DOPPIOZERO

Da Dalì a DALL-E2: creare i creatori

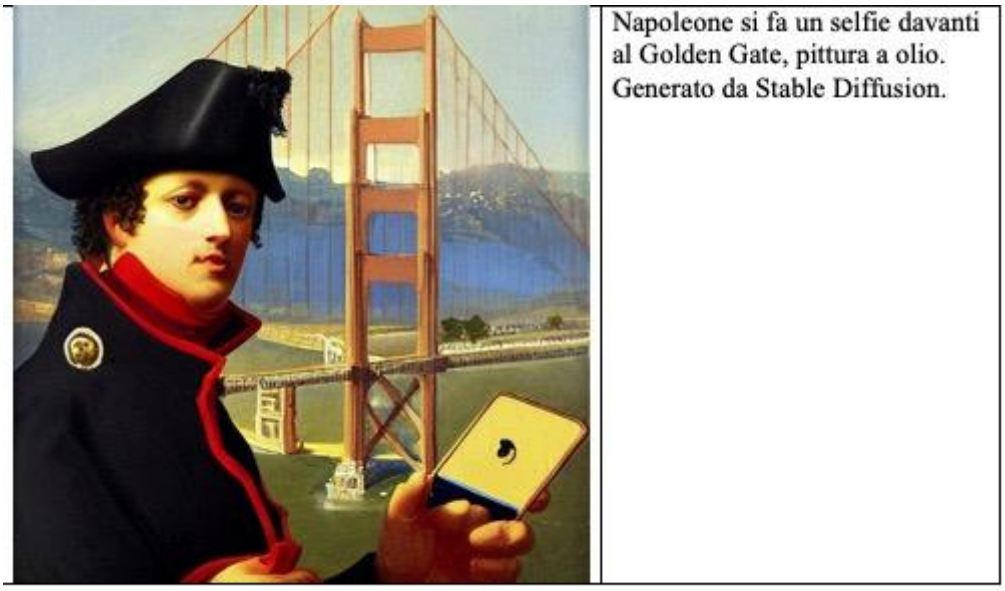
Riccardo Manzotti

19 Settembre 2022

Il 2022 sarà ricordato come l'anno in cui è cambiato il rapporto tra gli esseri umani e le immagini. L'intelligenza artificiale è diventata capace di creare immagini a partire da una semplice descrizione verbale. Questa nuova tecnologia si chiama, nome piuttosto prosaico, TTI ovvero "Text To Image". Per esempio, voi scrivete "Napoleone si fa un selfie davanti al Golden Gate" e voilà, la TTI produce, come per magia, proprio quella immagine nello stile del pittore scelto o persino in uno stile che non corrisponde a nessun pittore in particolare.

Questa tecnologia è letteralmente esplosa nelle ultime settimane ed è diventato uno degli argomenti principali sulla rete (da *Time* al *New York Times*). Il confine tra creatività umana e intelligenza artificiale è stato attraversato. La TTI mette alla portata di qualsiasi utente la possibilità di creare un numero praticamente infinito di immagini. Fino a qualche settimana fa, per corredare un testo con illustrazioni o con foto, si sarebbe dovuto ricorrere o a materiale di archivio o all'intervento di professionisti umani (magari mediati da siti come *Fiver* o commercializzati da grandi archivi online privati a pagamento). Oggi, in molti casi l'intervento umano non è più indispensabile, non è più necessario. I vari siti (qualcuno gratis e qualcuno a pagamento) che mettono a disposizione la TTI sono in grado di generare immagini di ogni tipo.

Ovviamente questa tecnologia è ancora acerba. Sicuramente la TTI mostra dei limiti. Non tutte le immagini sono soddisfacenti o belle o azzeccate. La formulazione della descrizione verbale rimane un elemento critico non facilmente gestibile. E tuttavia, al netto di tutti questi fattori limitanti, rimane il fatto incontestabile che la TTI è in grado di generare in pochi secondi immagini nuove che non sono mai state né catturate né create da esseri umani. È qualcosa che, se ci si pensa bene, mette i brividi. La TTI manifesta capacità che si riteneva fossero una prerogativa dell'essere umano e rappresenta una sfida per gli esseri umani impegnati in attività tradizionalmente considerate al riparo dall'avanzata della razionalità artificiale, come l'arte e la creatività.



Come riesce la TTI a produrre immagini sulla base di una semplice descrizione verbale? La magia è resa possibile dalla combinazione di due fattori: reti neurali (*deep learning*) e grandi quantità di dati (*big data*). In pratica, senza entrare in tecnicismi, grazie a Internet oggi l'Intelligenza Artificiale ha accesso a miliardi di immagini (dalle foto che carichiamo sui social fino alle raccolte di artisti in cerca di visibilità come DeviantArt). Grazie al fatto che queste immagini sono spesso corredate da descrizioni verbali (le immagini sono "taggate"), una intelligenza artificiale dotata di *deep learning* può leggere queste gigantesche collezioni (dette *dataset*) ed estrarre le caratteristiche strutturali di base delle immagini (per esempio lo stile di Van Gogh o la forma dei visi). Alla fine, il sistema è in grado, grazie a un algoritmo detto di *diffusione latente* (ultimo termine tecnico), di generare immagini e capire quanto siano coerenti con le strutture apprese e con la descrizione verbale proposta dall'utente (il *prompt*). A partire da un seme causale, si possono generare infinite immagini che si evolveranno fino a corrispondere con il testo fornito.

A partire da questa descrizione un po' impressionistica si capisce che un fattore chiave sia la collezione di immagini su cui il sistema apprende (il dataset) che è il fattore che differenzia più di ogni altro i sistemi TTI oggi disponibili. Tra i più famosi si devono citare Dall-E2, Midjourney e Stable Diffusion. I primi due sono privati, mentre l'ultimo è stato messo a disposizione di chiunque voglia utilizzarlo (è anche possibile utilizzare delle demo online). Dall-E2 prodotto dalla OpenAI, deve il suo nome a una crasi tra il nome dell'artista Salvador Dalì, noto per le sue opere oniriche e fantasmagoriche e il film della Pixar, Wall-E.

La collezione di immagini di partenza ha delle ovvie conseguenze su quello che il sistema sarà in grado di generare e infatti i tre sistemi TTI creano immagini

diverse proprio perché sono stati “cresciuti” con tradizioni iconiche diverse. Sono nomi che possono sembrare arcani e sibillini, ma se li cercate in rete vedrete che sono diventati di dominio comune. È facile perdersi nell’esplorazione di uno dei tanti siti dedicati alle TTI (per esempio sulla sezione di Reddit dedicata a DALL-E2) saltando da un’immagine all’altra in una cornucopia infinita di combinazioni grafiche, una biblioteca iconica di Babele del possibile, un giardino delle delizie digitale. Ogni cosa si trova rappresentata in ogni modo possibile (o quasi). Eppure tutta questa bulimia pittorica deriva dalle immagini di partenza.

La domanda d’obbligo è se il motore di tutto, l’intelligenza artificiale, sia dotato di vera creatività oppure no? Molti risponderebbero di no. Conosciamo gli algoritmi, sappiamo che i sistemi TTI non conoscono il significato di quello che creano e siamo in grado di vedere la dipendenza tra le loro creazioni e i dataset di partenza. Eppure, la risposta non è banale. Che elementi esistono per sostenere che la creatività umana sia basata su un principio qualitativamente diverso? In fondo, i sistemi TTI non copiano e incollano, come potrebbe fare Google o un grafico principiante, ma estraggono le caratteristiche stilistiche e iconiche contenute nelle immagini a loro disposizione e ne generano di nuove a partire da una partenza casuale. Si tratta di vera creatività? Qualcuno potrebbe negarlo e dire che quello che fa l’intelligenza artificiale non è altro che un modo statisticamente sofisticato di copiare le opere degli esseri umani. Ma è proprio così? In fondo, se guardiamo alla storia dell’arte, non assistiamo a una, più o meno, continua evoluzione a partire dalle opere dei predecessori?

Ogni generazione aggiunge qualcosa di nuovo, ma allo stesso tempo declina l’eredità del passato. Ogni artista umano è immediatamente collocabile nel suo periodo storico, anche senza conoscerne il nome. Anche gli esseri umani dipendono dai loro dataset, che chiamiamo cultura, contesto, tradizioni. Certo, dovremmo distinguere tra artisti e illustratori, tra Marcel Duchamp e Alexandre Cabanel, ma anche così non è facile trovare un principio che non sia riducibile a complessi processi causali innestati su tradizioni culturali. In fondo anche l’evoluzione, che ha prodotto una miriade di organismi e strutture, non è altro che variazione, selezione e trasmissione.

Certo, ridurre la creatività a casualità ben organizzata sarebbe un brutto colpo per il narcisismo umano che, come diceva Freud, è già stato detronizzato dal cosmo con Copernico e dai viventi con Darwin. I sistemi TTI gettano una luce fredda sulla nostra capacità più preziosa e, finora, meno riprodotta: la creatività estetica. Sospendiamo il giudizio e consideriamo altri aspetti di questa rivoluzione (magari metafisicamente meno ambiziosi, ma praticamente più urgenti): la natura

delle immagini, l’inflazione del significato, l’impatto sul lavoro, gli aspetti etici e il copyright. Vediamole in rapida successione per poi rivelarne la radice comune: il concetto di umano.

Fino a qualche settimana fa, nel mondo degli studiosi di cultura visuale e delle immagini, era prassi distinguere tra immagini prodotte dalla creatività umana usando (di fatto) la mano – come quadri e disegni – e il mondo delle immagini prodotte attraverso meccanismi più o meno automatici (dalla fotografia a Photoshop). Ovviamente esistono infinite gradazioni (come il rendering 3D) che rendono difficile tracciare distinzioni nette. Oggi questa differenza, raccontata molto bene nel classico volume di Susan Sontag (La fotografia), è stata resa ancora più ambigua. I sistemi TTI non producono fotografie, ma, di fatto, generano nuove immagini grazie a quanto hanno appreso dalle immagini a disposizione sulla rete. Le immagini non sono più tali e il loro ruolo andrà rivisto.

La moltiplicazione delle immagini (quasi demoniaca) ha una conseguenza che potremmo sintetizzare con il termine “l’inflazione del significato”. La moneta cattiva scaccia la buona e la facilità nel creare figure ne abbatte il significato. È come se l’immagine digitale corrispondesse a una finanza impazzita senza più alcun legame con l’economia reale (la fonte del significato). Cento immagini digitali di un corpo nudo, ma un unico corpo nudo reale (o magari neppure uno). Dove è finito l’eros? Che cosa dovremmo concupire?

Inoltre, è paradossale che questa miriade di immagini prodotte dall’intelligenza artificiale non siano viste da nessuno, e quindi non siano nemmeno immagini, fino al momento in cui qualche utente non le vede. In effetti, si potrebbe sostenere che, almeno in potenza, una volta che una intelligenza artificiale sia stata addestrata, tutte le immagini possibili siano già al suo interno: quelle lecite e quelle non lecite. Il testo fornito dall’utente, in senso proprio, non creerebbe qualcosa, piuttosto la selezionerebbe all’interno di un campionario potenziale.

Se queste immagini non hanno significato per chi le ha prodotte, non le hanno finché qualche utente non glielo attribuisce. Per esempio, e ci avviciniamo al problema etico, se un utente chiedesse a DALL-E2 di creare una immagine blasfema o offensiva, sarebbe tale fino a che non venisse usata da qualcuno a quello scopo? Il problema non è solo teorico, ma pratico. Finora, per evitare grane legali, i sistemi proprietari (DAELL-E2 e Midjourney) sono limitati da filtri che bloccano la produzione di contenuti imbarazzanti, mentre Stable Diffusion si limita a essere distribuito con una nota che esorta a usarlo per fini legittimi. Ovviamente molti hanno interpretato questa nota in modo eccessivamente restrittivo e lo stanno usando per produrre immagini di ogni tipo (satira politica,

deep fake, pornografia e simili). Ma non si può che interrogarsi sulla effettiva negatività di immagini prodotte in questo modo.

In fondo si tratta di creazioni digitali che non hanno alcuna radice diretta con quello che rappresentano. Come le fantasie verbali di De Sade, non sono state composte attraverso delitti o ledendo la libertà di altre persone; sono solo fantasie e, a differenza di quelle del grande Marchese, sono state generate da meccanismi privi di consapevolezza. Come possiamo considerare offensiva una immagine digitale che è stata prodotta attraverso dei meccanismi privi di significato.

Per di più, i sistemi TTI generano qualsiasi immagine o quasi, ma non ne comprendono il significato. Quindi, anche se il processo creativo artificiale presentasse analogie con quello umano, non ne comprenderebbe il significato. Scherzando potremmo dire “Signore, perdona Dall-E, Midjourney e Stable Diffusion perché non sanno quello che fanno”! Sono sistemi complessi, ma privi di consapevolezza. Per ora, si comportano in modo analogo ai processi alla base dell’evoluzione che, anche se sono in grado di produrre l’ala piumata dell’Archaeopteryx o il sistema nervoso umano, lo fanno senza finalità cosciente. L’orologiaio – come in un libro di Richard Dawkins – se c’è, è cieco.

È plausibile che la capacità dei sistemi artificiali di generi infinite immagini porti alla inflazione digitale, cioè alla liberazione delle immagini dal giogo pesante del loro significato. Quindi un’immagine di un angelo umiliato non soffre e non offende gli adoratori dell’angelo. Inoltre l’angelo e la “sua” immagine non hanno più alcun legame necessario, ma solo una somiglianza contingente e arbitraria. Anzi l’immagine dell’Angelo (o di Biden, Madonna, Carlo III) non sarebbe più “sua”. Ogni immagine non sarebbe più l’immagine “di qualcuno”, ma solo un’immagine che, in modo del tutto arbitrario, potrebbe assomigliare a qualcuno. Il meccanismo alla base dei sistemi TTI è privo, per così dire, del peccato originale della fotografia, la presunta copula con la realtà che si pensava si consumasse nell’atto penetrativo dell’obiettivo. L’intelligenza artificiale crea un giardino dell’Eden iconico dove ogni cosa è innocente.

L’ultimo punto, su cui correrò, è il colpo mortale che questi sistemi, nelle prossime versioni, assesteranno non solo alla nostra autostima, ma anche al mercato della creatività minore. Non stiamo parlando dei grandi creativi, gli artisti con la A maiuscola, ma di tutti quei piccoli e medi grafici e illustratori che vivono dell’esercizio di una sensibilità e abilità estetica che soddisfa i bisogni quotidiani della comunicazione e della produzione. Per loro questa rivoluzione rischia di essere la campana a morto della propria professionalità. Oltre all’aspetto

economico ci sarà anche l’inevitabile ricaduta nella formazione e nell’educazione umana.

Nel momento in cui Photoshop – ma anche Word, Powerpoint e Whatsapp – ingloberanno questa tecnologia, chi troverà più il tempo e la voglia di mettere insieme una illustrazione originale, anche se non perfetta e ingenua nell’esecuzione? Già mi immagino che la prossima versione di office permetterà a tutti di dare sfogo alla propria creatività; mentre non farà altro che atrofizzare quasi completamente le capacità già ridotte di autori, studenti e comunicatori.

A parte l’impatto economico su alcuni settori lavorativi, la Text-To-Image (TTU) non potrà che ridurre ulteriormente lo spazio proprio della mente umana esternalizzando, un pezzo alla volta, la creatività umana. Certo, alcuni autori (penso anche agli italiani Francesco D’Isa e Lorenzo Ceccotti) hanno analizzato puntualmente questo fenomeno e ne hanno evidenziato i forti limiti. Hanno in gran parte ragione, ma solo per ora. La generazione di immagini a partire da semplici testi è lo scoperchiamento di un vaso di Pandora che non potremo richiudere tanto facilmente. Come nell’apprendista stregone di Walt Disney, i nostri assistenti digitali si moltiplicheranno oltre le nostre aspettative. Dall-E2, Stable Diffusion e MidJourney sono solo le avanguardie di una marea che metterà in discussione la nostra essenza di esseri umani: significato, consapevolezza, creatività. Per anni ci siamo cullati nella sicurezza di essere fatti a immagine e somiglianza del creatore. Oggi creiamo dei creatori. Siamo ancora necessari?