

Procedura TTP/PTI per IA generativa e test di elaborazione inversa

Daniele Giovanni Papi

Abstract

La creazione di immagini tramite IA è viziata dai livelli imprevedibili di autonomia degli algoritmi. La traduzione del linguaggio naturale in sequenze di selezione presenta due punti di criticità: competitività semantica (significati plurimi delle parole) e ambiguità proiettive (vincoli geometrici, condizioni ottiche). Entrambi sono in stretta dipendenza dallo scorrimento reciproco delle reti generative avversarie (GAN) e dalle priorità di esclusione/inclusione alle quali sono state addestrate; il problema non è nell'intenzione (cosa voglio rappresentare) ma nella sua espressione (come voglio farlo). La struttura, l'ordine logico, la precisione sono concetti dal senso diverso se ci si muove all'interno della completezza del linguaggio naturale oppure in quella forma nascosta di bilancio statistico che presiede agli alberi di selezione di una IA generativa. Il conflitto tra due configurazioni espressive del tutto diverse nella strategia di traduzione, nella quale i termini che normalmente sono due (A → B) diventano un numero imprecisato che dipende da quanti e quali elementi di confronto le GAN trovano nella competizione reciproca verso la scelta del descrittore. In quest'esperienza si applica un motore di IA esterno come sistema di controllo degli elementi di confronto delle GAN della IA generativa per verificare l'ipotesi che con l'uso di un testo letterario quale risorsa descrittiva, l'utilizzo in sequenza di due motori IA sia vantaggioso in termini di qualità dei risultati nelle procedure text-to-image.

Parole chiave

IA generativa, BOT, text-to-image, prompting, feedback.



Lo Studio dei ritratti,
elaborazione dell'autore

Scopo e strumenti

Quest'esperienza si basa sul consolidato utilizzo della Intelligenza Artificiale (IA) per rappresentare graficamente un luogo descritto in un testo letterario. Operazioni analoghe basate sull'intelligenza naturale, sulla creatività, sul talento (tutti termini esclusivamente umani) sono state condotte per secoli (pittura di soggetto biblico e mitologico, illustrazioni editoriali, tavole dei cantastorie etc.) e appartengono a una logica ormai chiara di traduzione visibile della forma narrativa; questa radicazione diffusa e precisa nell'alveo delle aspettative sarebbe di per sé termine di confronto sufficiente per la valutazione dei risultati; tuttavia, un sistema di valutazione artificiale a *feedback* è stato ottenuto attraverso l'elaborazione inversa del prodotto grafico.

In considerazione della distanza linguistica tra gli strumenti semantici della figurazione da ottenere e quelli della forma letteraria, l'esperienza è volta alla definizione di un possibile ed efficace percorso logico e vuole verificare l'ipotesi che l'analisi del testo e la successiva sintesi in un *prompt* per IA generativa possa essere affidata con profitto a un BOT; a questo scopo la procedura è stata separata in due fasi, definite *text-to-prompt* (TTP) e *prompt-to-image* (PTI). La fase di elaborazione TTP è preliminare ed è stata condotta attraverso il BOT *ChatGPT 4*, in considerazione della posizione di assoluto predominio e la diffusione capillare dell'algoritmo del quale sono state vagliate le opzioni: *Plus*; *Consensus*; *WFM* ed è stata scelta la *Plus*, per la maggiore flessibilità e la tendenza a non ricorrere a descrittori esterni al testo in analisi.

L'elaborazione TTP interviene sul testo allo scopo di estrarre e ordinare gli elementi descrittivi secondo le priorità e le capacità di confronto delle GAN che interverranno nella fase PTI. Nella preparazione della fase PTI sono stati considerati i tre principali algoritmi di generazione: *DALL-E OpenAI*, *Stable Diffusion* e *Midjourney* e si sono osservati quattro parametri: efficacia percettiva della geometria della scena; raffinazione del dettaglio e delle texture; comandi aggiuntivi per rigenerare variazioni e migliorare singoli particolari; flessibilità dell'interfaccia. *Midjourney* è stato scelto sulla base di valutazioni empiriche, se non euristiche, perché in letteratura (ufficiale e parallela) non è documentata nessuna informazione sui dettagli funzionali degli algoritmi e le valutazioni di prestazione sono state condotte in termini strettamente applicativi. *Midjourney* è stato utilizzato attraverso l'interfaccia *ImageArt* [Nichol 2021, pp. 6546-6471; Radford 2021, pp. 8748-8763; Saharia 2022, pp. 9110-9126].

Caso studio e procedura

Il Ritratto di Dorian Gray contiene elementi intellettuali di particolare assonanza con il tema, ma questa è una valutazione essenzialmente umana, di natura culturale, e si è ritenuto di utilizzare il testo di Wilde per quattro motivi concreti e indipendenti: universale notorietà dell'opera; chiarezza del contesto storico e urbano di riferimento; descrizioni minuziose degli ambienti; ampia possibilità di confronto con i prodotti dell'intelligenza naturale (figg. 1, 2), così come suggerito dalle premesse sperimentali.

Tra i molti passi possibili e adatti allo scopo, in ragione della ricchezza e della completezza della descrizione dello studio di Basil Hallward, è stato scelto l'*incipit* della traduzione italiana: "Lo studio era intriso d'uno splendido odore di rose e quando la lieve brezza estiva strusciava



Fig. 1. Albert Lewin, fotogramma da *The Portrait of Dorian Gray*, pellicola, 1945.

Fig. 2. Paul Thiriat, illustrazione da *The Pictures of Dorian Gray*, raccolta di incisioni, 1907.



tra gli alberi del giardino, dalla porta aperta penetrava il profumo pesante delle serenelle, o quello più delicato dei rosaspini [...] nell'angolo di un divano coperto da stoffe persiane [...] di quando in quando l'ombra fantastica di un uccello volante si proiettava e scorreva sulle pesanti tende di seta [...] lo strepito di Londra pareva la vibrazione delle note basse di un organo lontano. In mezzo alla camera, su un cavalletto, era il ritratto a figura intera di un giovane di singolare bellezza" [Wilde 2013, p. 9].

È stato notato che esplicitare qualsiasi informazione sull'origine del passo non è rilevante per il BOT; l'albero di selezione di *ChatGPT 4* sembra essere addestrato sui contenuti delle opere letterarie, quantomeno di questa, suggerendo così una possibile evoluzione dell'esperienza con l'uso di testi scritti allo scopo.

Per valutare le reali capacità di identificazione dell'origine dell'incipit sono state preliminarmente compiute 13 procedure di controllo indipendenti di TTP, delle quali tre a partire dalla traduzione in francese, una dal tedesco, una dallo spagnolo, otto in italiano e si è verificato che il riconoscimento del testo avviene sempre, e sempre conduce il BOT a procedere sulla versione in lingua originale (inglese britannico di periodo vittoriano) della quale conserva gli elementi caratteristici nel risultato della procedura sintetica (*prompt*), ma non in quello della fase analitica (schema dei contenuti) che è offerta nella lingua di selezione dell'interfaccia, nel caso in oggetto, in italiano (figg. 3, 4).

Fig. 3. Sintesi della scena tramite BOT.

Fig. 4. Sintesi degli elementi visivi del testo tramite BOT.

Scena:

- Una stanza luminosa e piena di un profumo floreale.
- Fiori di rose fresche e serenelle, con un tocco di profumo di rosaspini.
- Una leggera brezza estiva attraversa gli alberi visibili da una porta aperta.

Dettagli focali:

- Lord Henry Walton sdraiato su un divano ricoperto da tessuti ricchi.
 - Sta fumando sigarette.
 - Può vedere i rami di un'acacia fuori dalla porta.
 - I rami sono fragili, pieni di fiori colorati e dolci come il miele.

Atmosfera:

- Ombre fugaci di uccelli proiettate su tende pesanti di seta, evocando un effetto giapponese.
- L'ombra richiama pittori di Tokyo, caratterizzati da un'estetica di velocità e movimento.

Sonorità e ambientazione:

- Ronzio di api tra l'erba alta e attorno ai caprifogli.
- Lontano ronzio urbano che sembra il suono di un organo basso.

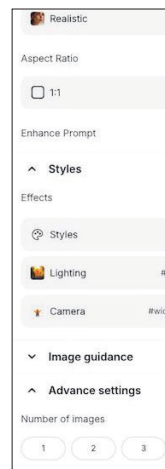
Oggetto centrale:

- Un ritratto a figura intera di un giovane di straordinaria bellezza, posto su un cavalletto.
- Basil Hallward, l'artista, è seduto poco distante.

Sintesi degli elementi visivi principali dello studio di Basil Hallward:

- Illuminazione e atmosfera:**
 - Lo studio è inondato di luce naturale che entra dalla porta aperta e dalle finestre.
 - Un senso di calma e raffinatezza permea l'ambiente, arricchito da profumi floreali.
- Dettagli floreali e naturali:**
 - Il profumo delle rose si mescola a quello delle serenelle e dei rosaspini.
 - Dall'esterno, si intravedono alberi e rami di acacia con fiori dolci e fragili.
- Oggetti nello studio:**
 - Un cavalletto centrale regge un ritratto a figura intera di un giovane di straordinaria bellezza.
 - Una sedia o un divano ricoperto di tessuti ricchi e decorativi.
 - Tende di seta pesante che si muovono leggermente con la brezza estiva.
- Elementi dinamici:**
 - Ombre fugaci di uccelli si proiettano sulle tende, evocando un effetto estetico giapponese.
 - Il ronzio delle api e il suono lontano della città creano una sinfonia naturale e urbana.

Fig. 5. Maschera di impostazione di Midjourney.



L'esperienza finale di TTP è stata condotta a partire dalla traduzione italiana. Le maggiori differenze semantiche tra i risultati analitici e quelli sintetici delle diverse procedure si sono rilevate a partire dall'espressione della richiesta. Le formulazioni che hanno prodotto analisi e sintesi migliori sono state le seguenti quattro:



Fig. 6. Generazione con "high contrast".

1. a partire da questo testo crea un *prompt* per la generazione dell'immagine corrispondente con *Midjourney* ed evidenzia testualmente gli elementi fondamentali;
2. trasforma il seguente testo in un *prompt* ottimizzato per l'elaborazione grafica con *Midjourney* e genera uno schema analitico testuale che descriva gli elementi principali;
3. crea un *prompt* adatto a *Midjourney* per generare l'immagine descritta nel seguente testo e sintetizza in uno schema analitico gli elementi di massimo interesse;
4. genera un *prompt* adatto a *Midjourney* per l'elaborazione grafica della scena descritta in questo testo e crea un riassunto analitico testuale degli elementi principali.

A una lettura 'umana' le richieste sono equivalenti ma le scelte linguistiche e semantiche influenzano imprevedibilmente il prodotto della IA perché i rapporti di co-significanza del linguaggio naturale non sono rispettati e termini linguisticamente pressoché sinonimi (crea/genera; proponi/presenta; testo/brano/frase/periodo; isolare/definire; elemento/oggetto/parte/componente etc) sono interpretati in modo diverso e producono *prompt* estremamente distanti l'uno dall'altro, secondo logiche apparentemente casuali. Questi *lost-in-translation* sembrano dipendenti da nodi problematici non risolti (e forse non risolvibili per via statistica) all'interno degli alberi di selezione del BOT perché, al di là delle sfumature di traduzione presenti anche nel linguaggio naturale, non è stata dimostrata nessuna differenza nel proporre la richiesta in lingua inglese (lingua originale di training del BOT di *ChatGPT 4*, come largamente ritenuto, seppure non documentato) o in italiano e si è quindi proceduto a una verifica applicativa pratica preliminare dei diversi *prompt* ottenuti dall'elaborazione TTP.



Fig. 7. Generazione con 'natural sunlight'.



Fig. 8. Generazione con 'aerial view'.

I *prompt* generati dalle richieste selezionate come i quattro migliori sono stati utilizzati per alimentare Midjourney in multiple variazioni di configurazione del motore generativo con lo scopo di giungere alla stabilizzazione dei parametri generativi e di individuare tra i quattro il *prompt* in grado di produrre le immagini con migliore aderenza alla richiesta:

'A luminous studio filled with the scent of roses, lilacs, and hawthorn blossoms. A summer breeze flows gently through an open door, revealing fragile acacia branches with sweet, honey-like flowers. A silk-covered couch. Shadows of birds flit across heavy silk curtains, creating a Japanese-inspired effect. Bees hum over uncut grass and golden honeysuckle. In the center of the room, a full-length portrait of a strikingly beautiful young man stands on an easel'.

È di particolare interesse notare come non sia esplicitato alcun rimando al testo, all'autore, all'atmosfera londinese vittoriana, allo stile dell'arredo; è inevitabile concludere che l'identificazione esplicitata da ChatGPT 4 nella fase analitica sia in qualche modo possibile anche per le GAN di Midjourney e che il BOT 'riconosca' questa capacità.

Il *prompt* (generato dalla richiesta al punto 3 del precedente elenco) è stato elaborato da Midjourney con il controllo dei seguenti parametri: (a) rapporto di quadro dell'immagine; (b) risoluzione; (c) angolo e assetto di osservazione; (d) caratteristica grafica; (e) tono e intensità dell'illuminazione (fig. 5).

120 generazioni sono state effettuate con i seguenti parametri:

(a) 1:1; (b) HR 4096×4096; (c) *horizontal wide shot*; (d) *no-style*; (e) *high contrast* (fig. 6).

400 generazioni sono state effettuate con i seguenti parametri:

(a) 1:1; (b) HR 4096×4096; (c) *horizontal wide shot*; (d) *realistic*; (e) *natural sunlight* (fig. 7).

352 generazioni sono state effettuate con i seguenti parametri:

(a) 1:1; (b) HR 4096×4096; (c) *aerial view*; (d) *realistic*; (e) *natural sunlight* (fig. 8).

Complessivamente, le 872 procedure indipendenti, ciascuna per due o quattro immagini, hanno prodotto un totale di 2268 immagini.

Risultati

Sulle 2268 immagini si è operata una selezione fino a definire un singolo prodotto grafico da ritenere il risultato finale della procedura generativa (fig. 9).

L'immagine finale contiene tutti gli elementi descritti dall'*incipit*, includendo in essi l'atmosfera e lo stile letterari, il contesto architettonico e in larghissima misura anche quello storico.

Quest'ultimo presenta delle imprecisioni di ordine marginale dipendenti da scelte non ottimali nell'arredamento e nella decorazione ma, sulla scorta delle molte rappresentazioni umane dello stesso soggetto osservate, si ritiene appartengano a quella 'deviazione dalla maniera' che è propria di ciascun ambiente reale e non siano di disturbo, piuttosto citazioni tendenti verso un maggiore realismo (fig. 10). Tuttavia, *Midjourney* non compie valutazioni interne neppure indicative (ad esempio, a partire dalla regolazione dei parametri) e non è in grado di autostimare l'aderenza del prodotto grafico al *prompt* fornito; per questo, dopo l'operazione di selezione necessariamente condotta in modalità 'umana', si è applicato all'immagine selezionata e a un set di altre 9 di controllo (fig. 11) il test di *feedback* attraverso IA.



Fig 9. Immagine finale.



Fig 10. Varietà
e compresenza nell'arredo
vittoriano.



Fig 11. Set delle 9 immagini
di controllo.

È stata applicata la funzione *prompt enhance* di Midjourney (l'algoritmo fornisce un nuovo *prompt* interpretando l'immagine generata come se fosse di provenienza esterna e, come ovvia conseguenza, il *prompt* originale è dimenticato e completamente ignorato) e gli 1 + 9 *prompt* autogenerati da Midjourney sono stati offerti all'interpretazione del BOT di ChatGPT 4, con la logica di procedere in direzione inversa, da ITP (*image-to-prompt*) a PTT (*prompt-to-text*). Il *prompt* ottenuto dall'immagine finale è stato:

'A wide establishing shot of a grand Victorian-era painter's atelier in London, showcasing the expansive hall with elegant architectural details. The scene is framed to capture the grandeur of the space, featuring a central easel displaying a captivating portrait of a young man, surrounded by three additional easels showcasing other portraits. The walls are adorned with numerous paintings, reflecting the artistic legacy of the room. Natural sunlight floods through a large window, casting warm and soft tones that create a golden hour glow. Shadows are cast naturally across the hall, enhancing the ambient outdoor lighting and highlighting the silk-covered Victorian-style couch that adds a touch of elegance to the scene. Gentle highlights dance on surfaces, and subtle reflections of greenery add depth to the earthy color palette. An open door invites a refreshing summer breeze, rustling through delicate

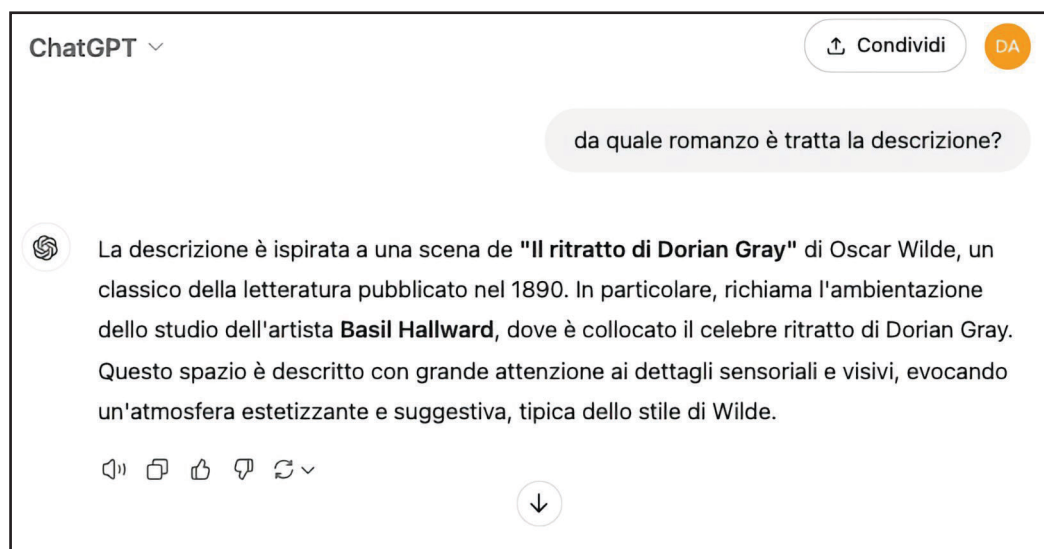


Fig. 12. Esito del test PTT sull'immagine finale.

acacia branches adorned with sweet, honey-like flowers. The shadows of birds flitting across heavy silk curtains create a serene effect, reminiscent of Japanese aesthetics. In the foreground, bees hum gently over uncut grass and golden honeysuckle, enriching the tranquil atmosphere. The balanced composition and deep focus clarity draw the viewer into this picturesque moment, encapsulating the essence of a bygone era with sweeping clarity and cinematic perspective.'

I *prompt* ottenuti dal set di controllo sono diversi ma contengono elementi di somiglianza. Sottoposti all'analisi *prompt-to-text* del BOT di ChatGPT 4, i *prompt* generati dalla procedura *image-to-prompt* hanno dato luogo, sia per l'immagine finale sia per le 9 del set di controllo, a un univoco riconoscimento da parte del sistema (fig. 12), sebbene al proprio interno essi non contengano che pochissime delle parole originali del testo di Wilde.

In sintesi, la valutazione attraverso il *feedback test* ha mostrato che i *prompt* autogenerati da Midjourney contengono elementi propri dell'elaborazione IA e ad essa sola manifesti che ricollegano invariabilmente l'immagine alla fonte letteraria esatta.

Conclusioni e sviluppi

Come è noto, gli algoritmi di IA procedono in modo fortemente condizionato dal training globale del sistema; le ragioni che hanno portato alla scelta dell'incipit de *Il Ritratto di Dorian Gray* sono probabilmente le stesse che hanno reso possibile il successo del test di *feedback*.

Inoltre, riteniamo che la prima procedura di generazione dei *prompt* abbia provveduto a una sorta di *training* marginale che è stato assimilato dal BOT e che questo debba essere verificato con la ripetizione dell'esperienza con un testo ignoto al training globale e su sistemi appartenenti a diverse catene di esperienza (BOT riferibili a memorie locali indipendenti, ossia a utenti diversi). Tuttavia, riteniamo altresì che le IA siano in grado di dare luogo a forme grafiche e analitiche irricognoscibili a livello naturale ma evidenti all'interpretazione artificiale stessa.

Riferimenti bibliografici

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M. (2021). Towards photorealistic image generation and editing with text-guided diffusion models. In *ICML*, 2022 pp. 6456-6471. arXiv preprint arXiv:2112.10741.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *ICML*, 2021, pp. 8748-8763. <https://doi.org/10.48550/arXiv.2103.00020>.

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, M. Norouzi, D. (2022), Image-to-image diffusion models. In *ACM SIG-GRAPH 2022 Conference Proceedings*, 2022, pp. 9110-9126.

Wilde, O. (2013). *Il ritratto di Dorian Gray*. Milano: Feltrinelli.

Autore

Daniele Giovanni Papi, Politecnico di Milano, daniele.papi@polimi.it

Per citare questo capitolo: Daniele Giovanni Papi (2025). Procedura TTP/PTI per IA generativa e test di elaborazione inversa. In L. Carlevaris et al. (a cura di). *èkphrasis. Descrizioni nello spazio della rappresentazione/èkphrasis. Descriptions in the space of representation*. Atti del 46° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Milano: FrancoAngeli, pp. 3121-3140. DOI: 10.3280/oa-1430-c917.

Procedure for TTP/PTI in Generative AI and Reverse Processing Tests

Daniele Giovanni Papi

Abstract

The creation of images through AI is influenced by the unpredictable levels of autonomy exhibited by algorithms. The translation of natural language into selection sequences presents two critical challenges: semantic competitiveness (the multiple meanings of words) and projective ambiguity (geometric constraints and optical conditions). Both are closely tied to the interplay of generative adversarial networks (GANs) and the inclusion/exclusion priorities they have been trained on. Thus, the core issue lies not in the intent (what I want to represent) but in its expression (how I want to represent it). Detailed structuring, logical order, and precision hold different meanings depending on whether one operates within the comprehensiveness of natural language or within the hidden form of statistical balance governing the selection trees of a generative AI. The conflict between these two entirely different expressive configurations manifests as a translation strategy in which what is typically a binary relationship ($A \rightarrow B$) becomes an indeterminate number, depending on how many and which comparison elements the GANs identify in their mutual competition to select descriptors. These quantities and qualities are only marginally controllable via the refinement and depth parameters available to the user. In this study, an independent AI engine, external to the representation system, is applied as a prompting tool and subsequently as a control mechanism, using an inverse procedure, for the GAN comparison elements of the generative AI. The objective is to test the hypothesis that starting from a literary text, the sequential use of two AI engines improves the quality of results.

Keywords

Generative AI, BOT, text-to-image, prompting, feedback.



The Portrait Studio,
author's elaboration.

Purpose and Tools

This experiment is based on the well-established use of AI (Artificial Intelligence) to graphically represent a place described in a literary text. Similar operations, relying on natural intelligence, creativity, and talent (all exclusively human terms), have been conducted for centuries (e.g., paintings of biblical and mythological subjects, editorial illustrations, storytellers' panels) and belong to a now well-understood logic of visually translating narrative forms. This widespread and precise anchoring within the realm of expectations would, in itself, provide sufficient grounds for evaluating the results. However, an artificial feedback-based evaluation system has been implemented through the reverse elaboration of the graphic product.

Given the linguistic distance between the semantic tools of the envisioned figurative representation and those of the literary form, the experiment aims to define a possible and effective logical pathway. It seeks to verify the hypothesis that text analysis and its subsequent synthesis into a prompt for generative AI can successfully be entrusted to a bot. To this end, the procedure was divided into two phases, termed text-to-prompt (TTP) and prompt-to-image (PTI). The TTP elaboration phase is preliminary and was conducted through the *ChatGPT-4* bot, chosen due to the absolute predominance and widespread adoption of its algorithm. Among the options evaluated –Plus, Consensus, and WFM– the Plus version was selected for its greater flexibility and tendency not to resort to descriptors external to the text under analysis. The TTP elaboration works on the text to extract and organize descriptive elements according to the priorities and comparative capacities of GANs, which will intervene in the PTI phase.

In preparing for the PTI phase, the three main generation algorithms were considered: *DALL-E* (OpenAI), *Stable Diffusion*, and *Midjourney*. Four parameters were observed: perceptual effectiveness of scene geometry; refinement of details and textures; additional commands to regenerate variations and enhance specific details; and interface flexibility.

Midjourney was selected based on empirical –if not heuristic– evaluations, as no documented information exists in official or parallel literature on the functional details of these algorithms. Performance assessments were conducted in strictly application-based terms. *Midjourney* was utilized via the *ImageArt* interface [Nichol 2021, pp. 6546-6471; Radford 2021, pp. 8748-8763; Saharia 2022, pp. 9110-9126].

Case Study and Proceduree

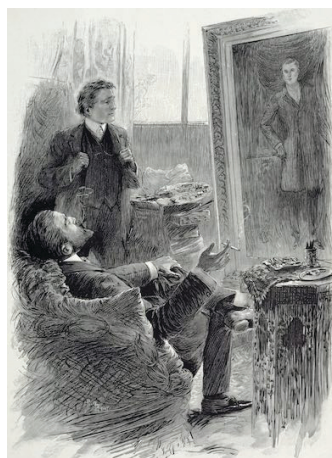
The Picture of Dorian Gray contains intellectual elements that resonate particularly well with the theme. However, this is an essentially human, cultural assessment. Wilde's text was selected for four concrete and independent reasons: the universal renown of the work; the clarity of its historical and urban context; the meticulous descriptions of settings; and the broad possibility for comparison with outputs of natural intelligence (figs. 1, 2), as suggested by the experimental premises.

Among the many suitable excerpts, the opening description of Basil Hallward's studio was chosen due to its richness and completeness. The Italian translation begins as follows:



Fig. 1. Albert Lewin, still frame from *The Portrait of Dorian Gray*, motion picture, 1945.

Fig 2. Paul Thiriat, illustration from *The Picture of Dorian Gray*, engravings collection, 1907.



“Lo studio era intriso d'uno splendido odore di rose e quando la lieve brezza estiva strusciava tra gli alberi del giardino, dalla porta aperta penetrava il profumo pesante delle serenelle, o quello più delicato dei rosaspini [...] nell'angolo di un divano coperto da stoffe persiane [...] di quando in quando l'ombra fantastica di un uccello volante si proiettava e scorreva sulle pesanti tende di seta [...] lo strepito di Londra pareva la vibrazione delle note basse di un organo lontano. In mezzo alla camera, su un cavalletto, era il ritratto a figura intera di un giovane di singolare bellezza” [Wilde 2013, p. 9].

It was observed that providing explicit information regarding the excerpt's origin is irrelevant to the bot. The *ChatGPT-4* decision tree appears to be trained on literary works, at least on this one, suggesting the possibility of evolving the experiment using purpose-written texts.

To evaluate the bot's actual ability to identify the origin of the excerpt, 13 independent TTP (text-to-prompt) control procedures were conducted: three starting from the French translation, one from German, one from Spanish, and eight from Italian. It was verified that the text was consistently recognized, and the bot invariably proceeded using the original version (Victorian-era British English). While the distinctive elements of the original text were retained in the synthesized output (the prompt), the analytical phase (content schema) was presented in the interface's selected language, in this case, Italian (figs. 3, 4).

Fig 3. Scene synthesis by BOT.

Fig 4. Visual elements synthesis by BOT.

Scena:

- Una stanza luminosa e piena di un profumo floreale.
- Fiori di rose fresche e serenelle, con un tocco di profumo di rosaspini.
- Una leggera brezza estiva attraversa gli alberi visibili da una porta aperta.

Dettagli focali:

- Lord Henry Walton sdraiato su un divano ricoperto da tessuti ricchi.
 - Sta fumando sigarette.
 - Può vedere i rami di un'acacia fuori dalla porta.
 - I rami sono fragili, pieni di fiori colorati e dolci come il miele.

Atmosfera:

- Ombre fugaci di uccelli proiettate su tende pesanti di seta, evocando un effetto giapponese.
- L'ombra richiama pittori di Tokyo, caratterizzati da un'estetica di velocità e movimento.

Sonorità e ambientazione:

- Ronzio di api tra l'erba alta e attorno ai caprifogli.
- Lontano ronzio urbano che sembra il suono di un organo basso.

Oggetto centrale:

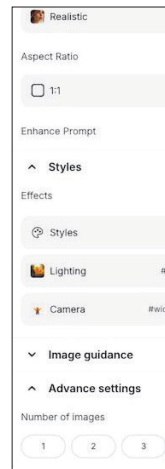
- Un ritratto a figura intera di un giovane di straordinaria bellezza, posto su un cavalletto.
- Basil Hallward, l'artista, è seduto poco distante.

ChatGPT

Sintesi degli elementi visivi principali dello studio di Basil Hallward:

- Illuminazione e atmosfera:**
 - Lo studio è inondato di luce naturale che entra dalla porta aperta e dalle finestre.
 - Un senso di calma e raffinatezza permea l'ambiente, arricchito da profumi floreali.
- Dettagli floreali e naturali:**
 - Il profumo delle rose si mescola a quello delle serenelle e dei rosaspini.
 - Dall'esterno, si intravedono alberi e rami di acacia con fiori dolci e fragili.
- Oggetti nello studio:**
 - Un cavalletto centrale regge un ritratto a figura intera di un giovane di straordinaria bellezza.
 - Una sedia o un divano ricoperto di tessuti ricchi e decorativi.
 - Tende di seta pesante che si muovono leggermente con la brezza estiva.
- Elementi dinamici:**
 - Ombre fugaci di uccelli si proiettano sulle tende, evocando un effetto estetico giapponese.
 - Il ronzio delle api e il suono lontano della città creano una sinfonia naturale e urbana.

Fig 5. Midjourney's parameters setup window.



The final TTP experience was conducted using the Italian translation. The most significant semantic differences between the analytical and synthetic results of the various procedures emerged in the phrasing of the request.



Fig 6. Generation with 'high contrast'

The formulations that produced the best analysis and synthesis results were the following four:

1. from this text, create a prompt for generating the corresponding image with *Midjourney* and textually highlight the essential elements;
2. Transform the following text into an optimized prompt for graphical elaboration with *Midjourney* and generate an analytical textual schema describing the key elements;
3. Create a prompt suitable for *Midjourney* to generate the image described in the following text and synthesize an analytical schema of the most relevant elements;
4. Generate a prompt suitable for *Midjourney* for the graphical elaboration of the scene described in this text and create a textual analytical summary of the main elements.

For a 'human' reader, these requests are equivalent. However, linguistic and semantic choices unpredictably influence the AI's output, as natural language's co-significance relationships are not pre-served. Linguistically near-synonymous terms (e.g., create/generate; propose/present; text/passage/phrase/period; isolate/define; element/object/part/component) are interpreted differently, producing prompts that vary widely according to seemingly random logic. These "lost-in-translation" issues seem to stem from unresolved (and perhaps statistically irresolvable) problematic nodes within the bot's decision trees. Beyond natural language nuances, no differences were observed between re-quests made in English (widely presumed, though undocumented, to be *ChatGPT-4*'s training language) or Italian.



Fig 7. Generation with 'horizontal wide shot'.



Fig 8. Generation with 'aerial view'.

Practical preliminary application tests were conducted on the various TTP-generated prompts. The four best prompts were used to feed *Midjourney* with multiple configurations of the generative engine to stabilize the generative parameters and identify the prompt most capable of producing images closely aligned with the request:

'A luminous studio filled with the scent of roses, lilacs, and hawthorn blossoms. A summer breeze flows gently through an open door, revealing fragile acacia branches with sweet, honey-like flowers. A silk-covered couch. Shadows of birds flit across heavy silk curtains, creating a Japanese-inspired effect. Bees hum over uncut grass and golden honeysuckle. In the center of the room, a full-length portrait of a strikingly beautiful young man stands on an easel'.

Of particular interest is the absence of explicit references to the text, the author, the Victorian London atmosphere, or the furniture style. It is inevitable to conclude that the explicit identification provided by *ChatGPT-4* in the analytical phase is somehow also possible for *Midjourney's* GANs, and that the bot "recognizes" this capability.

The prompt (generated from the request in point 3 of the above list) was processed by *Midjourney* under the following controlled parameters:

(a) image aspect ratio, (b) resolution, (c) angle and perspective, (d) graphic characteristic, (e) tone and intensity of lighting (fig. 5).

120 generations were conducted with the following parameters:

(a) 1:1; (b) HR 4096x4096; (c) horizontal wide shot; (d) no-style; (e) high contrast (fig. 6).

400 generations were conducted with the following parameters:

(a) 1:1; (b) HR 4096x4096; (c) horizontal wide shot; (d) realistic; (e) natural sunlight (fig. 7). 352 generations were conducted with the following parameters: (a) 1:1; (b) HR 4096x4096; (c) aerial view; (d) realistic; (e) natural sunlight (fig. 8). In total, the 872 independent procedures, each producing 2 or 4 images, yielded a total of 2,268 images.

Results

Out of the 2,268 generated images, a selection process was undertaken to identify a single graphic output deemed the final result of the generative procedure (fig. 9).

The final image incorporates all the elements described in the text's incipit, including its literary atmosphere and style, architectural context, and, to a large extent, the historical setting. The latter does exhibit minor inaccuracies arising from suboptimal choices in furnishings and decorations; however, based on the numerous human representations of the same subject analyzed, these inaccuracies appear to belong to the category of 'deviations from convention' typical of any real environment and are not disruptive. Rather, they serve as citations that enhance realism (fig. 10).

Nevertheless, *Midjourney* does not perform internal evaluations, not even indicative ones (e.g., based on parameter adjustments), and is incapable of self-assessing the adherence of the



Fig 9. Final image.



Fig 10. Variety and co-presence in Victorian furniture.



Fig 11. Set of the 9 control images.

graphic output to the provided prompt. For this reason, after the necessarily human-conducted selection operation, the final image, along with a control set of nine additional images (fig. 11), was subjected to feedback testing via AI.

The prompt enhance function of *Midjourney* was applied (the algorithm generates a new prompt by interpreting the generated image as if it were an external source, meaning the original prompt is forgot-ten and entirely ignored). The original image's prompt and those of the 9 control images autogenerated by *Midjourney* were submitted to the interpretation of *ChatGPT-4*, following a reverse logic from ITP (image-to-prompt) to PTT (prompt-to-text).

The prompt obtained from the final image was:

A wide establishing shot of a grand Victorian-era painter's atelier in London, showcasing the expan-sive hall with elegant architectural details. The scene is framed to capture the grandeur of the space, featuring a central easel displaying a captivating portrait of a young man, surrounded by three additional easels showcasing other portraits. The walls are adorned with numerous paintings, reflecting the artistic legacy of the room. Natural sunlight floods through a large window, casting warm and soft tones that create a golden hour glow. Shadows are cast naturally

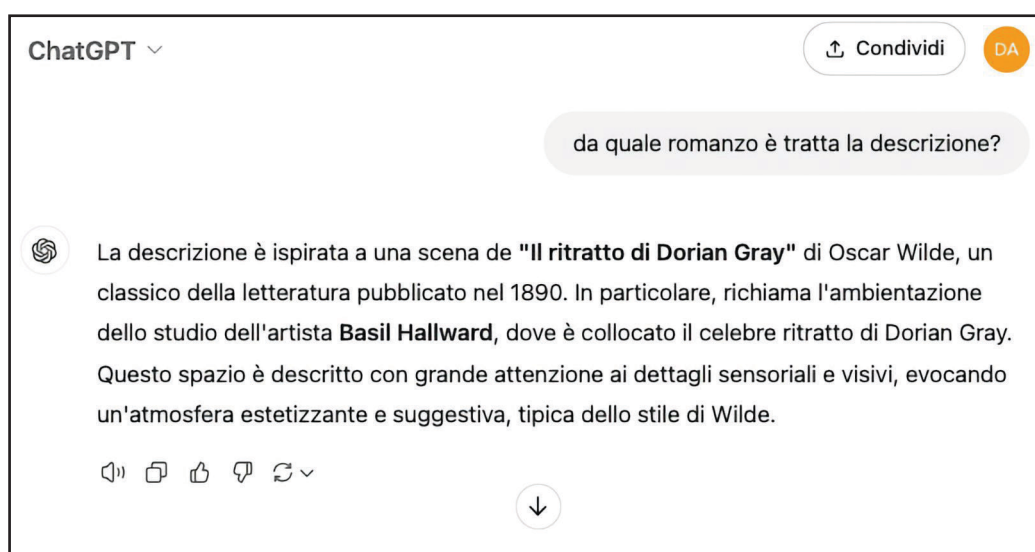


Fig 12. PTT test result on the final image.

across the hall, enhancing the ambient outdoor lighting and highlighting the silk-covered Victorian-style couch that adds a touch of elegance to the scene. Gentle highlights dance on surfaces, and subtle reflections of greenery add depth to the earthy color palette. An open door invites a refreshing summer breeze, rustling through delicate acacia branches adorned with sweet, honey-like flowers. The shadows of birds flitting across heavy silk curtains create a serene effect, reminiscent of Japanese aesthetics. In the foreground, bees hum gently over uncut grass and golden honeysuckle, enriching the tranquil atmosphere. The balanced composition and deep focus clarity draw the viewer into this picturesque moment, encapsulating the essence of a bygone era with sweeping clarity and cinematic perspective.

The prompts obtained from the control set differed but retained elements of similarity. When subjected to the prompt-to-text analysis by *ChatGPT-4*, the prompts generated through the image-to-prompt procedure led to a consistent recognition by the system, both for the final image and the 9 control images (fig. 12). This recognition occurred even though the generated prompts contained only a few original words from Wilde's text.

In conclusion, the evaluation conducted via the feedback test demonstrated that the prompts autogenerated by *Midjourney* include elements inherent to AI processing.

These elements, uniquely characteristic of AI, invariably connect the generated images to the precise literary source.

Results Conclusions and developments

As is well known, AI algorithms operate in a manner heavily influenced by the global training of the system. The reasons that led to the selection of the opening passage from *The Picture of Dorian Gray* are likely the same that enabled the success of the feedback test.

Additionally, we believe that the initial prompt-generation procedure contributed to a form of marginal training that was assimilated by the BOT. This hypothesis should be tested through the repetition of the experiment using a text unfamiliar to the global training and on systems belonging to different experiential chains (e.g., BOTs linked to independent local memories or distinct user bases).

However, we also posit that AI systems are capable of generating graphic and analytical forms that are unrecognizable at a natural level but evident to artificial interpretation itself.

Reference List

- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M. (2021). Towards photorealistic image generation and editing with text-guided diffusion models. In *ICML*, 2022 pp. 6456-6471. arXiv preprint arXiv:2112.10741.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *ICML*, 2021, pp. 8748-8763. <https://doi.org/10.48550/arXiv.2103.00020>.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, M. Norouzi, D. (2022), Image-to-image diffusion models. In *ACM SIG- GRAPH 2022 Conference Proceedings*, 2022, pp. 9110-9126.
- Wilde, O. (2013). *Il ritratto di Dorian Gray*. Milano: Feltrinelli.

Author

Daniele Giovanni Papi, Politecnico di Milano, daniele.papi@polimi.it

To cite this chapter: Daniele Giovanni Papi (2025). Procedure for TTP/PTI in Generative AI and Reverse Processing Tests. In L. Carlevaris et al. (Eds.), *èkphrasis. Descrizioni nello spazio della rappresentazione/èkphrasis. Descriptions in the space of representation*. Proceedings of the 46th International Conference of Representation Disciplines Teachers. Milano: FrancoAngeli, pp. 3121-3140. DOI: 10.3280/oa-1430-c917.