**Introduzione a Midjourney ed ai tool T2I – novembre 2024**

Fonti:

* Webinar di Vincenzo Cosenza - marketer e copywriter. (su AI360) aprile 2024
* Video di Raffaele Gaito - settembre 2024
* varie

Midjourney a settembre 2024 è il leader di mercato Text2Image, con moltissime funzionalità (in questa introduzione vedremo solo le principali), un mercato comunque che è in continua evoluzione.  
  
Midjourney inoltre è lo strumento T2I più chiaro nella tutela dell'utente finale. Tutti i suoi piani sono a pagamento, nella licenza d'uso si attesta esplicitamente che l'utente può utilizzare le immagini generate da midjourney a scopi commerciali.



In Italia la prima azienda a usare midjourney a scopi commerciali è stata “Il Mulino Bianco” per il marketing. Il loro team legale ha fatto una ricognizione dei contratti / clausole di vari tool ed ha scelto midjourney proprio perché afferma esplicitamente il diritto all’uso commerciale degli utenti. Sotto questi aspetti legali e di copyright, chatGPT/Dall-e 3 sono un po' meno chiari (?!).

**Due modalità di accesso:**

1. Da [midjourney.com](http://midjourney.com/), ora disponibile a tutti anche in creazione (sino a pochi giorni fa, per creare immagini dal sito web, occorreva prima averne create alcune da discord; altrimenti il sito era in sola consultazione delle immagini degli altri).

**E’ la versione 6.** Tutti possono vedere le immagini generate da tutti gli altri (è questa una caratteristica tipica di midjourney); è utile per imparare, facendo click sull'immagine si possono vedere il prompt che l'ha generata e chi. Diversi tab:



Il tab per visualizzare immagini random, hot (del momento), top day, taggate da me.

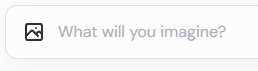


Il folder che contiene le immagini che abbiamo creato. Dunque serve per vedere ciò che abbiamo creato precedentemente (anche tramite Discord – vedi la seconda modalità di accesso) ed anche l’immagine il cui prompt abbiamo inserito nel box in alto.



Per organizzare le immagini create in folder e per usare filtri di ricerca.

Infine, il box in alto serve ad inserire il nuovo prompt o caricare un’immagine (click sull’icona, declick).



La seguente icona, a destra del box del prompt, permette di impostare alcuni parametri in modo grafico.



Come si vede, midjourney.com fornisce dunque una bella User Interface con comandi grafici (fino a pochi mesi fa non era disponibile).

1. Da [discord](https://discord.com/invite/midjourney) (un social per gruppi di discussione [[1]](#footnote-1)): è la modalità di accesso tradizionale, meno user-friendly ma più ricca di funzionalità (upscale e vary). Occorre registrarsi prima su discord e poi su midjourney.  
   Ci sono varie "stanze" midjourney: benvenuto, annunci/novità, **newbies per la generazione immagini**. Gli utenti di ogni stanza newbies sono visibili in alto a dx.  
   Per avere una stanza privata, non affogata nel mare magnum collettivo e quindi senza distrazioni rinfrescate in tempo reale, **nella lista utenti di destra si individua il midjourney bot (uno degli utenti, di fatto), si preme tasto dx del mouse sopra di esso e si sceglie la voce “messaggio”.**

**L’interazione (tramite discord)**  
L'interazione da prompt non è semplice come con i chatbot. midjourney infatti ha sì 20 milioni di utenti (pochi comunque rispetto ai principali chatbot), ma sono tutti tecnici specializzati in grafica, illustrazioni e fotografia od appassionati.  
Dal prompt si interagisce con midjourney tramite lo **slash**. Per generare immagini si fa:  
**/imagine** prompt (con la autocompletion, prompt e cursore lampeggiante), invio: parte l’esecuzione (completata entro 60 secondi – vedi claim di midjourney)  
  
Es. /imagine prompt women portrait. Si attiva il midjourney bot. La modalità utilizzata di default può essere di vari tipi. Per cambiarla /settings (ce ne sono molte) --> fast mode. Dipende dal piano sottoscritto.  
Si ottengono 4 immagini.

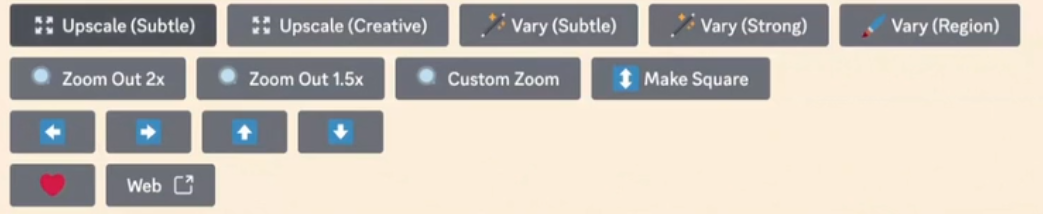
La generazione è progressiva. I modelli usati dagli strumenti T2I come midjourney (anche Dall-e 3 e Stable Diffusion) sono differenti da quelli usati dai chatbot T2T. Usano una tecnica detta "diffusione" che consiste nel prendere un'immagine, aggiungere rumore (pixel a caso), e poi togliere rumore a caso a seconda del prompt che gli abbiamo dato. Cioè in pratica distruggendo e poi ricreando l'immagine. In parte questo processo è visibile. E’ un po' come uno scultore che dal blocco di marmo (la materia informe) ricava poco per volta la scultura. I modelli T2I sono preallenati su milioni di immagini. (midjourney è stato chiamato in causa da 5 aziende per presunta violazione di copyright).  
  
Con i tool T2I è meglio scrivere i **prompt in inglese**, anche se funzionano anche con altre lingue.

Se si clicca sull'icona circolare si riesegue la generazione.  I risultati sono sempre diversi perché, ad ogni generazione, midjourney aggiunge un seme (seed) al prompt, che varia da 1 a 17\*10^9. Dunque le immagini potenzialmente generabili da midjourney a fronte di un prompt specifico (il medesimo!) sono **17 miliardi!**  
  
Stile dell’immagine, personaggio, colori (o bianco e nero) sono casuali.  
Da fine marzo 2024 è possibile forzare la medesima persona con "character reference" (ne parliamo più avanti).  
  
Possiamo ovviamente **arricchire** il precedente prompt, molto scarno. Ad es.  
*Italian woman portrait, short black-haired, fotography*  
dove abbiamo anche indicato lo stile dell'immagine (fotorealistico anziché pittorico).  
  
Sino alla v5, midjourney funzionava meglio con **prompt semplici e schematici, cioè keyword o sequenze di keyword**. I prompt generati dagli LLM come possibili input ai tool T2I (un classico aiuto alla nostra mancanza di fantasia od esperienza nel campo) rischiano di essere troppo discorsivi. Ad es. "2 personaggi, in un giardino, la luna sullo sfondo" è un prompt che rischia di confondere midjourney. **I tool T2I infatti non hanno le capacità di comprensione del testo degli LLM.** Tuttavia, dalla v6 questo limite sembra parzialmente superato, anche se comunque è consigliabile sfrondare un po' il prompt generato dal chatbot.

Come **personalizzare** ulteriormente la generazione? In vari modi:  
- arricchendo il prompt (copiato ed incollato da prima), ad esempio con il tipo di fotografia: *studio* fotography, oppure *street* potography, oppure *documentary* fotopraphy.

Proviamo studio potography  --> la nitidezza, le luci e lo sfondo (neutro) sono quelli tipici delle foto in studio.  
- specificando nel prompt dei parametri, preceduti dal doppio trattino e poi uno spazio ad es. -- AR 16:9.  -- style raw  
raw genera immagini più scarne e realistiche, spesso migliori. I modelli di diffusione hanno infatti la tendenza a creare immagini troppo ricche e sofisticate.  
Un parametro interessante è la macchina fotografica. È bravo a riprodurre la marca della macchina, ad es Canon o Laika, molto meno a riprodurre il singolo modello di una data marca.  
  
  
**Tasti upscale (ingrandimento)**

  
4 in ordine di lettura.  
Upscale aumenta la grandezza e definizione dell'immagine selezionata. Il default di midjourney è 1024 x 1024 (con ar 1:1).  
Compaiono nuovi tasti.

  
Upscale di due tipi (Subtle o Creative), si raggiunge una definizione 4X, adatta anche alle **stampe**.

Upscale Subtle genera un’immagine (ingrandita) fedele all’originale, Upscale Creative aggiunge qualche dettaglio. Dipende dai nostri obiettivi.  
Vary di tre tipi. Con i primi due si può variare l’immagine poco o molto. La terza opzione Vary (Region), detta "impainting" (aggiunta ad inizio aprile 2024 anche da dalle3), permette di selezionare una regione, ad esempio i capelli, e inserire nel prompt una caratteristica in più della immagine, ad es. *hat*. Al click può comparire una pop-up per una eventuale modifica del prompt.

Questa funzionalità ci mette un poco di tempo e non sempre funziona bene.

I due bottoni di zoom out allargano l’immagine (e quindi creano nuove parti di essa). Le frecce inseriscono spazio a sx, dx, alto, basso.

: va sul sito con l’immagine selezionata.  
  
**Tasti vary (variazioni)**

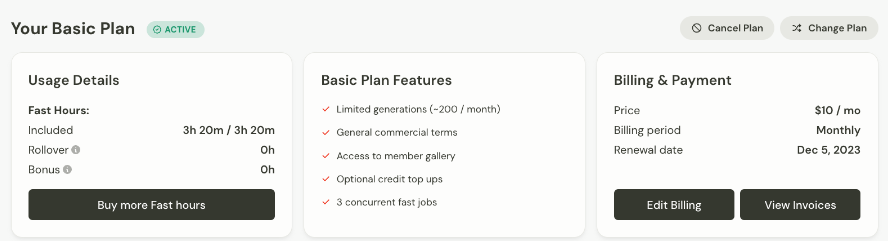
  
4 in ordine di lettura.  
Ripetono le funzioni di variazione viste prima, che cioè possono essere attivate indipendentemente dall’upscale oppure dopo l'upscale, come visto prima.  
Variazione subtle o strong. Utili per arrivare all'idea di immagine che avevamo in mente. Cioè non bisogna arrendersi.

In generale, i ritocchi alle immagini alla Photoshop (i cosiddetti “foto-ritocchi”) non sono possibili con midjourney. Serve Photoshop!  
  
**Testi nell’immagine generata.**  
L'aggiunta di testo alle immagini è da sempre il punto debole dei modelli T2I, che non sono nati per questo e sono stati allenati su immagini. La diffusione non è adatta al testo. Con la v6 questo skill è comunque migliorato. Provare ad esempio:  
a girl wearing a t-shirt with the text "hello world". Insistere, non arrendersi, fare variazioni.  
Tra i vari tool T2I il migliore per la generazione di immagini con contenuti testuali è *ideogram*, gratuito, che riesce quasi sempre a generarle, purché il testo sia breve (2 parole). Anche in italiano! [Ideogram è una startup fatta da alcuni ex-dipendenti di Google]

Dopo avere creato / modificato queste immagini, è utile andare sul sito midjourney: sotto il tab “create” vediamo l’elenco, in ordine cronologico inverso, di tutti i risultati del nostro lavoro eseguito prima tramite discord.

**Piani di abbonamento**  
Sono 4, come visto all’inizio.

Il piano Basic: 200 prompt generati (al mese), che corrispondono circa a 3.5 ore. Non ha la modalità rilassata (molti prompt eseguiti a bassa priorità, utile a chi ci lavora molto). Ecco il prospetto del mio piano basic:



NB I limiti di utilizzo del piano Basic sono espressi in prompt. Cioè le immagini generate sono da moltiplicare per 4 (800).

Il piano Standard: ha 15 ore, dunque circa 1000 immagini al mese. Sono tante. Questo piano è dunque adatto a molti utenti. Inoltre le ulteriori immagini sono comunque generate, anche se in modalità rilassata, cioè con tempi in base al traffico del momento. Ma al mattino la modalità relaxed va bene (non ci sono ancora gli utenti nordamericani).  
  
Gli ultimi due piani (Pro e Mega) offrono più immagini veloci al mese e soprattutto garantiscono la privatezza (non disclosure) delle immagini generate, detta "stealth image generation". Ad esempio utile per nuove idee, lancio di prodotti o servizi, bilanci e conti economici. Con i piani Basic e Standard le immagini generate possono, in teoria, invece finire nella community del sito midjourney vista prima (le immagini sotto il primo tab “Explore”)  
  
**Image2Image**  
Dal bot midjourney su discord:  
- premere + e poi "carica file"  
- selezionare l'immagine sul nostro PC.  
 essa viene caricata sul server midjourney.  
- specificare il prompt, nel solito modo, ad esempio: "man portrait in the style of impressionist paintings" e poi aggiungere la url dell'immagine (dove la prendo?), sempre nel prompt  
E’ più facile farlo dalla versione del sito. Carica l’immagine e scrivi il prompt. [Cioè un prompt testuale serve sempre, anche per Image2Image!] Come si vede, i risultati non sono buoni. Cioè, midjourney non va bene per generare varianti di una nostra foto.

Funzionalità interessante disponibile sia su discord che sul sito: la **character reference**. Vedi [questo bel video](https://www.youtube.com/watch?v=9HnayWhfJJA). Tramite discord:

* copiare l’URL dell’immagine prima generata, in questo modo: click sull’immagine, tasto dx del mouse, copia “image address”. Supponiamo sia una delle versioni del ritratto di donna italiana prima generato.
* /imagine prompt woman in a white dress –cref imageurl
* il parametro “cref” sta proprio per character reference
* midjourney genererà le immagini in accordo, cioè dello stesso personaggio (a seconda del prompt fornito).
* non sempre funziona bene.

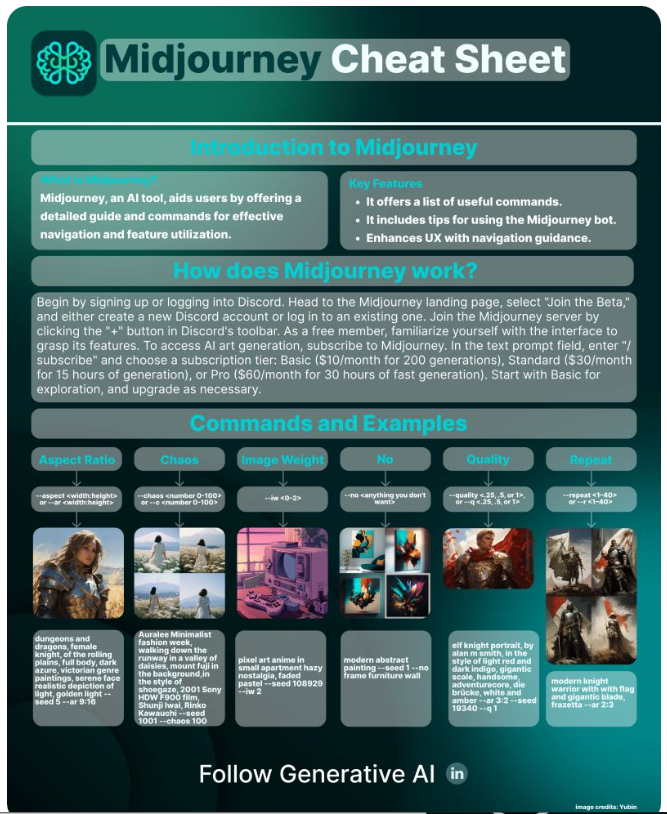
La "character reference" è una funzione della primavera 2024, molto interessante.

[Qui](https://www.youtube.com/watch?v=zgeI_ffCudY) un video sulla **consistenza dei personaggi** (*character consistency*).

Un'altra funzionalità di febbraio 2024 è la "**style reference**". Le immagini generate dal prompt rispetteranno lo stile dell'immagine selezionata. Ad esempio, da discord:

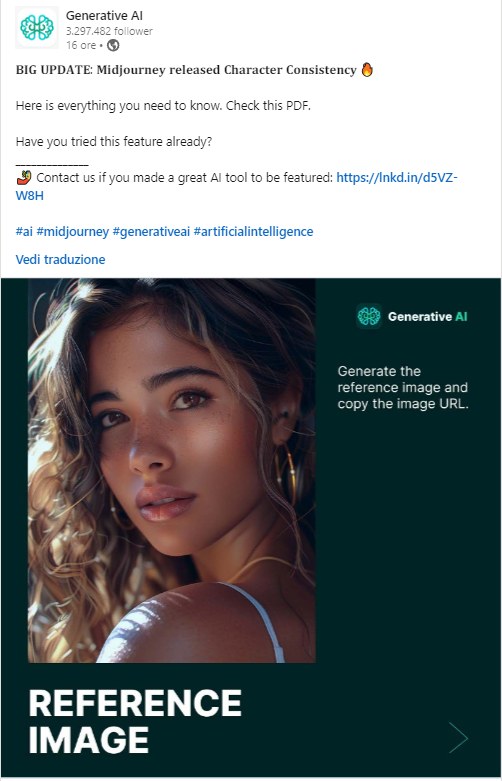
* /imagine prompt rock mountains in gray and brown
* copiare l’URL dell’immagine prima generata, in questo modo: click sull’immagine, tasto dx del mouse, copia “image address”.
* /imagine prompt city landscape –sref imageurl
* il parametro “sref” sta proprio per style reference
* genera panorami urbani con tonalità grigio e rosa.
* Funzionalità molto utile per creare immagini consistenti, oppure con un tocco personale, in base ad un’immagine “seme” costruita ad esempio con una certa palette di colori ed un certo stile.

Midjourney ha anche una API.



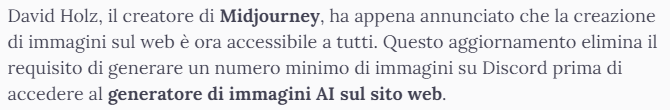
Febbraio 2024: character consistency

[Vedi il PDF](midjourney%20-%20character%20consistency.pdf).



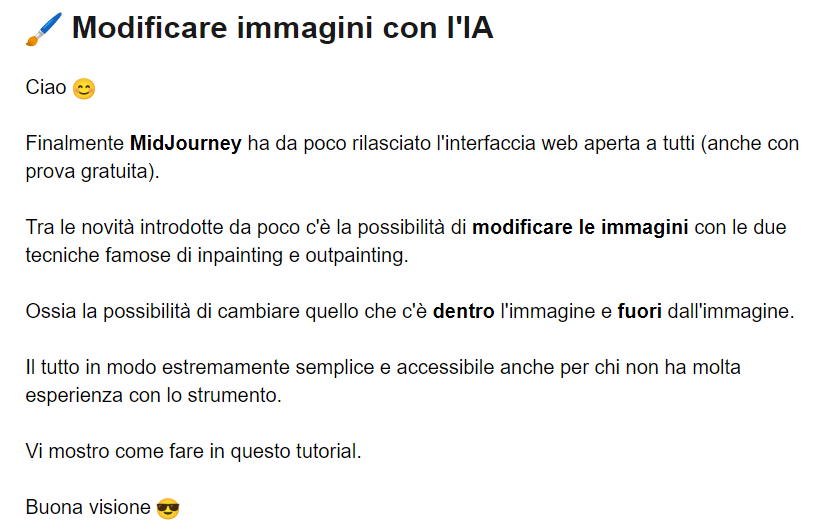
Agosto 2024



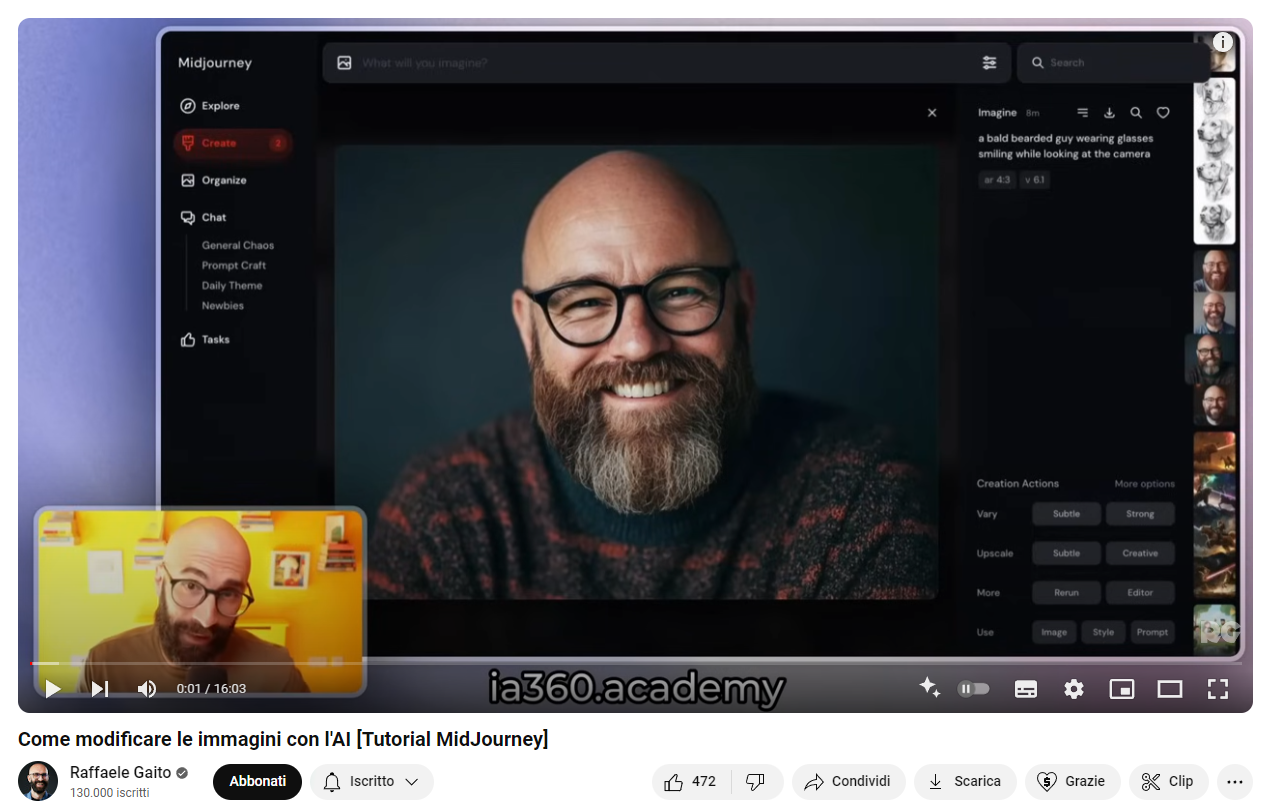


Vedi [questo link](https://www.punto-informatico.it/midjourney-sito-web-aperto-a-tutti-25-generazioni-ai-gratis/).

Agosto 2024



[](https://www.youtube.com/watch?v=oyuDKkGYmY0)



**Nota di Gaito sui prompt negativi della live del 3 settembre 2024.**

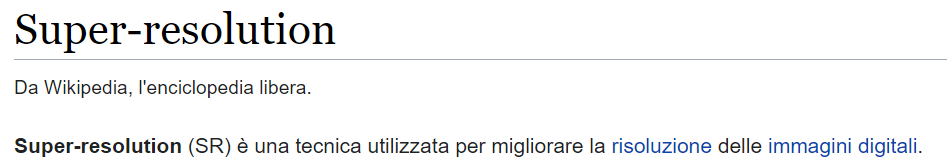
I modelli T2I non sono sensibili agli aspetti negativi dei prompt – fanno il contrario

Ad esempio: “Disegna una stanza di un ufficio di progettisti informatici, non mettere orologi sul muro”

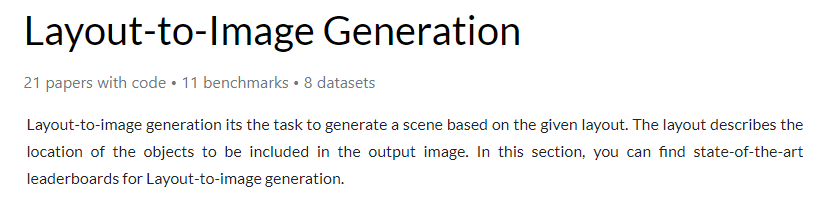
Le negazioni devono essere messe in apposite sezioni o dopo appositi flag.

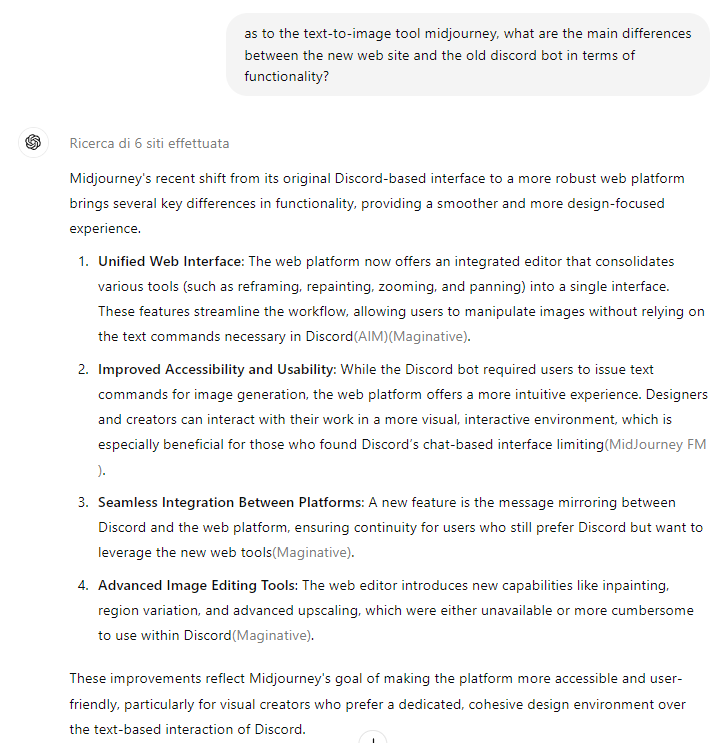
Ad esempio in midjourney: no blue, yellow (<https://youtu.be/iyV615lcrM8>)

**3 concetti:**

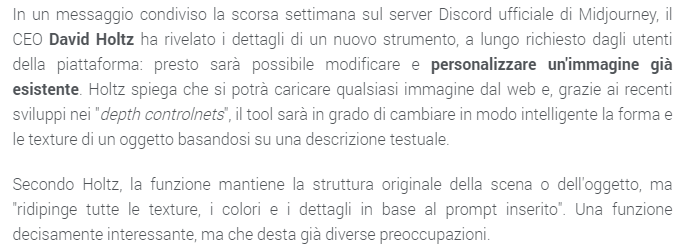


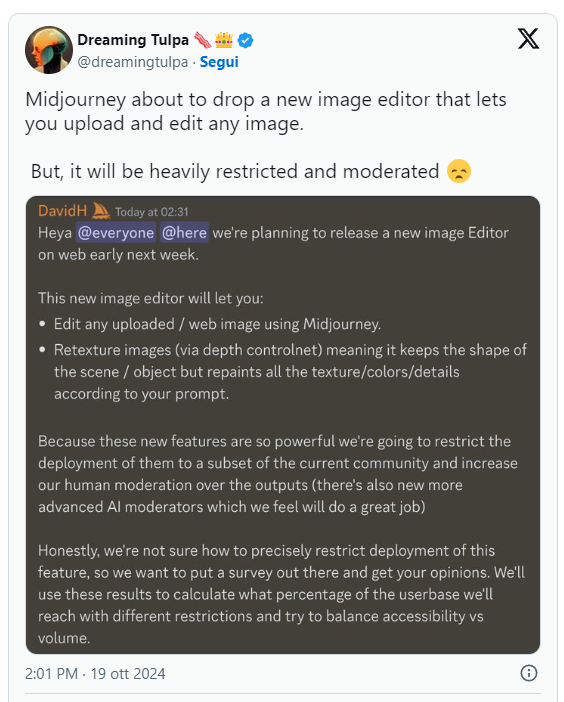
**Impainting**: to render something entirely new in any part of an existing image. In midjourney è implementata con “Vary (Region)”.











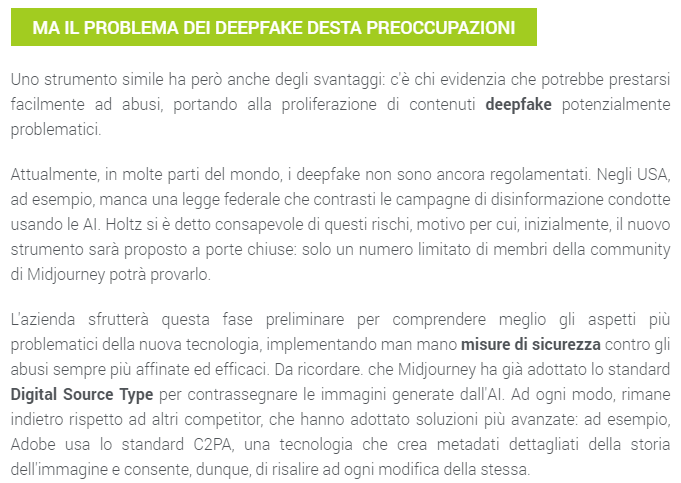


Immagine che contiene testo, schermata, design

Il contenuto generato dall'IA potrebbe non essere corretto.

Immagine che contiene testo, schermata, Carattere

Il contenuto generato dall'IA potrebbe non essere corretto.

1. L’installazione di Discord su Windows 10 purtroppo lo imposta con avvio automatico. Se si va nella sezione “App di avvio”” delle “Impostazioni di sistema” ciò non risulta! Occorre operare da Discord stesso, in questo modo: Discord --> Impostazioni utente (la rotella in basso) --> Impostazioni di Windows 🡪 Togliere la spunta alla prima opzione: "Apri discord" [↑](#footnote-ref-1)