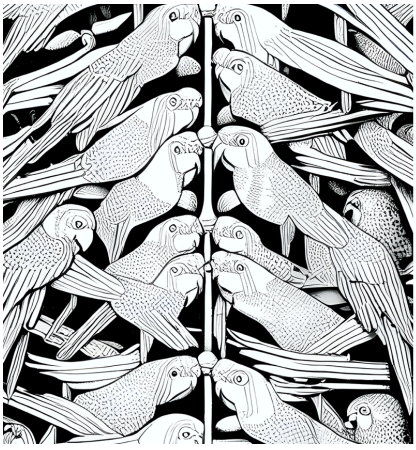
La Potenza della Latenza

Tre studi sull'AI generativa

Cosimo Accoto (maggio 2023, draft v.4)

Testi, Immagini, Agenti

Stiamo entrando in una nuova era mediale inflazionaria, quella dell'intelligenza artificiale generativa. A partire dall'esplosione della parola sintetica. La capacità di simulare la lingua non rappresenta, infatti, solo un avanzamento tecnico nel processamento macchinico del linguaggio naturale. È piuttosto un passaggio di civiltà che scardinerà le economie, le imprese e i mercati. Così come le politiche e le etiche del dire e del potere. E non solo l'ordine del discorso. Dalla produzione artistica alla diagnostica medica, dal marketing digitale al design industriale, l'era inflattiva dell'AI generativa farà leva anche sulle opportunità (e i rischi) dell'immagine sintetica. Con ridisegno dei significati culturali profondi e degli impatti strategici trasformativi derivanti dall'impiego massivo delle reti neurali artificiali e della potenza della latenza. Infine, dopo le parole macchiniche e le immagini sintetiche, esploriamo la questione emergente dell'adozione pervasiva di agenti autonomi. All'orizzonte si profila, dunque, un esercito computazionale fatto di intelligenze artificiali operative che, nei molti progetti in divenire, assolveranno ai compiti più svariati automatizzando ecosistemi di business, processi industriali e servizi erogati. Al di là dell'hype, ne valutiamo culturalmente e strategicamente potenzialità e vulnerabilità. È in costruzione un nuovo mondo in sintesi fatto di scritturazioni, raffigurazioni, agentificazioni sovrumane (o more-than-human).



- **La parola macchinica**
- **L'immagine sintetica**
- **L'agente autonomo**

▪ La parola macchinica -

Una nuova scrittura automatica

Affrontare la questione dei linguaggi sintetici, simulativi e inflattivi equivale a fronteggiare un passaggio di civiltà epocale e non episodico. Un passaggio molto commentato al momento, ma poco esplorato e compreso nella sua portata. Tecnicamente, il dispositivo che istanzia un "modello linguistico su larga scala" (LLM o *large language model*) è un assemblaggio socio-tecnico generativo fatto di abilità diverse connesse a molteplici architetture computazionali e risorse informative. La capacità di simulare il linguaggio nella sua forma testuale, di aggiustarlo in modalità contestuale, di archiviare conoscenza e informazione, di eseguire istruzioni e compiti linguistici, di sintetizzare temi con affinamento scalare, di originare sequenze di argomentazioni e tentativi di ragionamento per step, di articolare risposte e costruire dialoghi sono il frutto di un'orchestrazione complessa fatta di programmi software, dati e archivi informativi, algoritmi di apprendimento profondo anche a rinforzo umano, modelli matematico-stocastici della lingua. Si tratta, dunque, di un insieme di tecnicalità e operatività ingegneristico-computazionali intrecciate (*training on code, transformers, pre-training modeling, instruction tuning, words tokenization, reinforcement learning* con *human feedback*…) in grado di sequenziare statisticamente il linguaggio naturale umano. Il tutto, in molti dicono, per bilanciare e contrastare l'*hype* del momento- senza relazione di senso col reale. Vale a dire, cioè, senza che quel linguaggio macchinico sappia in realtà nulla del mondo e senza che abbia una qualche comprensione dei suoi significati. L'espressione usata "*pappagalli stocastici*" evoca questa scrittura simulativa verosimile, ma insensata.

Dentro le meccaniche di un LLM

Ma cos'è, in ultima istanza, un *large language model*? Possiamo dire che un LLM è un sequenziatore linguistico-probabilistico a bassa crossentropicità. Dunque, ridotto ai suoi minimi termini, è un modello matematico della distribuzione di probabilità delle parole di una lingua scritta che si sforza di minimizzare la crossentropia (cioè lo scarto tra due potenziali distribuzioni di frequenza) massimizzando, di conseguenza, la sua capacità performativa come *text predictor*. Come ha raccontato Binder (*Language and the Rise of Algorithm*, 2022), questo approccio è il frutto di un lungo percorso nella storia moderna del processamento del linguaggio naturale (NLP) che, partendo a inizio Novecento dalle catene di Markov applicate alla letteratura (sequenza di vocali e consonanti di un romanzo) e passando per i lavori di Shannon e Weaver a metà anni Cinquanta sulla misura dell'entropia e la distribuzione delle probabilità (*n-grams* e sequenza probabilistica di parole nella lingua), arriva a inizio anni Duemila con Bengio e colleghi all'applicazione delle reti neurali artificiali per il processamento del linguaggio naturale (*neural* NLP). Anche con importanti sviluppi recenti come l'impiego dei trasformatori (*transformers*) in grado di incorporare nell'analisi probabilistica del linguaggio la dimensione contestuale delle parole nelle frasi.
È molto importante però comprendere bene tecnicamente qual è il lavoro tecnico-operativo - invisibile ai più - dei modelli linguistici computazionali. E capire la loro relazione e differenza con il linguaggio naturale

umano. Riprendendo le avvertenze di Shanahan (*Talking About Large Language Models*, 2022), quando si interroga un sistema di questo tipo chiedendo di completare una frase (ad esempio, "l'autore della Divina Commedia è …") e ottenendo una determinata risposta ("…Dante"), in questo dialogo noi e la macchina intendiamo due cose molto diverse. Noi vogliamo sapere chi ha scritto nella realtà storica il famoso poema. La macchina, invece, intende "quale parola è statisticamente più probabile che segua nella sequenza della frase "l'autore della Divina Commedia è…"? Dentro gli archivi informativi con cui è alimentato il modello troverà che "Dante" è la parola più frequentemente associata nella sequenza di parole della frase in questione. Nel caso di specie e più filosoficamente, dunque, con la sua interrogazione, l'umano intende dire e chiede di conoscere un elemento di concreta "verità storica" del mondo. Dal suo canto la macchina, invece, intende processare e può solo restituire un risultato di pura "probabilità linguistica" del testo.

Lingua, pensiero, mente e mondo

Tuttavia, e qui è il punto critico, l'umano - preso tra antropomorfismi e sociomorfismi - immagina che la macchina comprenda la domanda e che arrivi alla risposta nello stesso modo in cui fa lui. Dunque, per non rimanere vittime di *hype* (ma anche per non perdere delle opportunità di business), occorre distinguere - come mostra un lungo studio sulla "dissociazione tra linguaggio e pensiero nei LLM" (Mahowald, Ivanova e altri, 2023) - le competenze linguistiche "formali" dalle competenze linguistiche "funzionali". Le prime (quelle formali) si riferiscono alla capacità del processamento macchinico del linguaggio naturale in grado di riconoscere la struttura sintattica di una lingua, le sue regole grammaticali, le sue regolarità nella costruzione delle frasi. E, quindi, poi di riprodurla e simularla probabilisticamente. Le seconde (quelle funzionali) riguardano le capacità proprie del cervello umano di costruire un linguaggio che è in relazione col mondo e che ci consenta cognitivamente di agire in esso impiegando la percezione e i sensi, la comunicazione e gli altri, il ragionamento e le interazioni. I successi raggiunti dai LLM nelle competenze formali non devono trarci in inganno rispetto alle seconde che, ad oggi, rimangono lontane da quelle umane. Da qui anche la necessità e l'importanza di nuove pratiche disciplinari come il *prompt engineering e design*. Interrogazioni, istruzioni, dati, esempi sono di norma gli input impiegati per sollecitare la macchina a produrre, attraverso un modello matematico ottimizzato su token linguistici, l'output desiderato (una conversazione, un testo, un riassunto …). Per una buona produzione dell'output, l'ingegneristica dello spunto (*prompt engineering*) necessita di avere una qualche comprensione del meccanismo/modello impiegato dalla macchina, oltre che una qualche conoscenza del dominio disciplinare di riferimento. In ogni caso, ad oggi potenzialità e meraviglie, ma anche limitazioni, allucinazioni, inventive, errori lessicali, sintattici, semantici e retorici di ChatGPT *et similia* sono conseguenti a questa peculiare modalità operativa di processamento computazionale, probabilistico e simulativo, della lingua. In prospettiva, si stanno però già prefigurando e testando integrazioni di capacità elaborative neuro-simboliche e funzionali nei modelli linguistici a larga scala per ovviare alle attuali, evidenti limitazioni.

Solo pappagalli stocastici?

In questo frangente, qualcuno velocemente viene riproponendo il ban platonico delle arti imitative ("*della cosa imitata l'imitatore non sa nulla che valga nulla*" scriveva Platone nella *Repubblica*) nella sua versione contemporanea degli *stochastic parrot*, dei pappagalli probabilistici, come anticipavo. Altri ingenuamente si stupiscono delle nuove meraviglie tecnologiche simulacrali e del grado di verosimiglianza raggiunto e via via sempre più affinato a superamento di soglie un tempo immaginate invalicabili (e tra l'altro siamo in attesa, dopo GPT-3, di GPT-4 di molte magnitudini superiore). Di volta in volta, l'umano fronteggia questa presa di parola della macchina o con palese sufficienza (non c'è comprensione del senso) o con facile entusiasmo (una svolta nella generazione del linguaggio). Sono tuttavia visioni filosofiche deboli del momento e del passaggio strategico che viviamo perché cercano di depotenziare o banalizzare l'impatto culturale spaesante dell'arrivo dei linguaggi sintetici. Che non riguarda la questione di assegnare e riconoscere o meno intelligenza, coscienza, senzienza alle macchine. Piuttosto e in prospettiva, l'arrivo del "linguaggio sintetico" (come scrivono Bratton e Aguera Y Arcas, *The Model is The Message*) scardina e decostruisce (Gunkel) in profondità gli apparati, i domini e i dispositivi istituzionali del discorso, della parola e del parlante così come della scrittura e dell'autorialità. La presa di parola della macchina sarà un'operazione più profonda e spaesante nel lungo periodo (e *disruptive* su industrie e mercati: dall'educazione all'intrattenimento, dal giornalismo al marketing). Anche le big tech, Google in primis, sono in allarme rosso. Più culturalmente e strategicamente, dobbiamo però marcare meglio questa discontinuità. In primis, il fatto che non ci sia "comprensione di senso" (punto da approfondire e da non dare per già facilmente sciolto) non significa, ad esempio, che non ci sia comunque produzione/circolazione di senso e di impatto per l'umano coinvolto nell'assemblaggio sociotecnico. Il senso circola sempre in qualche forma attraverso l'intelligenza, o non intelligenza, dell'umano che leggerà (anche inconsapevole di ingannarsi sul processo simulativo in atto). La cosiddetta "intelligenza artificiale" non è pensabile *in sé* e *per sé* (come mero artefatto tecnico) come spessissimo viene intesa, ma sempre *con altri* e *per altri* (come assemblaggio sociale). E, qui, antropomorfismi e sociomorfismi sono sempre all'opera con i loro pregi (empatia e efficienza) e i loro rischi (intrasparenza e manipolazione).

Macchine che prendono la parola

D'altro canto, dire che è una svolta nella produzione del linguaggio lascia inesplorata la natura di questa operazione senza precedenti di "strutturalismo sperimentale", come l'ha definita Rees nel suo *Non-Human Words* (2022). Quindi, sostenere a proposito dei LLM che si tratta di meri pappagalli stocastici significa non comprendere la portata culturale epocale di questo passaggio alla "parola non-umana". La prerogativa storica della parola (simulata) ai soli umani mostra segni di cedimento. Passaggio che la teoria letteraria e la filosofia continentale avevano anticipato. Ad esempio, tutta la riflessione sulla "morte dell'autore" con Barthes (*La mort de l'auteur*) e Foucault (*Qu'est-ce qu'un auteur?*) come ci ha ricordato il filosofo Gunkel in una sua serie di post su Twitter a fine 2022. In questa prospettiva, precisa Gunkel, la parola/scrittura della macchina rappresenterebbe la fine dell'autorialità (per come l'abbiamo conosciuta, trasformata e operazionalizzata storicamente finora) e l'inizio di un nuovo percorso/discorso della parola,

del linguaggio, della scrittura, della proprietà intellettuale e così via. Con tutte le sue opportunità e tutte le sue inquietudini, vulnerabilità e rischi. Dunque, continua Gunkel, non sarebbe la fine della scrittura, ma la fine dell'autore (nella sua forma storica attuale). Ma, insieme all'autorialità che entra in questione e in crisi, siamo anche all'avvio più complessivamente di una nuova era inflazionaria della parola (e dei media più in generale). Che, come tutti i passaggi mediali inflattivi, scardina per un verso e istituzionalizza per l'altro nuovi ordini del discorso, nuovi regimi di verità e falsità, nuove logiche e dinamiche di economia politica e di potere. Come ha scritto Jennifer Petersen nel suo *How Machines Came to Speak* (2022) «…molti impieghi dei bot e dell'apprendimento automatico ristrutturano il discorso, riorganizzando le posizioni di chi parla, del testo e del pubblico – e, così facendo, cambiano ciò che significa essere un soggetto parlante ... il momento attuale potrebbe essere un'occasione per ripensare alcuni dei nostri assunti fondamentali sul discorso». La parola è potere. Come direbbe Foucault, in che forme sorprendenti e arrischiate verremo allora parlati dalla nuova lingua sintetica?

Imprese e nuove uncanny valley

Quel che è certo è che con i linguaggi sintetici non siamo di fronte solo a nuovi problemi tecnologici, ma anche e soprattutto a nuove o rinnovate provocazioni culturali e sorprendenti paradossi (tra il dentro e il fuori del testo, tra il linguaggio e la sua relazione col mondo, tra la presa di parola della macchina e l'esperienza dell'umano che viene parlato). E, se i problemi tecnici richiedono una soluzione ingegneristica, le provocazioni intellettuali ci sollecitano piuttosto all'innovazione culturale. Di questa le imprese hanno un urgente bisogno per attraversare, abitare e prosperare in queste nuove *uncanny valley.*

▪ L’immagine sintetica -

Dalla parola all’immagine sintetica

Insieme alla parola, l’immagine sintetica rappresenta l’altra espressione letteralmente più visibile e oggi sempre più presente della capacità generativa dell’AI. Quella del processamento macchinico delle immagini (*image processing*) è stata un’evoluzione storica lunga: scientifica, industriale e artistica insieme. A partire dagli anni Venti del Novecento, è stato un percorso che ha portato l’immagine ad essere prima digitalmente processata e poi, col primo decennio degli anni Duemila, ad essere generativamente sintetizzata. Così, nel tempo, attraverso una serie di discontinuità ontologiche (Nail, *Theory of the Image*, 2019; Thomson-Jones, *Image in the Making*, 2021), quello che chiamiamo “immagine” è stato prima ri-rappresentato con produzioni, strutture e interfacce digitali e poi, infine, da ultimo proprio ri-creato attraverso l’impiego di reti neurali artificiali profonde. Ma cosa rappresentano filosoficamente le immagini sintetiche tipo quelle prodotte dai modelli di diffusione stabile (SDM, *stable diffusion model* come per Stable Diffusion), ma anche quelle create in forme generative varie da DALL-E, Midjourney, Imagen? E come si producono poi tecnicamente? Possiamo iniziare allora da questa ultima domanda. Dunque, qual è l’ingegneria di un’immagine di sintesi?

Le meccaniche di una genAI visuale

Un modello generativo a diffusione stabile ha in genere all’origine un'immagine (interpretata dalla macchina come trasposizione numerica che è poi il “suo” modo di “vedere” il mondo) corrotta e degradata progressivamente iniettando del rumore gaussiano. L’iniezione diffusiva di rumore nei dati dell’immagine continua fino alla distruzione totale della stessa che diviene, a quel punto, interamente rumore (processo di *forward diffusion*). Una volta terminata questa diffusione degradativa dell’immagine scomposta in pixel caotizzati, la tecnica generativa capovolge il processo addestrando invece una rete neurale artificiale a ricreare l'immagine impiegata in ingresso e prima “rumoreggiata” (processo di *reverse diffusion*). Così, attraverso l’operazione di *denoising* (eliminazione del rumore) si procede a invertire la fase di perturbazione al fine di generare inedite immagini a partire dallo stato di rumore casuale. Se il processo di denoising avviene impiegando lo “spazio latente” di un’immagine (come in Stable Diffusion) piuttosto che l’immagine in sè si parla di modello a diffusione latente (LDM o *latent diffusion model*). Come vedremo ora, la potenza inflattiva dell’immagine sintetica deriva da questa capacità macchinica di scandagliare e valorizzare lo spazio latente del dato osservato, ma invisibile all’umano. Così, in un flusso operativo *text-to-image* (dal prompt all’output) il processo macchinico generativo inverte tecnicamente il processo classificatorio. Il modello non classifica immagini date assegnandole ad una categoria (*classifier*), ma dato un input testuale genera (*generator*) una nuova immagine.

Dai token linguistici ai pixel grafici

L’assemblaggio computazionale che genera l’immagine a partire da un testo è variamente composto: *text prompt, tokenization, embedding, text transformer, noise predictor* e molto altro. Ciascuno di questi momenti e tecnicalità del flusso generativo ha funzioni specifiche come, ad esempio, convertire il prompt testuale iniziale in token linguistici comprensibili dalla macchina (che non riconosce le parole umane in quanto tali), ridurre la

dimensionalità rappresentativa vettoriale dei dati ricercandone e preservandone le similarità contestuali (come le prossimità semantiche e di senso), predire il rumore latente nell'immagine latente per poi sottrarlo in maniera iterata e campionata per step (producendo così una nuova immagine latente), trasformare infine l'immagine latente in immagine-pixel e restituirla al prompt iniziale come nuovo prodotto visivo di sintesi. Come si può intuire da questa semplificazione illustrativa, la trasformazione dei "token linguistici" in "pixel grafici" è un'operazione stratificata di assemblaggi algoritmici che prima decostruiscono e poi ricostruiscono in forma nuova un'immagine. In questo modo da un prompt testuale (es. "donna che raccoglie fiori nello stile di Picasso") ma anche sempre più multimodale, si creerà un'immagine visivamente nuova. Questo approccio è destinato strategicamente ad allargarsi a vari domini: voce, suoni e musica, diagnostica medica per immagini, robotica sociale e collaborativa, design industriale per la prototipazione ingegneristica generativa (MIT Technology Review, *Generative AI in Industrial Design*, 2023).

Valore strategico dello spazio latente

Questo nuovo rapporto tra segnale (immagine) e rumore (degradazione) è decisivo. Per un'immagine digitale classica il rumore è il disturbo causato dalla totalità delle varie degradazioni fisiche del segnale. Se in un'immagine digitale si procede semplicemente alla sua rimozione, nell'immagine sintetica (e in particolare nel suo spazio latente) il rumore prima si aggiunge e poi si sottrae. Si procede in questo modo perché è più facile per le reti neurali artificiali ricostruire partendo da una struttura d'immagine degradata piuttosto che costruire da zero. Inoltre, lavorare sullo spazio latente delle immagini (che è ridotto rispetto allo spazio ad alta dimensionalità delle immagini originarie) consente di contenere e efficientare lo sforzo computazionale dell'iniezione di rumore. Naturalmente non è solo una questione di efficienza. È rilevante anche dal punto di vista dell'esplorazione e dell'esercizio artistico ed economico della creatività (*Art in the Age of Machine Learning*, Audry 2021; *Latent Spaces: A Creative Approach*, Yee-King, 2022). Ma è importante anche e soprattutto da un punto di vista più culturale e filosofico. Lo spazio latente è lo spazio che ospita e mappa tutte le dimensioni (*features*) possibili dei dati in input. Sono le dimensioni (*pattern* come colore, angolatura, grandezza, orientamento, ecc.) estratte automaticamente da una rete neurale artificiale addestrata. Per mercati e imprese sarà allora vitale esplorare, competitivamente e filosoficamente, questo "spazio im/possibile dell'inosservato latente" (Accoto). Un concetto di immagine latente si trova anche nel discorso fotografico più classico. Ma la distanza semantica e ontologica tra i due concetti segna un punto di non ritorno. Se l'immagine 'latente' in un processo meccanicamente fotografico era prodotta chimicamente, l'immagine 'latente' in un processo artificialmente generativo è prodotta algoritmicamente.

Arriva l'era dell'immagine di sintesi

Anche da questa veloce ricognizione è evidente che l'immagine sintetica non è più semplicemente una "trascrizione isomorfica del reale" come è, invece, un'immagine fotorealistica (Rodowick). Non è più, cioè, la rappresentazione realistica visuale di oggetti, ambienti o persone reali. Con l'AI generativa (GenAI), l'immagine sta continuando a ritmo accelerato il suo cammino trasformativo verso nuove nature, culture, statuti e domini. L'idea di una "immagine tecnica" (*technical image*, Flusser) o di una "immagine operativa" (*operative image*, Farocki) aveva già cominciato a circolare negli anni passati. Ora diversi saggi in uscita torneranno a riflettere più direttamente su questo passaggio inflattivo epocale all'immagine sintetica. E sulle sue caratteristiche di novità (immagini di macchine solo per macchine, natura operazionale e non rappresentazionale del visivo, finalità mediali simulativo-predittive). Ne cito tre rilevanti. Nel suo saggio in uscita *Operational Images* (2023)

Parikka narra di questa trasformazione prodotta da una visualità divenuta oramai postumana (*post-human visuality*). Zylinska continua il lavoro iniziato con *Nonhuman Photography* nel suo prossimo *The Perception Machine* (2023) analizzando l'impatto delle tecnologie generative nella costruzione delle immagini e nella nostra percezione delle stesse. Anche *Computational Formalism* (Wasielewski, 2023) affronta la questione in particolare con riferimento alle tecniche di deep learning e computer vision nell'arte visiva e alle implicazioni storiografiche ed epistemiche connesse.

L'innovazione culturale necessaria

Non dobbiamo farci trarre in inganno: le immagini future avranno un'ontologia diversa da quella del passato pur restando in superficie, all'inizio almeno, simili a quelle di una volta. Al punto che forse dovremo cominciare a usare anche dei neologismi come "algorealismo" in luogo del più classico "fotorealismo" quando, ad esempio, visualizzeremo volti ultra-realistici di umani inesistenti. Per questo, attraversare l'uncanny valley (sexy e risky insieme) dell'immagine sintetica richiederà uno sforzo culturale. Accogliere dentro le nostre società in maniera sicura, prospera, inclusiva e solidale questi sviluppi tecnologici non sarà semplice. Le vulnerabilità sono molteplici e significative a partire dalla proliferazione dei "deep fakes" (Lyon, Tora, 2023) e, più in generale, delle implicazioni critiche connesse (politica, sicurezza, lavoro per citarne alcune). Come ha scritto Parikka "…*ci sono immagini che principalmente operano; non sono necessariamente rappresentative o pittoriche. Le immagini operative mettono in crisi ciò che è un'immagine nella misura in cui passano dalla rappresentazione alla non-rappresentazione, dal primato della percezione umana di corpi, movimenti e cose alla misurazione, al modello, all'analisi, alla navigazione e altro ancora. Cambiano le scale e i termini di riferimento*…" (2023, preface). Siamo dentro una nuova era mediale inflazionaria, quella dell'AI generativa. E se è vero che le ere mediali inflazionarie (dalla parola sintetica all'immagine sintetica) sono tali non semplicemente perché arrivano nuove tecnologie espansive di produzione e circolazione della conoscenza, ma "*quando la portata della loro rappresentazione del mondo minaccia i confini delle precedenti nozioni culturali di realtà*" (Castillo, Egginton, *Medialogies*, 2017), allora regolamenti giuridici e principi etici non saranno sufficienti. Saranno necessari, ma non sufficienti. Avremo bisogno anche e soprattutto allora di (fare) vera "innovazione culturale".

▪ L’agente autonomo -

GenAI: dalla medialità alla produttività

L’era inflattiva dell’AI generativa sempre più evoca all’orizzonte non solo una nuova ecologia mediale sintetica (testi, immagini, suoni, video), ma più profondamente anche una nuova economia sintetica popolata e animata da ‘agenti autonomi’. Siamo solo agli inizi naturalmente e l’hype è montante, ma la progressiva introduzione da parte di imprese e istituzioni di agenti autonomi artificiali si candida a scardinare e riconfigurare antichi modi di produzione e vecchie divisioni del lavoro. In tutti i settori e le industry. Dunque, sarebbe in arrivo un’armata di agenti computazionali che immagina di (ri)organizzare in modo automatizzato il lavoro necessario a completare compiti assegnati molteplici e articolati (non solo, quindi, a produrre una singola immagine o uno specifico testo come accade con le forme attuali dell’AI generativa). Potremmo sintetizzare, allora, il passaggio come un nuovo orientamento della generative AI dalla medialità alla produttività. I nomi di questi nuovi agenti cominciano a circolare: AutoGPT, BabyAGI, Microsoft’s Jarvis, CAMEL, HyperWrite, AgentGPT, Copilot. La lista è destinata ad allungarsi velocemente. A diverso titolo, si possono etichettare come ‘*autonomous agents*’ (AGE) o anche, con una mia proposta alternativa, *autonomous generative entities*. Nell’era della ‘transazione infinita’ come l’ho definita, l’arrivo degli agenti autonomi consente di avviare la sperimentazione di una *artificial economics* in forme nuove reimmaginando ecosistemi di cocreazione di valore e architetture di business in una logica di servizio *AI agent-based*. Ma cos’è, anzitutto, un agente artificiale autonomo?

L’era emergente degli agenti autonomi

In una definizione di mercato data di recente: “*autonomous agents are programs, powered by AI, that when given an objective are able to create tasks for themselves, complete tasks, create new tasks, reprioritize their task list, complete the new top task, and loop until their objective is reached*” (Schlicht, 2023). Schematizzando e in astratto: dato un determinato obiettivo, un agente autonomo definisce i compiti iniziali attingendo anche alla sua memoria (corta e lunga) e creando sottotask/goal, li mette in esecuzione evocando strumenti e risorse necessari e ne raccoglie i primi feedback, sulla scorta di questi genera nuovi compiti mettendoli in scala di priorità selettivamente per poi continuare a iterare il processo, per cicli migliorativi, fino al conseguimento finale dell’obiettivo. Questo fatto singolarmente (*autonomous agents* intesi sovente come copiloti), ma anche in modalità collettiva (nella forma computazionale di *multi-agents systems*). In un esperimento di Park e colleghi (2023), sono state immaginate ad es. aggregazioni di agenti con coordinamento autonomo emergente. Una ventina di agenti artificiali (con l’input di organizzare un party per San Valentino) hanno iniziato a simulare, in autonomia, varie attività connesse all’evento. Questi agenti autonomi sono: “*computational software agents that simulate believable human behavior. Generative agents wake up, cook breakfast, and head to work; artists paint, while authors write; they form opinions, notice each other, and initiate conversations; they remember and reflect on days past as they plan the next day*". Questa capacità di pianificazione è un tratto caratteristico dell’essere un agente autonomo.

Non solo linguaggio, ma pianificazione?

Dopo il successo nell'individuazione della sequenza di parole (modelli linguistici su larga scala) siamo passati ora all'individuazione della sequenza delle azioni (agenti pianificanti *step-by-step*). Per acquisire questa capacità di pianificazione passo dopo passo, sono state chiave tre dimensioni: a) una qualche capacità di "ragionamento" realizzata in modalità 'catena di pensieri' (*chain of thought*) che indirizza il modello linguistico verso la soluzione cercata; b) una qualche capacità di individuare/eseguire le azioni/sottotask da intraprendere e reiterare in autonomia fino ad arrivare a risolvere il compito assegnato quando le informazioni prodotte dal primo prompt non siano sufficienti e siano necessarie ulteriori azioni e osservazioni; c) una qualche capacità di mettere in priorità e dare un ordinamento sequenziale progressivo (incluse dipendenze e concatenamenti relativi tra i vari task) orientato verso il completamento del compito. Dunque, *reasoning* e *acting* al modo della macchina sono al centro di questa nuova intelligenza generativa agente. Questo è un cambio significativo nella storia della programmazione: *"The real unlock that makes agents an entirely new software paradigm lies in the modern LLM's ability to take in a goal, along with a set of facts and constraints, and then* ***create a step-by-step plan*** *for achieving that goal. Before LLMs, the* programmer *had to make the plan - a computer program is really just a step-by-step set of actions the machine will need to take to accomplish a goal. But in the LLM era, machines' newly acquired ability to make their own plans has everyone in a frenzy of either fear or greed"* (Stokes, 2023). In prospettiva più astratta e filosofica, il *prompt engineering* è una forma immanente di *chaos engineering*. Così una "catastrofe crossentropica" -antropomorficamente nota come 'allucinazione'- si mostra essere un sorprendente e illuminante *point-of-failure* (POF) dell'assemblaggio utente-agente.

Dissezionando l'anatomia di un copilota

Ma cosa significa "ragionamento" e "pianificazione" nel caso degli agenti autonomi? Anche qui, per evitare facili e fuorvianti antropomorfismi e sociomorfismi è bene entrare un po' nella loro meccanologia. In primo luogo, quello che chiamiamo *agent* è, in realtà, un assemblaggio distribuito, stratificato e coordinato di funzioni/agente molteplici (ad es. *execution agent, task creation agent, context agent, prioritization agent*), ciascuna incaricata di effettuare specifiche operazioni e di attivarsi e dialogare iterativamente e ricorsivamente con le altre tra strumenti, risorse, memorie, istruzioni (Wang, 2023). In secondo luogo, la dimensione del *reasoning* nei LLM è considerata una proprietà emergente della 'catena di pensiero' (*chain-of-thought* o COT), il meccanismo metacognitivo con cui l'utente umano conduce l'agente artificiale a discutere sempre linguisticamente intorno all'input iniziale, ma per intermedi piccoli passi (*let's think step by step*). In terzo luogo, la dimensione dell'*acting* dell'agente è nella sua capacità di auto-espandere e auto-riproporsi il prompt d'inizio con integrate osservazioni, spiegazioni, suggerimenti. Così facendo, l'agente autonomo ricorsivamente affina l'input/prompt muovendosi linguisticamente nella direzione cercata (Wang, 2023; Stokes, 2023). È bene avere chiare queste tecnicalità per evitare hype, delusioni o fraintendimenti. Quello che è importante qui filosoficamente rilevare, come ha ben scritto Yuk Hui nel suo saggio *Recursivity and Contingency*, è che: "*Contrary to automation considered as a form of repetition, recursion is an automation that is considered to be a genesis of the algorithm's capacity for self-positing and self-realization*". Richiamando Bateson, ci ricorda anche che la nozione di ricorsività è centrale nella definizione di "autonomia" di un sistema.

Tra code economy e artificial economy

L'avvio di questa fase nuova dell'AI generativa agente impatterà su modi di produzione e dinamiche di organizzazione. Il senso e la forma dell'esperienza d'impresa vivono, lo sappiamo, una morfosi profonda. Tra dati, algoritmi e protocolli, le trasformazioni organizzative innescate dall'irrompere della *code economy* che apre ed evolverà in *artificial economy* (Mercado, 2021) si sono di fatto appena avviate. Ridisegneranno imprese e mercati, strategie e leadership, competenze e comportamenti. Nell'era imminente degli ecosistemi di servizi, dei marketplace a piattaforma, dei business multilaterali, dei criptosistemi su reti decentralizzate, la cocreazione di valore si configurerà sempre più come un infinito processo *catallattico* (di scambio automatizzato anche via agenti di servizi-per-servizi), *simpoietico* (di coevoluzione con integrazione di risorse operanti e operande nell'assemblaggio utente-agente) e *prolettico* (di predizione e anticipazione a feedforward di bisogni, necessità, volontà di beneficiari umani e non umani). In questo nuovo orizzonte (Accoto 2021), l'arrivo degli agenti autonomi ridefinisce e rilancia, in forme sorprendenti, la storia più lunga dell'economia artificiale. "*What is an artificial economy? It is a computational representation of an economic system, which allows us to simulate the interaction of artificial agents. Artificial agents are the basic units that make up an artificial economy. These agents are computational objects containing information and rules for processing it. They can deploy very simple and silly behavior, or display sophisticated forms of artificial intelligence*" (Mercado, 2021).

Neoautomazione: mani, menti, mercati

Dal simulare virtualmente un comportamento economico con vite artificiali all'attivare generativamente un'economia artificiale con agenti computazionali, la cosiddetta *machine economy* sta così proseguendo il suo percorso dentro le civiltà umane. Tra robot che producono, agenti che pianificano, dati che quantificano, sensori che controllano, protocolli che disintermediano, la neoautomazione si conferma forza dirompente di trasformazione planetaria. Come ha scritto il filosofo Benjamin Bratton (2021): "*we define automation not just as the synthetic transference of natural human agency into external technical systems, but as the condition by which action and abstraction are codified into complex adaptive relays through living bodies and non-living media. It is both a direct physical ripple and an association of semiotic signaling with its reception; it includes language as well as mechanical information storage and communication. This more ecological conception of automation is one of the conditions revealed by the contemporary intensification of artificial algorithmic intelligence today. It speaks to the already entangled condition of our species, agency, industry, and cultural dramas more than it does to the contemporary concern of proper humans being improperly replaced by machines".* Così, l'automazione presente e prossima si dispiega oggi all'incrocio di tre stratificazioni ingegneristiche: è meccanica, è algoritmica, è protocologica. Ho connotato questa come l'automazione delle 3M ovvero: mani, menti, mercati. Un'automazione che si avvia a creare, conservare e circolare valore digitale in modalità neo-automatizzate e neo-aumentate sorprendenti quanto arrischiate. Ma che si avvia anche a esperire e agire il mondo in forme nuove.

Della cognizione sintetica del mondo

Chi e come conosce oggi il mondo? L'intelligenza umana incorporata ha dominato il campo dell'osservazione e della cognizione del reale per lungo tempo reclamando a sè il primato della conoscenza del mondo. Lo ha fatto, almeno alle nostre latitudini, escludendo da queste possibilità animali e piante (eludendo, tra gli altri, il tema complesso delle 'altre menti' o delle 'menti possibili' oltre a quello delle 'reti di attanti'). L'arrivo della cosiddetta intelligenza artificiale sembra aver nuovamente risvegliato questa supponenza umanocentrata. Almeno secondo due prospettive. Una prima tendenza incarna un umanesimo impaurito o arrabbiato. La seconda è, per converso, superficialmente entusiasta e tecnicamente galvanizzata. Quello timoroso o infuriato, a seconda dei casi, è un umanesimo che di fatto si ripiega su di sé (teso com'è a ricercare la sua essenza distintiva unicizzante e ad aggrapparsi a quella per cercare di preservare un nucleo eccezionale fondativo dagli attacchi della tecnica). Quello acclamante ed eccitato è viceversa un umanesimo in molti casi vittima in/consapevole dei propri antropomorfismi e sociomorfismi (si limita, cioè, a leggere l'AI come rispecchiamento/simulazione delle proprie capacità o, meglio, di quello che si pensa essere il modo dell'umano di percepire, osservare, conoscere, agire il mondo). Entrambe queste prospettive sono deboli, pregiudizievoli e limitate e di parte. Ad osservare, conoscere e operare il mondo non è più solo "l'umano" (se mai fosse stato l'unico). Forme di osservazione, di cognizione e di esecuzione 'more-than-human' (macchiniche, rizomatiche, prolettiche, simpoietiche, chimeriche, alterosomiche) sono da qualche tempo in divenire e in dispiegamento. Ci chiedono di allargare il nostro sguardo oltre i confini culturali del naturale, del nativo, dell'autentico, dell'essenziale, del corporeo umanomorfico. Finanche nella fisica quantistica, l'idea così centrale dell'osservatore del sistema da intendere come primariamente umano è stata oggetto di critica e chiarimento. Lo ha ribadito di recente (IAI 2022) il fisico Carlo Rovelli: <*The notion of "observer" should not be misunderstood. In quantum physics parlance, an "observer" can be a detector, a screen, or even a stone. Anything that is affected by a process. It does not need to be conscious, or human, or living, or anything of the sort ... In the example of the process where you kick a ball and break a window, the "observer" is the glass of the window. It is the physical thing that is affected by the process. In this general sense, the notion of "observer" plays a role. It's not a human observer, it is the physical system affected by a phenomenon*">. Se proprio abbiamo bisogno di un umanesimo, questo dovrà saper promuovere al suo meglio l'innovazione culturale (non la restaurazione culturale).

L'idea ingenua dell'umano nel loop

Parlando di umanesimo, c'è un'espressione umanamente accomodante e però anche filosoficamente ingenua che circola nei discorsi più comuni intorno alla nostra relazione -culturalmente e ideologicamente dicotomica- con le macchine. L'espressione è *human-in-the-loop* (HITL). In genere è tradotta e intesa come mantenere "l'umano nel ciclo". Prevede alcune varianti (come "on-the-loop") e naturalmente anche il suo contrario ("off-the-loop"). Il suo senso più generale è quello di un invito a policy maker, service designer, software developer, brand marketer a lasciare sempre all'umano la decisione e il controllo ultimi su apparati, architetture, dispositivi, macchine. E più complessivamente sui processi dell'automazione spinta delle cd. intelligenze artificiali. Il discorso ha molteplici valenze: politiche, economiche, morali, legali. Prospettiva da molti condivisa, ma superficiale e ingenua direi. Di fatto, l'umano è già sempre presente nel processo di costruzione, produzione e uso delle tecnologie: quando progetta, testa, addestra, seleziona, usa, corregge, etichetta e così

via. Ma direi di più: essendo l'assemblaggio umano-macchina sempre un sistema o reticolato sociotecnico, l'umano è sempre in controllo sia che sia "in-the-loop" (che qualche altro umano lo abbia incluso nel ciclo) sia che sia "off-the-loop" (che qualche altro umano lo abbia escluso). Ma dire questo non è ancora dire tutto. Questo storico assemblaggio umano-macchina, di volta in volta aggiornato, con le sue nuove articolazioni, stratificazioni e distribuzioni cambia i modi di produzione della conoscenza e di divisione del lavoro tra le parti. In effetti, la cd. AI è insieme sorprendente λόγος, ma anche nuovo ἔργον. Non solo conoscenza (discorso), ma lavoro (opera). E determina anche, più profondamente e per ora invisibilmente, uno spostamento di potere e di agenti, di modi e luoghi dell'esercizio del decidere. Dire, allora, che la decisione rimane umana è una naïveté culturale e filosofica. È una prospettiva aspirazionale che rimane in superficie e non coglie il movimento dialettico del reale. Il denso lavoro ricostruttivo sulla storia della "computer vision" di Dobson appena uscito (2023) erode la superficialità di questa narrativa (ideologia) accomodante mostrando tutta l'articolazione di questo assemblaggio umano-macchina e di come si sia trasformato nel corso del tempo in questa coevoluzione (del vedere, del conoscere, del decidere). Coevoluzione che come quella biologica tra le specie ha forme diverse: mutualismo, parassitismo, competizione, predazione. E solo una di queste porta beneficio ad entrambe le specie coinvolte. Negli altri casi, il gioco è sempre arrischiato quando non proprio mortale. C'è, dunque, un modo filosoficamente digiuno e politicamente ingenuo di guardare all'intelligenza artificiale. È quello strumentale (si tratta solo tecnologia), dicotomico (noi umani vs le macchine), antropocentrato (tenere l'umano nel loop e in controllo), allineante (rispetto dei valori umani) e dominante (all'umano spettano le decisioni) di un certo umanesimo. Si accompagna sovente anche ad una facile e consolante <deriva anestetica dell'etica> (Accoto). E ce n'è, invece, uno più filosoficamente educato e planetariamente avvertito che interpreta in maniera complessa e sofisticata il passaggio epocale di cui stiamo facendo esperienza. Un umanesimo quest'ultimo capace di cogliere lo statuto di provocazione culturale e intellettuale dell'AI nella lunga durata delle civiltà umane e di leggere il momento attuale con l'occhio lungo dell'evoluzione culturale e cognitiva planetaria e finanche cosmica, direi. E in questa evoluzione, l'arrivo dell'AI.

L'AI né prodotto né servizio, ma fabbrica

Che non è un prodotto e neppure un servizio, come si dice spesso. Piuttosto, è un assemblaggio sociotecnico cognitivo-produttivo che istanzia e orchestra nuove forme (e dinamiche) di organizzazione della produzione e, dentro queste, le nuove forme (e dinamiche) della futura divisione del lavoro. Come ho scritto (su HBR) e ripeto qui in conclusione, la cd. intelligenza artificiale non è mai 'in sé' e 'per sé' (puro artefatto strumentale), ma sempre 'con altri' (è sistema sociotecnico) e 'per altri' (è costruzione sociomorfica). Così, mi sembra di poter dire che le controversie correnti sul governo dell'AI sono filosoficamente digiune e politicamente miopi. A fronte delle arrischiate uncanny valley che stiamo attraversando (e, in una qualche misura, anche già abitando), si sollevano richieste accorate -quando non proprio allarmate- di mettere a governo (giuridico, etico, economico, sociale, politico…), di volta in volta, il dato, l'algoritmo, il modello, il codice, il protocollo. A questo processo di 'reificazione' (vale a dire, l'isolare e il ridurre l'AI a cose e, in particolare, a mere tecnicalità), fa da contraltare, la 'personificazione' (cioè, l'antropomorfizzare macchine e interfacce assegnando loro, di volta in volta, intelligenza, coscienza, senzienza e così via). Sono entrambi approcci insufficienti, per l'appunto, che non colgono l'articolazione assemblativa che, ideologicamente, chiamiamo intelligenza artificiale. Che, come dicevo sopra, non è solo sorprendente λόγος, ma anche nuovo ἔργον. Non solo conoscenza, ma lavoro. Che non riguarda solo l'umano con i suoi antropomorfismi e sociomorfismi, ma anche il nostro orizzonte sempre

più more-than-human. Un orizzonte intricato fatto di assemblaggi, perturbazioni, aggrovigliamenti, speciazioni, automazioni, reticolati, derivati, stratificazioni, distribuzioni che turbano il sonno di molti. Anche il discorso regolatorio e giuridico comincia a prendere consapevolezza che occorrerà dotarsi di sguardi perturbati e inconsueti per riuscire a cogliere un divenire poco familiare. Di quanta innovazione culturale saremo capaci? Quanto riusciremo ad essere aberranti nel nostro pensare il nuovo mondo?"

Una nuova (sexy/risky) uncanny valley

Per tornare al nostro tema degli agenti autonomi, certamente, a leggere strategicamente il presente, i limiti ad oggi di un'automazione generativa agente sono ancora significativi. Ancora in buona misura basati sul processamento macchinico del linguaggio umano (solo formale, non funzionale), gli agenti risentono di questa simulazione del linguaggio e di non comprensione del mondo. Anche con le loro "allucinazioni" (o "catastrofi crossentropiche" come ho proposto di rinominarle per evitare facili e ingannevoli antropomorfismi). Limiti che si ripercuotono su ragionamento e pianificazione ovviamente. Di fatto, non sono macchine di *reasoning* e *acting* native, ma linguistiche di base, integrate variamente. A questo si aggiungano, i limiti degli strumenti, delle risorse, dei processi a cui devono attingere per poter completare i diversi compiti a cui vengono chiamate oltre a quelli delle vulnerabilità dei rispettivi modelli linguistici fondativi (dalle memorie computazionali alle fonti informative e così via). Nonostante ciò, sono in diversi oggi a ritenere l'orizzonte degli agenti autonomi tanto probabile quanto promettente. Non c'è dubbio: stiamo entrando nell'era dell'*hyperautomation* e dei *digital worker*, umani e non (Wilson, 2023). Con l'arrivo di questo 'lavoro sintetico' (dopo parola sintetica e immagine sintetica), si aprirà dunque una nuova uncanny (sexy/risky) valley tutta ancora da esplorare nelle sue potenzialità e vulnerabilità. Dovremmo oramai saperlo bene. La tecnica (che non è mai solo ingegneria macchinica, ma sempre anche economia politica) vive -e non sfugge a- una paradossale esistenza "farmacologica". È veleno e antidoto come dicono i filosofi. O, per risalire al mito greco di Prometeo, è insieme donazione e punizione della divinità, dono e danno. È al contempo, come scriveva Machiavelli a proposito della politica, "ruina" e "remedio". Allertati dalla sapienza degli antichi (*timeo Danaos et dona ferentes*), a noi spetta oggi nuovamente il compito di coltivare con speranza, in maniera solida e solidale, questo dono difettato.

Referenze

Accoto, *Il mondo ex machina* (2019)

Accoto, *Mani, Menti, Mercati*. In Bordoni, *Il primato delle tecnologie* (2020)

Beer, *The Tensions of Algorithmic Thinking* (2022)

Bratton, *The Terraforming* (2021)

Hui, *Recursivity and Contingency* (2019)

Mercado, *Artificial Economics. Methods, Models, and Interdisciplinary Links* (2023)

Li, *Language Models: Past, Present, Future* (2022)

Shanahan, *Talking About Large Language Models* (2022)

Rees, *Non-Human Words: On GPT-3 as a Philosophical Laboratory* (2022)

Bratton, Aguera Y Arcas, *The Model is The Message* (2022)

Gunkel, *ChatGPT is the event that 20th century continental philosophy had been preparing us for* (twitter post, 2022)

Petersen, *How Machine Came to Speak* (2022)

MIT Technology Review, *Generative AI in Industrial Design* (2023)

Nail, *Theory of the Image* (2019)

Thomson-Jones, *Image in the Making* (2021)

Audry, *Art in the Age of Machine Learning* (2021)

Yee-King, *Latent Spaces: A Creative Approach* (2022)

Parikka, *Operational Images* (2023 forthcoming)

Zylinska, *The Perception Machine* (2023 forthcoming)

Wasielewski, *Computational Formalism* 2023 (forthcoming)

Castillo, Egginton, *Medialogies* (2017)

Navas, *The Rise of Metacreativity* (2023)

Lyon, Tora, *Deep Fakes* (2023)

Nyholm, *This is Technology Ethics* (2023)

Accoto, *Imprese, piattaforme, ecosistemi e …quantum stack*? In Besana, *Future of Work* (2021)

Park et alii, *Generative Agents: Interactive Simulacra of Human Behavior* (2023)

Schlicht, *The Complete Beginners Guide to Autonomous Agents* (2023)

Stokes, *AI Agent Basics: Let's Think Step by Step* (2023)

Tella, *The New Automation Mindset* (2023)

Wilson, *Age of Invisible Machines* (2023)

Yao et alii, *React: Synergizing Reasoning and Acting in Language Models* (2023)

Wang, *The Anatomy of Autonomy* (2023)

Dobson, *The Birth of Computer Vision* (2023)

Rovelli, *Consciousness is Irrelevant for Quantum Mechanics* (intervista, 2023)