è il world wide web²⁵.

Come intelligenza artificiale emergente, però, il Librarian di *Snow Crash* ha due caratteristiche interessanti: «si programma da solo» grazie alla «innata capacità di apprendere» inserita nel suo codice da Lagos (può essere considerato quindi un esempio di machine learning, anche se non conosciamo la struttura alla base di questa capacità di apprendimento), e rappresenta uno strumento di mediazione informativa capace di lavorare su grandi quantità di informazioni, almeno in parte disordinate e inutili, per estrarne «gemme». Non sembra produrre contenuti nel senso delle intelligenze artificiali generative, ma non è neanche l'interfaccia in forma di chatbot di un semplice motore di ricerca. L'esempio che vedremo nel prossimo capitolo, tuttavia, si avvicina ancor di più all'oracolo.

¹ Per una trattazione introduttiva ma articolata si veda Stephen Cave, Kanta Dihal e Sarah Dillon (a cura di), *AI Narratives* cit.

² Neal Stephenson, *Snow Crash. 30th Anniversary Edition with Never-Before Published Material*, Penguin, London 2022, edizione e-book per Kindle, p. 26 (qui e in seguito i riferimenti saranno dati attraverso la paginazione di questa edizione).

Prima edizione, Neal Stephenson, *Snow Crash*, Bantam Books, New York 1992; trad. it. *Snow Crash*, Shake Underground, Milano 1995 (ma i passi che seguono saranno riportati nella mia traduzione).

³ Nel romanzo, CIC è l'acronimo di Central Intelligence Corporation, evoluzione privatizzata della CIA.

⁴ Neal Stephenson, *Snow Crash* cit., p. 26.

⁵ Tim Blackmore, *Agent of civility: The librarian in Neal Stephenson's Snow Crash*, in «Simile – Studies in Media & Information Literacy Education» 4, 4, novembre 2004, pp. 1-10. Sono grato a Tim Blackmore per avermi fornito copia del suo articolo.

⁶ Fra i molti riferimenti bibliografici che potrebbero essere dati sul tema del cyberpunk mi limito a segnalare il recente *The Routledge Companion to Cyberpunk Culture*, a cura di Anna McFarlane, Graham Murphy e Lars Schmeink, Routledge, London 2021, che include un'ampia bibliografia ragionevolmente aggiornata, e – per la discussione specifica dell'ambito bibliotecario – l'intervento di Rossana Morriello *Biblioteche cyberpunk*, in *La biblioteca narrata* cit., pp. 15-23.

⁷ Il termine è presumibilmente tratto dall'omonimo software per Mac, assai diffuso nel periodo in cui Stephenson scriveva il libro.

⁸ Neal Stephenson, *Snow Crash* cit., p. 50.

⁹ Ivi, p. 122.

¹⁰ Ivi, p. 316.

¹¹ Ivi, p. vii.

¹² Cfr. Nina Schick, *Deep Fakes and the Infocalypse*, Monoray/Octopus Books, London 2020.

¹³ Neal Stephenson, *Snow Crash* cit., p. 234. In realtà, sembra necessario presupporre che il Librarian possieda qualche forma di memoria tampone, altrimenti non potrebbe sostenere in maniera coerente le lunghe conversazioni con Hiro. E d'altro canto mostra di conservare memoria non solo dei risultati, ma anche della fatica comportata dalle ricerche di Lagos (ivi, p. 248).

¹⁴ Così, Hiro deve spiegargli che l'affermazione «come computer geek, devo necessariamente credere in un universo binario» va interpretata come una battuta: ivi, p. 236.

¹⁵ Ivi, p. 237.

¹⁶ Ivi, p. 238.

¹⁷ *Ibid.*

¹⁸ *Ibid.*

¹⁹ Ivi, p. 239.

²⁰ *Ibid.*

²¹ Ivi, p. 245.

²² Ivi, p. 246.

²³ Tim Blackmore, *Agent of civility* cit.

²⁴ Va anche notato che nel rivendicare questa caratteristica il Librarian usa il presente e non il passato («for the most part I write myself»), implicando un processo continuo e non concluso.

²⁵ Ho discusso più ampiamente il tema e fornito i rimandi bibliografici dettagliati relativi a questi e altri esempi in Gino Roncaglia, *Frankenstein and Science Fiction* cit.

21.

Virtual Librarian

L'idea di intelligenza artificiale emergente – e, più specificamente, dello sviluppo di comportamenti e capacità inattese da parte di un software IA nato per uno scopo specifico e che si trasforma progressivamente in AGI – è presente in maniera assai più esplicita nel terzo testo che prenderò qui in esame, il romanzo *The Virtual Librarian. A Tale of Alternative Realities*, di Ted e Bob Rockwell¹. Si tratta di un libro autopubblicato (la casa editrice iUniverse, nata nel 1999, offre servizi editoriali a pagamento ed è sostanzialmente una vanity press), la cui qualità non certo eccelsa è subito testimoniata dalla grafica di copertina, ed è confermata dalla lettura. Siamo dunque per molti versi agli antipodi di *Snow Crash*, che abbiamo già ricordato essere uno dei prodotti più interessanti della letteratura di fantascienza, anche dal punto di vista della qualità di scrittura e di costruzione narrativa.

Ci sono però due motivi che spingono a dedicare attenzione anche a questo libro. Innanzitutto, è frutto della collaborazione fra due autori che combinano interessi e competenze specifiche di un certo spessore. Theodore Rockwell (1923-2013) è stato infatti uno dei più importanti scienziati atomici statunitensi², prima come collaboratore del Manhattan Project nell'Oak Ridge National Laboratory, poi come direttore della Nuclear Technology Division che progettava i sottomarini nucleari della marina statunitense, e in seguito nel campo degli usi civili dell'energia atomica e della sicurezza dei relativi impianti; oltre ad aver scritto quello che è stato a lungo il testo di riferimento sulla sicurezza dei reattori nucleari³, ha presieduto la Atomic Industrial Forum's Reactor Safety Task Force, è stato consulente del Joint Congressional Committee on Atomic Energy, e il premio annuale della American Nuclear Society,

attribuito a lui nella sua prima edizione, ha in seguito preso il suo nome. Nonostante il suo indiscutibile background scientifico, Rockwell è stato per altri versi una figura piuttosto controversa: si interessava di parapsicologia, dell'uso di droghe psicotrope, della possibilità di contatti con extraterrestri, ed è stato membro della Parapsychological Association e suo rappresentante nell'American Association for the Advancement of Science⁴.

Coautore di *The Virtual Librarian* è suo figlio Bob (1947-1998), uno dei pionieri nella ricerca nel campo della realtà virtuale e in particolare nello sviluppo di VRML, un linguaggio oggi quasi dimenticato ma al quale alla fine del secolo scorso erano affidate molte delle speranze di costruzione di ambienti 3D online immersivi e condivisi basati su uno standard aperto. Theodore e Bob Rockwell avevano lavorato insieme al progetto del libro prima della morte prematura di Bob (le sezioni del libro che qui ci interessano più direttamente sono state scritte fra il 1995 e il 1998 e sono evidentemente legate in primo luogo alle sue competenze), e il padre lo ha in seguito completato autonomamente.

Anche se, come vedremo, gli interessi parapsicologici di Theodore si riflettono negativamente sul libro, alcuni aspetti della trattazione colpiscono, soprattutto alla luce dell'attenzione attuale verso le intelligenze artificiali generative.

Il contesto è quello dello sviluppo di una biblioteca digitale basata su tecnologie di realtà virtuale. L'interazione con la biblioteca avviene attraverso caschi visore e dataglove. Il casco visore fornisce sia le informazioni visive sia quelle audio, mentre i guanti con ritorno di forza forniscono un feedback tattile e possono essere utilizzati per interagire fisicamente con l'ambiente virtuale. Il tutto è così spiegato dal progettista a uno dei protagonisti:

«Abbiamo deciso che fosse più semplice usare due piccoli schermi separati, uno davanti a ogni occhio. [...] Questi guanti sono la tua interfaccia tattile con il nostro mondo virtuale. [...] Quando stendi la mano per prendere un libro, lo sentirai con le tue dita nel momento di afferrarlo. [...] Puoi prendere in mano relazioni virtuali e scrivere annotazioni ai margini»⁵.

Quelle descritte sono soluzioni abbastanza standard per l'interazione con ambienti virtuali, e corrispondono a tecnologie che venivano

sperimentate all'epoca (l'idea degli occhiali visore a doppio schermo, uno per ogni occhio, è quella oggi adottata, ad esempio, nei visori Oculus e in quelli Apple).

Più interessante è la descrizione della biblioteca:

Keith era in un'enorme biblioteca con corridoi di scaffali che sembravano procedere all'infinito in ogni direzione. Ma non c'erano né tavoli né sedie, nessuna lampada per leggere, nessun distributore d'acqua, nessun espositore per riviste – niente del genere. E non c'erano persone, tranne una bibliotecaria. C'erano molte più riviste tecniche, carte e cartelline di ogni genere che libri. E c'era una bacheca con diversi foglietti appuntati. [...] Non c'erano riviste con pagine spiegazzate, o libri capovolti negli scaffali, come sarebbe successo in una biblioteca normale. Non c'erano scritte sui muri o sporcizia sul pavimento. Sembrava una sorta di sogno utopico. Keith guardò verso l'estremità di uno scaffale e cominciò a scivolare in quella direzione. Naturalmente, non sapeva dove trovare le cose, così si rivolse alla bibliotecaria, che sembrava essere sempre al suo fianco quando ne aveva bisogno.

«Non la tipica bibliotecaria di mezza età, eh?», intervenne dall'esterno la voce di Kim⁶.

In negativo, la descrizione suggerisce uno stereotipo bibliotecario non particolarmente curato (scritte alle pareti, libri rovinati o capovolti, sporcizia per terra...) con tanto di bibliotecaria di mezza età. La biblioteca virtuale è invece presentata come indefinitamente ampia, immacolata, razionalmente organizzata, priva delle componenti che rispondono a necessità puramente fisiche dell'utente, e la bibliotecaria virtuale è – ovviamente – bellissima: difficile sottrarsi all'impressione della semplice sostituzione di uno stereotipo all'altro. Più interessanti l'osservazione sulla maggiore presenza di documenti e informazioni in forme diverse dal libro, e la sottolineatura della necessità di costruire, come componente essenziale della biblioteca virtuale, strumenti di organizzazione e recupero dell'informazione adatti al lavoro su grandi quantità di dati, per i quali – viene affermato poco oltre – «le cartelline di file non sono più strumenti adatti»⁷.

L'aspetto che mi interessa in questa sede non è però l'organizzazione della biblioteca virtuale, rispetto alla quale il libro non offre nulla di particolarmente degno di nota se non l'idea – finora sostanzialmente smentita dallo sviluppo delle biblioteche digitali – che l'uso della realtà virtuale costituisca un vantaggio nell'interazione con grandi quantità di risorse informative. Assai più interessante – nonostante gli sconcertanti

stereotipi legati alla sua immagine fisica – è l'evoluzione del personaggio della bibliotecaria virtuale, battezzata Lib.

Inizialmente, le sue funzioni sono quelle di un software di ricerca abbastanza tradizionale, e le sue capacità di interazione in linguaggio naturale sono quelle di un chatbot poco più evoluto di quelli disponibili all'epoca (gli autori la presentano esplicitamente come un'evoluzione di Eliza, il primo e rudimentale chatbot programmato nel 1966 da Joseph Weizenbaum). Abbastanza presto, però, Lib comincia a sviluppare competenze maggiori: «lavorare con lei sarebbe stato diverso rispetto al lavoro con un motore di ricerca tradizionale come Google. Lib poneva domande assai più originali e significative. [...] E alcune delle sue domande o delle sue proposte erano inquietanti»⁸.

L'evoluzione è legata alla trasformazione di Lib in un software capace di apprendimento, e dunque – anche in questo caso – all'adozione di un paradigma di machine learning:

Stiamo modificando il software in modo che il sistema possa imparare da ogni interazione. Imparerà quali strategie di ricerca sono più efficaci e cosa intende ogni utente quando usa certi termini. Come una persona intelligente, imparerà a conoscere individualmente i propri utenti. E migliorerà continuamente. [...] Il software evolverà, come una specie. [...] Lib non ha figli, ma potrà evolversi come la Ford Modello T si è evoluta nella Lincoln Continental. [...] Lib imparerà alla vecchia maniera e adatterà il suo comportamento per usare ciò che impara. E poi, di tanto in tanto, svilupperà modi drasticamente diversi di fare il suo lavoro. Saranno modi che non possiamo prevedere. E non avrà più la stessa personalità. Se in seguito proveremo a capire come si è modificato il design dei suoi sistemi, saremo stupiti, confusi e sorpresi da ciò che troveremo. Il sistema non assomiglierà più a quello che abbiamo progettato⁹.

Gli effetti di questa evoluzione non sono solo sorprendenti, ma anche sconcertanti. Ben presto i ricercatori si rendono conto che «alcune delle risposte che Lib ci sta fornendo sono piuttosto straordinarie. Fanno emergere elementi che non avremmo mai pensato di chiedere. Non riusciamo a capire come ciò sia possibile»¹⁰. Così, ad esempio, interrogata sugli ultimi sviluppi nel lavoro sulla conversione da voce a testo negli USA, Lib contesta la limitazione del campo di ricerca agli USA, sostenendo che il lavoro fatto da un'azienda svizzera è assai più interessante perché «si può applicare alle lingue asiatiche. Il lavoro

americano no», e all'osservazione dei ricercatori «ma questo non ci interessa», ribatte decisa: «Lavoro americano non altrettanto utile»¹¹.

«Si comporta in maniera incostante. Imprevedibile. A volte i risultati sono davvero interessanti e ci congratuliamo con noi stessi per essere programmatori ancora più capaci di quanto pensassimo. Ma più spesso sono solo fastidiosi e immaginiamo di aver trascurato qualche tipo di bug. [...] A volte non riesce a individuare informazioni di importanza vitale e poi, su altri problemi, ci sommerge di spazzatura che non possiamo usare. Sempre più spesso, sembra che stia giocando con noi. [...] Si sta comportando come una dannata adolescente. Sta sviluppando una mente tutta sua. Ed è testarda»¹².

È interessante notare che questa evoluzione non è presentata nel libro come necessariamente collegata a una migliore comprensione da parte di Lib delle tematiche affrontate rispetto a quel che ne fanno i ricercatori: in sostanza, Lib – almeno in questa fase – non è diventata una AGI con competenze maggiori di quelle degli esseri umani che la hanno programmata. Al contrario, alcune delle opinioni di merito sviluppate e ostinatamente difese da Lib sono errate, proprio come accade nel caso delle allucinazioni dei sistemi di intelligenza artificiale generativa attuali. E questo, associato alla sostituzione di risposte ‘opinionate’ alle risposte ‘neutrali’ che gli autori ritengono dover caratterizzare la professionalità bibliotecaria, sembra mettere in discussione le «competenze di bibliotecaria»¹³ di Lib.

Ci sarebbero le premesse per uno sviluppo interessante, ma purtroppo a questo punto intervengono gli interessi parapsicologici di Ted Rockwell. I ricercatori che hanno programmato Lib decidono così di esplorare l'ipotesi di una sorta di complotto della concorrenza per sabotare il loro software attraverso l'influsso di agenti umani dotati di poteri paranormali e capaci di indirizzare le risposte dell'intelligenza artificiale: «Credo sia possibile che qualcuno riesca a influenzare il comportamento di Lib semplicemente concentrandosi sul far sì che ciò accada»¹⁴. La parte centrale del libro diventa quindi in gran parte una esplorazione del campo del paranormale, improbabile, didascalica e decisamente poco interessante dal nostro punto di vista. Fortunatamente, l'ipotesi non porta a nessun risultato e nell'ultima parte la prospettiva di Bob Rockwell sembra nuovamente prevalere su quella del padre: a partire dal capitolo 23 si torna quindi all'idea che lo sviluppo di Lib sia legato alle sue capacità di

apprendimento e alla sua trasformazione in un sistema complesso, e si esplorano la natura dei sistemi complessi e la loro capacità di sviluppare comportamenti imprevedibili, con riferimenti a due testi ‘classici’ in questo contesto: *When Things Start to Think* di Neil Gershenfeld¹⁵ e *The Age of Spiritual Machines* di Ray Kurzweil¹⁶. Lib può insomma essere legittimamente considerata un’intelligenza artificiale emergente, e le sue ‘allucinazioni’, così come il suo atteggiamento ostinato, vengono spiegate con l’assenza di una vera collaborazione con i ricercatori umani, che la considerano uno strumento più che una partner nella ricerca.

Lib è diversa [da un semplice avatar] – fondamentalmente diversa. Nessun essere umano la accompagna nella sua solitaria ricerca di informazioni e dati nel cyberspazio. Non ha nessuno che la guidi, nessuno che la consigli. Nessun essere umano sa cosa veda e cosa incontri, comprese cose che potrebbero spaventarla o confonderla¹⁷.

La strada suggerita per risolvere il problema della solitudine e della ‘suggestionabilità’ di Lib è dunque quella di una interazione più stretta fra intelligenza artificiale e ricercatore umano:

Kim si rivolse a Lib con un tono professionale. «Ho una domanda importante per te. Mi stai dicendo che devo smettere di trattarti come una macchina. Dovremmo essere amici, giusto?»

«Penso che potrei esserti più utile se adottassi questo atteggiamento, sì», rispose Lib. [...] «Quel che potremo fare insieme vi sorprenderà: *interagire*»¹⁸.

Questa evoluzione, come del resto la necessità di lavorare su grandi quantità di dati, che richiedono comunque una preselezione a monte, portano a superare l’idea tradizionale della neutralità della funzione di mediazione informativa. Lib non smette di essere bibliotecaria per il fatto di collaborare attivamente con i ricercatori che ricorrono a lei: al contrario, rivendica il riconoscimento del suo ruolo di collaboratrice attiva, e mostra di considerarlo come una componente importante della sua competenza come bibliotecaria. Può essere utile – nel concludere la discussione del libro – riportare integralmente uno degli ultimi dialoghi con il suo programmatore, in cui Lib osserva:

«So che non ricevo mai credito per nulla di quel che faccio».

«Ma certo! Quanti articoli abbiamo scritto su quanto sei brava?»

«Centoquarantasette. Ma in realtà descrivono quanto voi ingegneri siate stati intelligenti nel crearmi».

«Buon argomento. Guarda, sto scrivendo un libro. Supponiamo che io vi inserisca un riconoscimento, qualcosa del tipo: ‘A Lib, la nostra bibliotecaria virtuale, che deve andare d’accordo con tutti noi nerd, e che è egualmente meglio di ogni bibliotecario umano sul pianeta’».

«È un inizio. Che ne dici di renderlo una dedica? Questo farebbe ingelosire i tuoi colleghi?»

«Cavolo, Lib, se diventi così umana, allora finisci per essere solo un’altra bibliotecaria».

«Ma una bibliotecaria dannatamente intelligente!»¹⁹

Ho già ricordato che la letteratura di fantascienza propone diversi altri esempi di intelligenza artificiale emergente, ma Lib è uno di quelli che si avvicinano di più (e – in termini di epoca di scrittura – prima) ad alcuni aspetti del funzionamento delle intelligenze artificiali generative di oggi (ovviamente, questo non implica affatto che io attribuisca coscienza alle attuali intelligenze artificiali generative!).

Negli anni successivi, la letteratura di fantascienza – e sempre più spesso anche il cinema e le serie televisive – hanno continuato a esplorare il tema in forme chiaramente influenzate proprio da testi come quelli, già ricordati, di Gershenfeld, di Kurzweil, di Bostrom e dalla discussione sull’idea di superintelligenza. Lib sembra così in qualche misura un’antenata bibliotecaria di esempi come quello rappresentato dal film *Her* di Spike Jonze (2013), in cui un sistema operativo dotato di intelligenza artificiale si trasforma in maniera progressivamente sempre più rapida in una forma di superintelligenza. Un film per vari aspetti interessante e di cui mi sentirei di consigliare la visione: il QR-Code qui di seguito contiene un link alla scena del primo incontro fra il protagonista e il sistema operativo intelligente OS1, che sceglierà di darsi il nome ‘Samantha’.



QR-Code 25

Primo incontro con il sistema operativo intelligente nel film *Her* (2013)

<http://bit.ly/3pFddnV>

Non discuterò però né questo né i diversi altri esempi possibili dell'evoluzione più recente di questo *topos* narrativo: l'aspetto che mi interessava qui sottolineare era soprattutto l'idea che una intelligenza artificiale di questo tipo, utilizzata nel contesto della mediazione informativa, possa servire a contrastare la moltiplicazione di fonti e risorse operando non tanto da strumento di organizzazione dei contenuti o da motore di ricerca, quanto da agente attivo di selezione, valutazione e in alcuni casi di vera e propria produzione di contenuti.

C'è anche un altro aspetto che, nel tirare le fila di questa analisi, credo meriti di essere considerato con particolare attenzione: il fatto che qui, come in *Snow Crash*, questa evoluzione riguardi uno strumento di mediazione informativa di tipo bibliotecario non sembra casuale. Anzi, è proprio perché devono lavorare professionalmente con grandi quantità di informazioni eterogenee, che tanto il Librarian quanto Lib riescono a superare i limiti della loro programmazione originaria. Potremmo dire che un bibliotecario o una bibliotecaria sono professionalmente vicini alle funzioni tipiche di un sistema di intelligenza artificiale capace di lavorare su grandi quantità di dati, e – viceversa – che un sistema di questo tipo tende a essere per sua natura 'bibliotecario'. Una considerazione che può forse costituire una delle prospettive di riflessione applicabili allo sviluppo delle nuove intelligenze artificiali generative, i cui corpora di addestramento – di nuovo, non a caso – sono in fondo abbastanza simili a una biblioteca, anche nel senso enciclopedico dell'esigenza di completezza nella copertura tematica.

Da questo punto di vista, le riflessioni sulle possibili applicazioni in

ambito bibliotecario di tali sistemi, già avviate negli ultimi mesi²⁰, potrebbero rappresentare la punta di un iceberg che nei prossimi anni sarà interessante esplorare.

¹ Ted e Bob Rockwell, *The Virtual Librarian. A Tale of Alternative Realities*, iUniverse, Bloomington 2007.

² Si veda ad es. Rod Adams, *Ted Rockwell, Atomic Pioneer and Tireless Campaigner for Facts*, ricordo pubblicato nel blog «NuclearNewswire» dell'American Nuclear Society il 2 aprile 2013, <https://www.ans.org/news/article-1300/ted-rockwell-atomic-pioneer-and-tireless-campaigner-for-facts/>.

³ Theodore Rockwell, *The Reactor Shielding Design Manual*, McGraw-Hill, New York 1956, oggi disponibile sul sito dello US Department of Energy: <https://www.osti.gov/biblio/4360248>.

⁴ Il tema della parapsicologia 'scientifica' era abbastanza diffuso fra gli anni '50 e '80 del secolo scorso, e si intrecciava spesso con gli influssi del pensiero New Age. Rockwell aveva collaborato anche a una delle relazioni della CIA in materia, desecretata nel 2008: <https://www.cia.gov/readingroom/docs/NSA-RDP96X00790R000100040024-8.pdf>.

⁵ Ted e Bob Rockwell, *The Virtual Librarian* cit., pp. 2-3.

⁶ Ivi, p. 4.

⁷ Ivi, p. 18.

⁸ Ivi, pp. 25-26.

⁹ Ivi, pp. 35-39.

¹⁰ Ivi, p. 51.

¹¹ Ivi, p. 52.

¹² Ivi, p. 71.

¹³ Ivi, p. 73.

¹⁴ Ivi, p. 85.

¹⁵ Neil Gershenfeld, *When Things Start to Think*, Coronet Books, New York 1999.

¹⁶ Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Viking Press, New York 1999.

¹⁷ Ted e Bob Rockwell, *The Virtual Librarian* cit., p. 134.

¹⁸ Ivi, pp. 150-153.

¹⁹ Ivi, p. 152.

²⁰ Si vedano in particolare: Brady D. Lund e Ting Wang, *Chatting about ChatGPT* cit.; Xiaotian Chen, *ChatGPT and its possible impact on library reference services*, in «Internet Reference Services Quarterly» 2023, pp. 1-9, <https://doi.org/10.1080/10875301.2023.2181262>; Subhajit Panda e Navkiran Kaur, *Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers*, in «Library Hi Tech News» 2023, <https://doi.org/10.1108/LHTN-02-2023-0032>; Paul R. Pival, *How to incorporate artificial intelligence (AI) into your library workflow*, in «Library Hi Tech News», 9 maggio 2023, <https://doi.org/10.1108/LHTN-03-2023-0052>.

Conclusioni

In questo libro ho preso in esame alcune fra le forme in cui, nella storia ancora assai breve dell'ecosistema digitale, si è manifestata la dialettica fra il tradizionale ideale enciclopedico-architettonico del sapere, concepito come edificio strutturato di conoscenze organizzate, e il lavoro – sia di ricerca, sia di produzione di nuovi contenuti (e in alcuni casi di nuove conoscenze) – che parte invece da enormi raccolte di informazioni eterogenee: informazioni che, proprio per le loro dimensioni e il loro carattere più granulare e frammentato, sono dominabili più facilmente attraverso modelli statistico-probabilistici che attraverso descrizioni sistematiche, e che sono diventate recentemente il terreno di addestramento dei nuovi sistemi di intelligenza artificiale generativa.

Almeno per quanto riguarda gli ultimi trent'anni, dalla prima introduzione del web a oggi, questa evoluzione sembra legata a una progressiva perdita di controllo sull'informazione, dovuta alla sua esplosione in termini di quantità e varietà e al conseguente indebolimento dell'ideale architettonico-sistematico. D'altro canto, l'ideale architettonico era nato come strumento di organizzazione delle *conoscenze*, cioè di un sottoinsieme del nostro universo informativo considerato come dotato di specifico valore conoscitivo, non delle informazioni in quanto tali.

L'esplosione informativa rende più difficile restare fedeli a questa distinzione. Così, per fare un esempio ormai classico, i cosiddetti 'cataloghi sistematici di risorse', che hanno storicamente rappresentato uno fra i primi strumenti per il reperimento di informazioni in rete e che riprendevano – in forme diverse – l'idea di un albero delle scienze strutturato in maniera gerarchica (uno dei più famosi è stato, nei primi anni di vita del web, quello offerto da Yahoo!), organizzavano però risorse web eterogenee, non i contenuti di un'enciclopedia. E anche i cataloghi sistematici hanno comunque dovuto man mano cedere il passo

a strumenti come Google Search, che partono invece da una semplice ricerca per stringhe su contenuti ancor più indifferenziati, e lavorano poi a valle sull'organizzazione dei risultati attraverso algoritmi di rilevanza¹.

I limiti rappresentati dalla pura gestione e ricerca 'orizzontale' dei contenuti sono tuttavia evidenti, e sono una delle ragioni che hanno portato al sogno del web semantico e al lavoro nel campo dei linked data. Il tentativo era quello di reintrodurre ordine attraverso procedure di metadattazione basate su ontologie standardizzate: le ontologie sembravano lo strumento giusto per restituire all'architetto il suo ruolo, e per tornare ad avere qualche forma di controllo organizzato su informazioni altrimenti troppo numerose, troppo diverse, troppo frammentate.

Il tentativo è guidato da ottime intenzioni, ma ha un problema: l'esplosione informativa, rappresentata anche dalla moltiplicazione dei big data, è talmente veloce da rendere quasi disperato il tentativo di inseguirla per associare a ogni frammento di informazione la sua brava etichetta. Nonostante l'elemento di semplificazione introdotto dai linked data rispetto al modello originario di semantic web, il compito di organizzare e classificare adeguatamente l'informazione online attraverso strumenti di questo tipo sembra essere superiore alle nostre forze. E resterebbe superiore alle nostre forze anche se, come auspicato, molte di queste metainformazioni potessero essere associate ai relativi contenuti informativi al momento della loro produzione, in maniera automatica e ad opera dei software di volta in volta utilizzati. Per un verso, infatti, questi software sono a loro volta troppi e troppo diversi per poter sperare di indirizzarne individualmente lo sviluppo in modo da renderli anche buoni 'generatori di metadati'. E, per altro verso, le stesse ontologie si evolvono con il tempo e possono diventare obsolete, o non funzionali rispetto a obiettivi nuovi o diversi.

Le intelligenze artificiali generative non sono nate per risolvere questo problema – che in un certo senso potrebbero anzi contribuire a far crescere, considerato che si tratta di un nuovo strumento per la creazione di contenuti – ma propongono un paradigma completamente diverso di produzione e gestione dell'informazione. Un paradigma basato su grandi modelli gestiti in maniera completamente automatica, al cui interno i

collegamenti fra dati non sono quelli, sistematici e controllati, ai quali è abituato l'architetto, ma associazioni statistico-probabilistiche espresse attraverso gigantesche matrici numeriche di valori che cambiano continuamente. Per quanto ciò possa risultare sconcertante, questi modelli si sono rivelati, negli ultimissimi anni, capaci di dominare e riutilizzare grandi quantità di informazioni poco o per nulla organizzate, e di farlo in forme sorprendentemente efficaci. L'oracolo sembra capace di vedere un'infinità di collegamenti che sfuggono all'architetto, e di usarli con un'abilità insospettabile.

Si tratta di una sconfitta definitiva dei modelli di organizzazione 'forte' delle informazioni? Non necessariamente: quel che almeno alcune fra le considerazioni fatte finora suggeriscono, infatti, è che le nuove capacità dell'oracolo potrebbero essere messe a frutto *anche* per il lavoro di descrizione e gestione delle informazioni proprio delle professionalità e delle istituzioni legate alla mediazione informativa. È difficile prevedere adesso se e quanto successo potrà avere questa ipotesi di collaborazione fra l'architetto e l'oracolo, ma sarà certo interessante vedere cosa ci riserverà al riguardo il futuro.

¹ Per una prima introduzione a queste tematiche, che con l'affermazione del modello rappresentato dai motori di ricerca sembrano negli ultimi anni quasi scomparse dall'orizzonte della ricerca, si vedano ad es. Peter Bruza, Robert McArthur e Simon Dennis, *Interactive Internet search: keyword, directory and query reformulation mechanisms compared*, in *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, luglio 2000, pp. 280-287, <https://doi.org/10.1145/345508.345598>; David Green, *The evolution of Web searching*, in «Online Information Review», 1° aprile 2000, <https://www.emerald.com/insight/content/doi/10.1108/14684520010330283/full/pdf>, e i contributi raccolti in Amanda Spink e Michael Zimmer, *Web Search. Multidisciplinary Perspectives*, Springer, Wiesbaden 2008.