



26 SETTEMBRE 2023

# DA DALL-E A MIDJOURNEY

## Guida alla scelta dei tool Text to Image

**Wavemaker**

# Guida alla scelta dei tool Text to Image

Il nostro viaggio per comprendere gli ambiti dell'intelligenza artificiale generativa (GenAI) continua. Dopo aver esplorato il mondo del "Text to Text" con **ChatGPT** e i suoi epigoni, questa volta ci addentreremo in quello del "**Text to Image**". Sarà un percorso affascinante, alla scoperta di strumenti quasi magici che permettono a chiunque di generare delle immagini, scrivendo delle semplici descrizioni testuali.

## L'IA dietro la "magia"

Era il gennaio del 2021 quando **OpenAI** svelò al mondo l'esistenza di **Dall-E**, un software di generazione sintetica delle immagini addestrato sulla base del modello linguistico **GPT-3**. Da quel momento, sono arrivati sul mercato una miriade di prodotti che hanno puntato a rendere possibile un sogno di molti: **riuscire a realizzare disegni, illustrazioni e foto usando solo le parole**.

## Ma come funzionano questi software text-to-image?

Nella **fase di apprendimento** il sistema, che utilizza reti neurali artificiali, inizia studiando **miliardi di immagini annotate**, ossia accoppiate ad una descrizione del contenuto e dello stile. In una prima fase del processo, queste immagini vengono gradualmente offuscate aggiungendo elementi estranei, noti come "**rumore**". Successivamente il sistema impara a rimuovere questo disturbo visivo per riportare l'immagine al suo stato originale. È un po' come un artista che scolpisce una figura rimuovendo il materiale inutile da un blocco di marmo.

Quando si passa alla **fase di utilizzo**, nel momento in cui scriviamo una descrizione di ciò che vogliamo (**prompt**), il sistema genera un codice numerico specifico (**seed** o vettore latente) che rappresenta questa idea. Utilizza quindi questo codice, insieme a ciò che ha imparato durante la fase di apprendimento, per creare una nuova immagine, non esistente, partendo da una sorta di "rumore" visivo.



## I più interessanti tool Text to Image

**Dall-E** (*gioco di parole tra il nome Wall-E, il robottino della Pixar, e il pittore Dalì*) è una **creatura di OpenAI**. È stato il primo software di questo tipo a far parlare di sé, ma l'azienda per lungo tempo non ha rilasciato aggiornamenti. A breve arriverà la versione 3 che promette un bel salto nella comprensione del linguaggio dei prompt. Dall-E si può usare gratuitamente fino a consumare un certo numero di crediti che vengono assegnati alla registrazione. Poi, è possibile acquistarne degli altri.

Anche se non produce immagini fotorealistiche, può essere d'aiuto per quelle di altro tipo. Due funzioni interessanti sono l'**inpainting** e l'**outpainting**. La prima permette di circoscrivere un'area dell'immagine ottenuta e di sostituirla con un nuovo elemento. La seconda dà la possibilità di estendere i confini dell'immagine aggiungendo ulteriori riquadri. Lo strumento consente anche di caricare una foto ed ottenere delle variazioni.

**Midjourney** è il prodotto dell'omonimo laboratorio di ricerca fondato da David Holz. È comparso sul mercato a febbraio 2022 in versione alpha per poi essere aperto gradualmente a tutti. Al momento non ha ancora una sua interfaccia grafica e quindi per utilizzarlo bisogna collegare la propria utenza a quella di Discord. In pratica, attraverso l'interfaccia della chat è possibile dialogare con il **bot di Midjourney** che, sulla base dei nostri prompt azionati dal comando /imagine, ci restituirà l'immagine richiesta. A questo possono essere aggiunti dei parametri ulteriori per ottenere dimensioni specifiche o stili diversi.

**I suoi output sono molto più sorprendenti e realistici degli altri tool presenti sul mercato**, tanto che è diventato lo standard per il lavoro dei creativi.

Midjourney viene aggiornato costantemente e la sua evoluzione è impressionante come dimostra questa immagine che rappresenta il risultato di uno stesso prompt a distanza di poco più di un anno.



Midjourney v. 1 (febbraio 2022)



Midjourney v. 5.2 (giugno 2023)

In questa nostra disamina non si può non parlare di **Stable Diffusion** che non è un prodotto, ma è un modello di intelligenza artificiale (un diffusion model) realizzato da **Stability AI, CompVis LMU e Runway**. La prima release pubblica è datata agosto 2022. Il suo codice è stato reso disponibile con licenza open source per cui è diventato la base di tanti prodotti che ne sfruttano la sua potenza.

Gli utenti più esperti lo installano sul proprio computer e lo usano gratuitamente, sfruttando anche le personalizzazioni effettuate da altri sviluppatori (il miglior repository di modelli è CivitAI).

I meno esperti possono utilizzare specifici servizi web, tra cui **DreamStudio**, creato dalla stessa Stability AI. Il software ha una comoda interfaccia utente nella quale inserire i prompt (per descrivere l'output desiderato) ed eventuali prompt negativi (per indicare cosa non vogliamo). Al primo accesso si ottengono dei crediti per generare una serie di immagini gratuitamente, poi è necessario pagare per averne di più. La comodità di questo software risiede anche nel suo periodico aggiornamento con nuovi modelli generativi.

**Firefly** è lo strumento di text to image creato da **Adobe**, leader del mercato dei tool per la creatività digitale. Svelato a marzo 2023, ora è disponibile a tutti gratuitamente per un numero limitato di risultati (che però avranno una filigrana e non potranno essere usati commercialmente). La sua peculiarità è che non è addestrato su immagini raccolte sul web, ma sulla banca immagini di Adobe Stock, quindi un tipo specifico di foto che hanno una loro riconoscibilità e uno stile ben definito. E questo può essere sia un vantaggio che un limite.

Al momento Firefly ha poche funzioni: quella di generazione di immagini da testo, quella di inpainting per trasformare una porzione di immagine e una di ricolorazione delle immagini. La più originale si chiama **"Text Effects"** e permette di creare velocemente un testo arricchito da diversi effetti che si possono generare, semplicemente descrivendoli. Altri software pensati per le immagini, invece, fanno fatica a generare testi precisi seguendo le indicazioni di un prompt.





**A cosa porterà il rapido sviluppo degli strumenti di text to image? Ridurrà le barriere all'ingresso per la produzione di immagini da parte di marketer, copywriter e di tutti coloro che non hanno una formazione specifica? Oppure alzerà l'asticella e aprirà nuove prospettive per artisti e art director che, anziché combattere la tecnologia, dovrebbero sforzarsi di capire come inserirla nel workflow quotidiano per potenziare la propria immaginazione?**



Any questions? We're here to help.  
Please contact

**Vincenzo Cosenza &  
Wavemaker Innovation Team**

**Wavemaker**

**Grow fearless**