èkphrasis

46° Convegno Internazionale
dei Docenti delle Discipline della Rappresentazione
Congresso della Unione Italiana per il Disegno

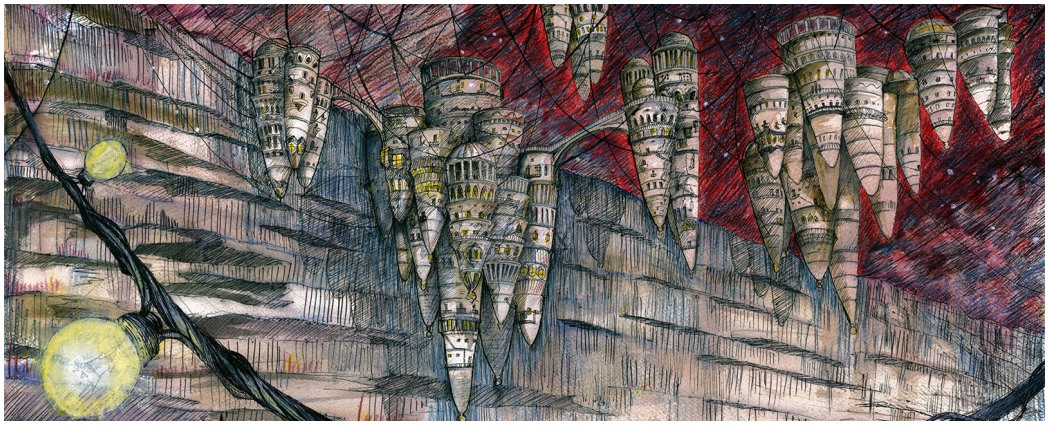
Èkphrasis e AI generativa: riflessioni analogico/digitali nell'immaginario de *Le città invisibili* di Calvino

Maria Grazia Cianci
Daniele Calisi
Stefano Botta
Sara Colaceci
Michela Schiaroli

Abstract

L'avvento dell'intelligenza artificiale (AI) ha avuto un impatto significativo su numerosi ambiti creativi, tra cui la generazione di immagini. L'esplosione tecnologica degli strumenti di generazione *text-to-image* ha permesso la produzione rapida di opere visive sulla base di semplici descrizioni testuali, democratizzando l'accesso alla produzione visiva. Tuttavia, questo progresso porta con sé interrogativi di natura tecnica, metodologica ed etica. La questione della riproducibilità dell'arte, già affrontata da Walter Benjamin con l'avvento della fotografia e del cinema, trova oggi nuove declinazioni nel contesto dell'AI, mettendo in discussione concetti come originalità, creatività e diritto d'autore.
Questo studio si propone di analizzare il fenomeno dell'AI generativa per la creazione di immagini attraverso un confronto con le tecniche analogiche tradizionali. Prendendo spunto dall'opera *Le città invisibili* di Italo Calvino del 1972, verrà esplorato come la rappresentazione grafica di città immaginarie vari a seconda del mezzo utilizzato, evidenziando le differenze stilistiche e metodologiche tra l'interpretazione umana dei testi rispetto a quella algoritmica dell'intelligenza artificiale.

Parole chiave
Generative AI, *spatial visualization*, Calvino, *representation*, *city*.



Città di Ottavia, china e pastello su carta (autore: Nicola d'Addario).

doi.org/10.3280/oa-1430-c944

Introduzione

Il settore dell'intelligenza artificiale sta subendo una vera e propria esplosione in numerosi campi, specie per quanto riguarda gli strumenti di generazione di immagini, grazie soprattutto alla diffusione e popolarizzazione di applicazioni sperimentali che permettono a chiunque, mediante poche parole e opzioni, di realizzare rapidamente opere digitali di ogni tipo. Ciò non senza scetticismo e numerosi interrogativi su vari temi, fra cui quelli tecnici e metodologici. L'avvento e l'utilizzo di queste tecnologie rimarcano altresì questioni etiche fondamentali sul diritto d'autore e sulla riproducibilità dell'opera d'arte. Proprio come Walter Benjamin poneva tali questioni con l'avvento della fotografia e del cinema [Benjamin 2014], oggi l'utilizzo dell'intelligenza artificiale utilizzabile praticamente da chiunque [Manovich 2018] rischia di annullare tutti quei codici universalmente riconosciuti fino ad ora utilizzati per definire e codificare l'arte e l'unicità dei vari artisti.
La ricerca vuole approfondire il tema della generazione *text-to-image* da intelligenza artificiale, mediante un confronto con la rappresentazione tramite tecniche analogiche tradizionali; lo studio trae ispirazione dal testo *Le città invisibili* di Italo Calvino, in cui l'autore riporta il racconto di cinquantacinque città d'invenzione [Calvino 1972].
Partendo dalle illustrazioni realizzate nell'ambito del corso di *Disegno dell'Architettura,* si propone un'analisi attraverso l'interpretazione grafica di una selezione di città immaginarie, in cui l'accento è posto non soltanto sulle differenze stilistiche delle visioni ottenute, ma soprattutto sui meccanismi che si instaurano nella traduzione di parole in segni, figure e scene. Tale processo, declinato alle necessità di sintesi derivanti dall'AI, pone altresì questioni importanti da tenere in considerazione per quanto riguarda la riduzione di una sintassi complessa in poche parole chiave capaci di restituire il medesimo messaggio.
Inoltre, si intende porre un accento sulla relazione fra *background* culturale e generazione di immagine. Può apparire evidente come, sulla base del medesimo testo, persone differenti possano dare vita a mondi particolarmente diversi; lo stesso Calvino, in un'intervista, sottolinea come siano "gli scenari dei primi anni della nostra vita quelli che danno forma al nostro immaginario" [1], e così anche il nostro intero bagaglio personale. È invece interessante analizzare quale sia l'estensione e il limite di tale affermazione quando si tratta di immagini generate dalle AI, in relazione al 'bagaglio iconografico' a cui tali strumenti fanno riferimento, non solo in confronto alle opere analogiche ma anche fra piattaforme differenti.

Lo stato dell'arte: AI generativa e tecniche d'immagine tradizionali

Concettualmente, i sistemi di generazione di immagini basati su AI operano attraverso reti neurali avanzate come le *Generative Adversarial Networks* (GAN) e i modelli di diffusione, capaci di sintetizzare visualizzazioni complesse a partire da descrizioni testuali [McCormack *et al.* 2019]. Questo processo si distingue dalle tradizionali tecniche artistiche per la sua capacità di elaborare enormi *database* iconografici, spesso addestrate su vasti *dataset* visivi, generando immagini in pochi secondi [Boden 1998]. Tra gli strumenti attualmente disponibili figurano *Midjourney, Stable Diffusion, Leonardo AI* e il più conosciuto *DALL·E* di *OpenAI*, che trasformano *input* testuali in immagini dettagliate. Questi sistemi, infatti, sono basati sulla capacità della macchina di interpretare dei testi o delle parole chiave per poi sintetizzarle visivamente secondo modelli statistici [Manovich 2018].
Le tecniche artistiche tradizionali, al contrario, sono caratterizzate da una maggiore soggettività e interpretazione personale. L'arte manuale richiede un processo di traduzione più articolato, che coinvolge il *background* culturale, la formazione artistica e la sensibilità individuale dell'autore [Grau 2002]. L'atto creativo umano è, in questo senso, influenzato dal vissuto personale e dal contesto storico-sociale dell'autore dell'immagine [Cianci, Calisi 2014].
Nel suo libro *Le città invisibili* Italo Calvino offre un modello di narrazione in cui il lettore è invitato a immaginare mondi alternativi attraverso descrizioni evocative. Secondo lo scrittore, le immagini mentali sono profondamente radicate nelle esperienze personali

di ciascun individuo, un concetto che solleva interrogativi sul modo in cui le AI, prive di esperienza diretta, possano interpretare simili descrizioni, ed è proprio questo uno dei fattori che sollevano domande sul ruolo dell'intelligenza artificiale nella creatività e nell'autorialità dell'opera d'arte [Zylinska 2020].
Questo studio si propone infatti di verificare come la traduzione automatica di tali descrizioni in immagini visive possa divergere dalla rappresentazione umana, influenzata dal bagaglio culturale e dall'esperienza individuale [Floridi 2014].

Metodologia della ricerca

La ricerca, basata sul confronto tra le illustrazioni realizzate dagli studenti del corso di *Disegno dell'Architettura* e le immagini generate da AI, è volta ad indagare come questa potesse rispondere se interrogata in merito alla creazione di immagini partendo da indicazioni testuali. L'obiettivo perseguito è determinare come il processo di traduzione da parola a immagine vari tra umano e intelligenza artificiale, evidenziando le sfide e i limiti dei due approcci. La metodologia di ricerca si è sviluppata in cinque fasi consequenziali (fig. 1), ed in alcuni casi contigue.
La prima fase ha riguardato la selezione delle illustrazioni degli studenti più idonee al confronto, andando a privilegiare non solo le immagini più accattivanti e suggestive dal punto di vista grafico/artistico ma soprattutto che fossero coerenti alla città invisibile che si riproponeva di raccontare.

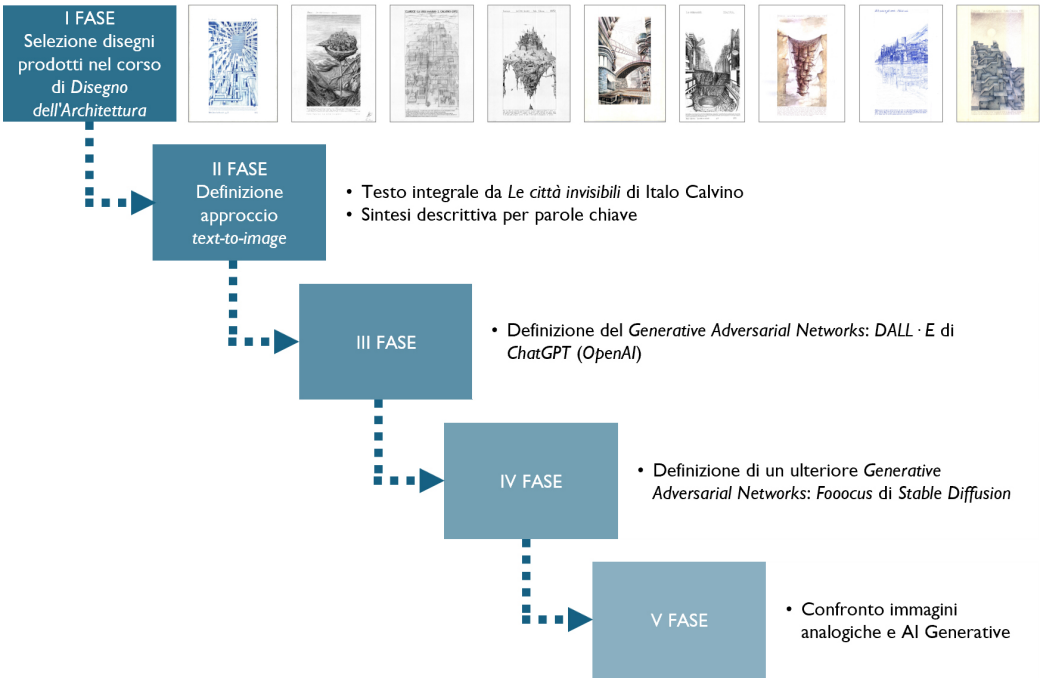


Fig. 1. Diagramma della metodologia di ricerca (fonte: immagine degli autori).

In seconda battuta è stato scelto il tipo di descrizione testuale da utilizzare; per questa fase, l'approccio *text-to-image* è stato duplice utilizzando sia il testo integrale tratto da *Le città invisibili* di Calvino, che attraverso l'utilizzo di parole chiave che si ritiene possano esprimere al meglio l'essenza della città descritta.
Successivamente sono state definite le *Generative Adversarial Networks* con cui effettuare le sperimentazioni, scegliendo di utilizzare *DALL·E* di *OpenAI*. Tuttavia, per avere una ulteriore lettura sono state aggiunte sperimentazioni anche con il *tool* per l'arte digitale con *Stable Diffusion*. La quarta fase ha riguardato l'analisi comparativa.

Le immagini ottenute sono state, pertanto, indagate seguendo diversi criteri valutativi, dalla fedeltà al testo originale o comunque alle parole chiave fornite, all'impatto estetico inteso come qualità visiva e armonia delle composizioni, ed infine alla capacità evocativa e cioè il grado di suggestione e profondità interpretativa raggiunte [Mitchell 1994].
Inoltre, è stato esaminato il grado di riduzione semantica che avviene nel passaggio dal linguaggio descrittivo al linguaggio visivo sintetizzato dall'AI [Grau 2002].

AI generativa ed interpolazione del linguaggio

L'efficacia di un sistema di intelligenza artificiale generativa *text-to-image* dipende non solo dalla qualità dell'algoritmo di generazione, ma anche da come il modello interpreta e suddivide il testo del *prompt*. Per questo motivo, la comprensione del processo di lettura e tokenizzazione diventa fondamentale per ottimizzare la scrittura dei *prompt* e ottenere risultati coerenti con l'intenzione dell'utente.
La tokenizzazione è il processo di suddivisione del testo in *token,* ovvero unità minime di significato, che possono essere parole intere, sotto-parole o persino singoli caratteri. Esistono diverse strategie di tokenizzazione:
- per carattere, più granulare, utile per lingue senza spazi come il cinese. Ciò a fronte di una maggiore lentezza nell'elaborazione e una più difficile interpretazione del significato semantico dei contesti in cui le parole sono inserite;
- *subword,* utilizzata nei modelli avanzati (*Byte-Pair Encoding* – BPE, *WordPiece*), che suddivide parole complesse in segmenti più piccoli per una migliore gestione del vocabolario,

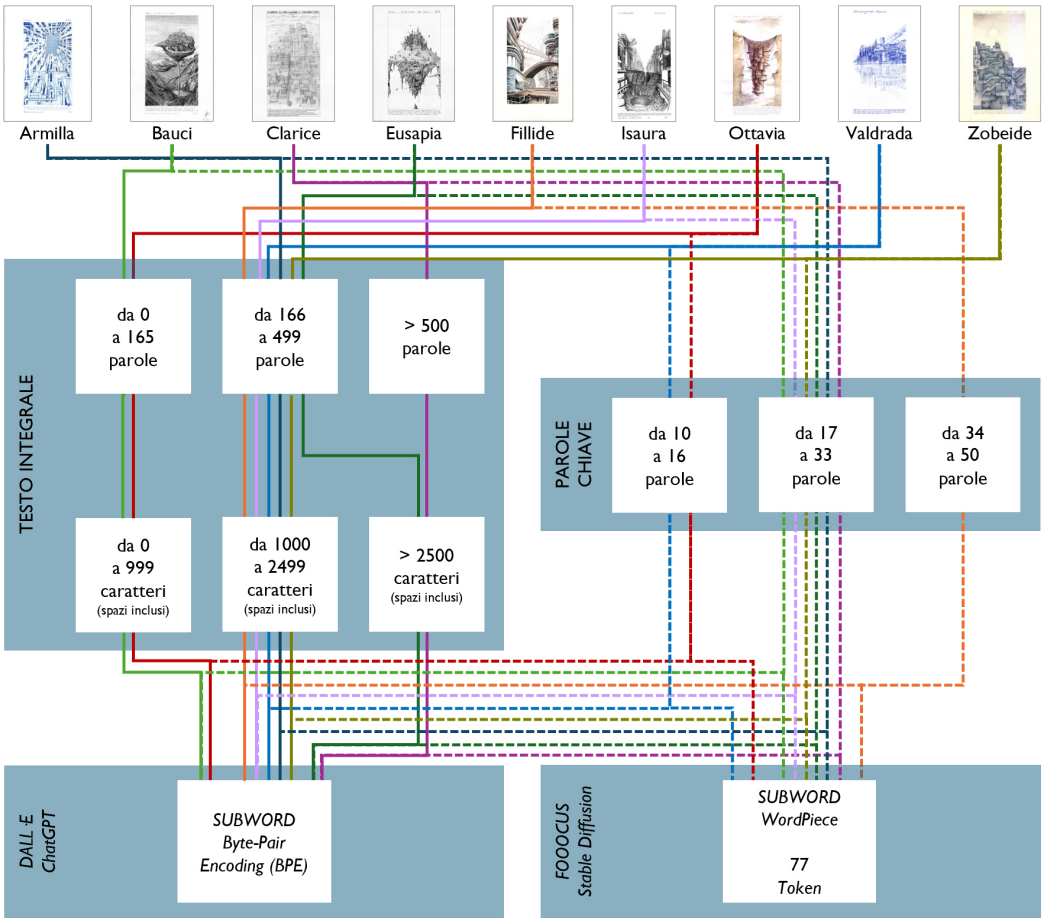


Fig. 2. Diagramma della struttura semantica di linguaggio della AI generativa indagata nella ricerca (fonte: immagine degli autori).

specie in presenza di parole rare o complesse. Ciò riduce il numero di parole sconosciute (*out-of-vocabulary,* OOV) ma richiede prestazioni maggiori;
- per parola, che separa il testo basandosi sugli spazi e sulla punteggiatura. Per quanto più semplice da implementare di altre, ha lo svantaggio di gestire in maniera poco ottimale la varietà grammaticale (mangiare, mangiato, mangiando saranno tre token diversi e non relazionati fra loro);
- *sentencePiece* considera più parole e caratteri insieme, gestendo una complessità grammaticale maggiore. Ciò rende il metodo più funzionale nel caso di modelli AI multilingua e/o con alfabeti diversi, con il problema però di avere spesso *token* molto lunghi [Mielke 2021].

Effettuando un confronto tra due modelli conosciuti di AI generativa *text-to-image, DALL·E* (*ChatGPT*) e *Fooocus* (basato su *Stable Diffusion*), si evidenziano le differenze nel modo in cui questi analizzano e comprendono un *prompt* testuale [Jamal 2024]. *DALL·E* utilizza B*yte-Pair Encoding* (BPE), un sistema *subword* ottimizzato per la comprensione contestuale, permettendogli di processare frasi articolate mantenendo coerenza semantica [OpenAI 2023]. *Fooocus,* invece, si basa su *CLIP* [OpenAI 2021], che usa anch'esso una tokenizzazione ma con un limite massimo di 77 *token*, riducendo la capacità del modello di gestire descrizioni lunghe e complesse (fig. 2).
Di conseguenza, mentre *DALL·E* riesce a interpretare *prompt* dettagliati con un alto grado di fedeltà, *Stable Diffusion* tende a dare più peso alle parole chiave, tralasciando parte delle informazioni in caso di *input* troppo estesi [Ramesh e*t al.* 2022].
Inoltre, la lingua (*natural languag*e) in cui il *prompt* viene scritto è un ulteriore fattore discriminante: *DALL·E* gestisce meglio lingue diverse dall'inglese grazie a un'ampia esposizione multilingue, mentre *Stable Diffusion* è fortemente ottimizzato per l'inglese [Zhang 2023]. Per ottenere i migliori risultati, diventa essenziale adattare il *prompt* al sistema utilizzato, scegliendo con attenzione la struttura linguistica e la disposizione delle parole chiave. Ciò può, quindi, richiedere un processo ulteriore di traduzione e filtraggio delle informazioni, che di fatto altera le sfumature di significato del *prompt* originale.
La scrittura dei *prompt* nelle AI generative non è perciò solo una questione creativa, ma anche e soprattutto tecnica, che richiede la conoscenza delle modalità di lettura del modello per ottenere risultati il più possibile aderenti alle intenzioni dell'utente.

Analisi e risultati della ricerca: dalle tecniche manuali alle differenti piattaforme di AI generativa

L'analisi comparativa ha evidenziato significative differenze tra le due modalità di rappresentazione, disegno analogico da 'intelligenza naturale' e prodotto digitale da intelligenza artificiale. Gli aspetti dei quali si è tenuto conto riguardano la sintesi visiva verso l'interpretazione soggettiva, il ruolo del bagaglio culturale e la fedeltà semantica.
Rispetto al primo punto – quello del confronto tra sintesi visiva e interpretazione soggettiva – va sicuramente sottolineato che, mentre le illustrazioni 'manuali' tendono ad enfatizzare alcuni dettagli piuttosto che altri in base alle inclinazioni personali dell'autore [Cianci, Calisi 2014], come ad esempio le sperimentazioni con le proporzioni e un utilizzo variegato di metodologie prospettiche differenti, l'AI generativa tende invece sempre a semplificare descrizioni complesse in elementi iconografici standardizzati, in base alle immagini presenti nei suoi *dataset* [Manovich 2002] e, nonostante i tentativi, non è stata in grado di soddisfare richieste testuali rispetto al metodo di rappresentazione desiderato.
Questo è stato il caso con le prove effettuate con le città di Armilla (fig. 3) e Fillide (fig. 4) dalle quali si evince una più sviluppata capacità prospettica della 'intelligenza naturale'.
La maggiore varietà stilistica dimostrata dagli autori delle immagini analogiche è la dimostrazione di come le esperienze personali ed il *background* culturale, mentre le AI producono tendenzialmente immagini più uniformi che suggeriscono il verificarsi di un *bias* derivante dai dati di addestramento [Nochlin 1971; Zylinska 2020]; si può quindi sostenere che entrambe le intelligenze presentino dei *bias*, ma questi portano a dei risultati opposti dal punto di vista della resa grafica e della rappresentazione nella creazione di immagini.

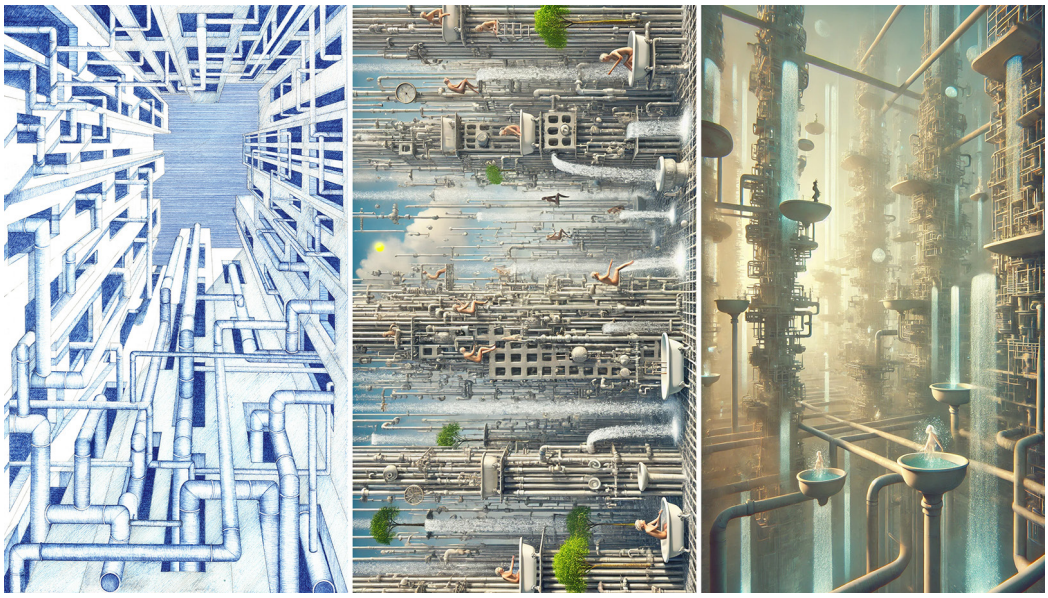


Fig. 3. Città di Armilla, pastello su carta (sx), immagine ottenuta utilizzando il testo integrale in *DALL·E (ChatGPT)* (centro), immagine ottenuta utilizzando parole chiave in *DALL·E (ChatGPT)* (dx) (fonte: autore del disegno anonimo a sx, immagini degli autori al centro e a dx).

Fig. 4. Città di Fillide, china e pastello su carta (sx), immagine ottenuta utilizzando il testo integrale in *DALL·E (ChatGPT)* (centro), immagine ottenuta utilizzando parole chiave in *DALL·E (ChatGPT)* (dx) (fonte: autore del disegno a sx Valerio Pasquali, immagini degli autori al centro e a dx).

Questo appare chiaro nell'esperienza fatta sulla città di Ottavia (fig. 5), nella quale la lettura integrale del testo di Italo Calvino l'AI generativa ha portato alla completa eliminazione della componente vegetazionale (elemento peraltro ignorato anche dall'autrice umana) presente nel racconto e, solo con un'attenta selezione di parole chiave, questo aspetto è stato ripristinato nell'immagine digitale creata.
La differenza sostanziale tra i due approcci avviene tuttavia quando si confronta la fedeltà semantica dei rispettivi risultati.
Se la 'intelligenza naturale' riesce a elaborare una sintesi tenendo presente gli elementi più significativi di un racconto ciò non avviene per le immagini generate con AI, che tende invece a privilegiare aspetti descrittivi espliciti e sintetici, non riuscendo a comprendere concetti più astratti o poetici, ciò porta ad una riduzione delle metafore calviniane a elementi più concreti [McCormack *et al.* 2019]. Nel caso della città di Isaura ad esempio (fig. 6), pur riuscendo a generare immagini suggestive, almeno in questo caso, l'intelligenza

Fig. 5. Città di Ottavia, china e acquerello su carta (sx), immagine ottenuta utilizzando il testo integrale in *DALL·E (ChatGPT)* (centro), immagine ottenuta utilizzando parole chiave in *DALL·E (ChatGPT)* (dx) (fonte: autrice del disegno a sx Giulia Gherardi, immagini degli autori al centro e a dx).

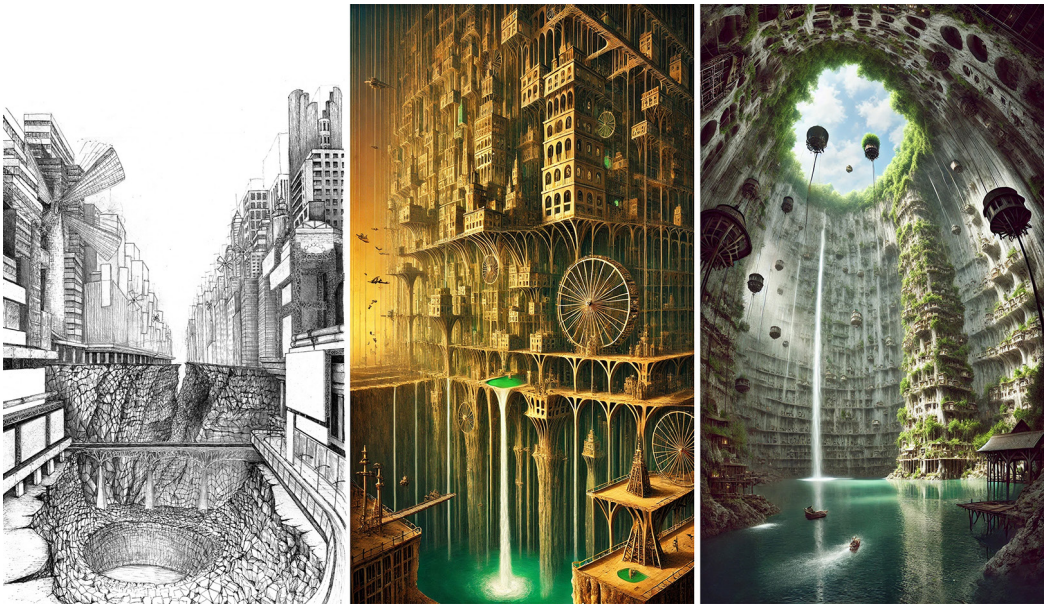


Fig. 6. Città di Isaura, china su carta (sx), immagine ottenuta utilizzando il testo integrale in *DALL·E (ChatGPT)* (centro), immagine ottenuta utilizzando parole chiave in *DALL·E (ChatGPT)* (dx) (fonte: autrice del disegno a sx Beatrice Evangelisti, immagini degli autori al centro e a dx).

artificiale ha prodotto delle immagini riproducendo gli aspetti strettamente descrittivi del racconto Calviniano. Tuttavia, la ricerca non appariva sufficientemente esaustiva utilizzando solamente *DALL·E* (*ChatGPT*) (fig. 7), pertanto si è deciso di verificare la capacità di generare immagini partendo da testi con un altro tipo di *Generative Adversarial Networks: Fooocus* (basato su *Stable Diffusion*).
Ogni piattaforma di AI presenta variazioni nell'approccio ai dati e negli *output* generati, a seconda dell'architettura del modello e della natura del *dataset* utilizzato.
Tuttavia, il confronto con *Fooocus* (basato su *Stable Diffusion*), pur enfatizzando il fotorealismo rispetto a *DALL·E* (*ChatGPT*), non ha generato immagini soddisfacenti dal punto di vista dalla sintesi visiva e interpretazione soggettiva oltre che nella fedeltà semantica; delle città generate con l'AI, di cui le quattro immagini riportate in fig. 8 sono una esemplificazione, solo la prima e la quarta si presentano minimamente suggestive, mentre quelle centrali si rivelano basiche nella descrittività e per nulla suggestive.

Fig. 7. Città di Bauci, china su carta (sx), immagine ottenuta utilizzando il testo integrale in *DALL·E (ChatGPT)* (centro), immagine ottenuta utilizzando parole chiave in *DALL·E (ChatGPT)* (dx) (fonte: autore del disegno a sx Marco Piccoli, immagini degli autori al centro e a dx).

Fig. 8. Città di Armilla (prima da sx), Ottavia (seconda da sx), Bauci (terza da sx), Isaura (quarta da sx), immagini ottenute utilizzando parole chiave in lingua inglese in *Fooocus*, basato su *Stable Diffusion* (fonte: immagini degli autori).

Conclusioni

La ricerca dimostra che, sebbene l'AI sia in grado di produrre immagini visivamente accattivanti, il suo processo differisce radicalmente da quello umano. La traduzione di testi letterari in immagini pone sfide significative, poiché l'intelligenza artificiale tende a ridurre la complessità semantica a *pattern* visivi statisticamente ricorrenti. I risultati sollevano tuttavia questioni rilevanti per il futuro della creatività assistita dalla AI generativa in diversi ambiti come le implicazioni tra creatività e riproducibilità, l'evoluzione del ruolo dell'artista, o ancora l'etica e il diritto d'autore.

LIntelligenza Artificiale offre nuove possibilità creative, ma pone anche interrogativi cruciali sulla natura della creatività stessa. Boden suggerisce che la creatività artificiale si basi su combinazioni e variazioni di elementi preesistenti, piuttosto che su una reale innovazione [Boden 1998].

La questione diventa quindi se la macchina possa mai essere veramente creativa o se rimanga uno strumento di supporto all'immaginazione umana.

Anche il ruolo dell'artista potrebbe essere ridefinito dalla crescente diffusione di strumenti di generazione AI. In un contesto in cui chiunque può creare immagini con pochi comandi testuali, si pone la questione di quale sia il valore dell'arte tradizionale e dell'intervento umano nel processo creativo [Zylinska 2020].

Infine, resta l'annosa questione dell'etica, l'utilizzo dell'AI nella generazione di immagini solleva infatti questioni legate alla proprietà intellettuale e al diritto d'autore. Le immagini prodotte da AI derivano da *dataset* di opere esistenti, sollevando dubbi sulla legittimità dell'uso di tali riferimenti senza il consenso degli autori originali [Floridi 2014].
In conclusione, l'analisi evidenzia il valore insostituibile dell'interpretazione umana nell'atto creativo, sottolineando il rischio di un'omologazione estetica dovuta all'uso di *dataset* preconfezionati. Tuttavia, la combinazione di AI e intervento artistico umano potrebbe aprire nuove prospettive per l'arte e il design. Future ricerche potrebbero approfondire l'interrogativo tutt'ora aperto su come integrare questi strumenti nel panorama artistico e su come regolamentare l'uso dell'AI per garantire una produzione visiva eticamente sostenibile. In un'epoca in cui le frontiere dell'arte sono in continua ridefinizione, il dialogo tra intelligenza artificiale e creatività umana si configura come una delle sfide più stimolanti per il mondo della cultura e della tecnologia.

Nota

[1] La citazione è stata estrapolata dall'intervista dal titolo *Italo Calvino - Un uomo invisibile* che Valerio Riva fece a Italo Calvino l'8 dicembre 1974 per la serie *Incontri - Fatti e personaggi del nostro tempo,* per RSI Radiotelevisione svizzera, con la regia di Nereo Rapetti.

Riferimenti bibliografici

Benjamin, W. (2014). *L'opera d'arte nell'epoca della sua riproducibilità tecnica.* Torino: Einaudi.

Boden, M. A. (1998). Creativity and artificial intelligence. In *Artificial intelligence,* vol. 103, n. 1-2, pp. 347-356. https://doi.org/10.1016/S0004-3702(98)00055-1.

Calvino, I. (1972). *Le città invisibili.* Torino: Einaudi.

Cianci, M. G., Calisi, D. (2014). Il mondo è un libro: visioni ispirate da "Le città invisibili" di Italo Calvino. In A. Garcia Melian (Ed.). *El Dibujo de Viaje de los Arquitectos*. Atti del XV Congreso Internacional Expresion Grafica Arquitectonica EGA2014. Las Palmas de Gran Canaria, 22-23 maggio 2014, pp. 751-759. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality.* Oxford: Oxford University Press.

Grau, O. (2002). *Virtual Art: From Illusion to Immersion.* Cambridge: MIT Press.

Jamal, S., Wimmer, H., Rebman, C. M. Jr. (2024). Perception and evaluation of text-to-image generative AI models: A comparative study of DALL-E, Google Imagen, GROK, and Stable Diffusion. In *Information Systems*, 25(2), pp. 277-292. https://doi.org/10.48009/2_iis_2024_123.

Manovich, L. (2002). *The Language of New Media.* Cambridge: MIT Press.

Manovich, L. (2018). *AI Aesthetics*. Moscow: Strelka Press.

McCormack, J., Gifford, T., Hutchings, P. (2019). Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. *Leonardo,* 52(3), pp. 285-291. https://doi.org/10.48550/arXiv.1903.02166.

Mielke, S. J., Alyafeai, Z., Salesky, E., Raffel, C., Dey, M., Gallé, M., Raja, A., Si, C., Lee, W. Y., Sagot, B., Tan, S. (2021). Between words and characters: A brief history of open-vocabulary modeling and tokenization in NLP. In *arXiv*. Cornell University. https://doi.org/10.48550/arXiv.2112.10508.

Mitchell, W. J. (1994). *The Reconfigured Eye: Visual Truth in the Post-Photographic Era.* Cambridge: MIT Press.

Nochlin, L. (1971). *Why Have There Been No Great Women Artists?.* New York: ARTnews.

OpenAI. (2021). *CLIP: Connecting text and images. OpenAI.* https://openai.com/index/clip/.

OpenAI. (2023). *Introducing ChatGPT and DALL·E. OpenAI.* https://openai.com/dall-e.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 10684-10695. https://doi.org/10.48550/arXiv.2204.06125.

Riva, V. Rapetti, F. (8 dicembre 1974). Intervista a Italo Calvino dal titolo: *Italo Calvino - Un uomo invisibile*. https://www.rsi.ch/archivi/Italo-Calvino-Un-uomo-invisibile--1643011.html.

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. (2023). Text-to-image Diffusion Models in Generative AI: A Survey. In *arXiv*.. https://doi.org/10.48550/arXiv.2303.07909.

Zylinska, J. (2020). *AI Art: Machine Visions and Warped Dreams*. Londra: Open Humanities Press.

Autori
Maria Grazia Cianci, Università degli Studi Roma Tre, mariagrazia.cianci@uniroma3.it
Daniele Calisi, Università degli Studi Roma Tre, daniele.calisi@uniroma3.it
Stefano Botta, Università degli Studi Roma Tre, stefano.botta@uniroma3.it
Sara Colaceci, Università degli Studi Roma Tre, sara.colaceci@uniroma3.it
Michela Schiaroli, Università degli Studi Roma Tre, michela.schiaroli@uniroma3.it

Per citare questo capitolo: Maria Grazia Cianci, Daniele Calisi, Stefano Botta, Sara Colaceci, Michela Schiaroli (2025). *Èkphrasis* e AI generativa: riflessioni analogico/digitali nell'immaginario de *Le città invisibili* di Calvino. 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. 3645-3664. DOI: 10.3280/oa-1430-c944.

 ISBN 9788835182412

èkphrasis

46th International Conference
of Representation Disciplines Teachers
Congress of Unione Italiana per il Disegno

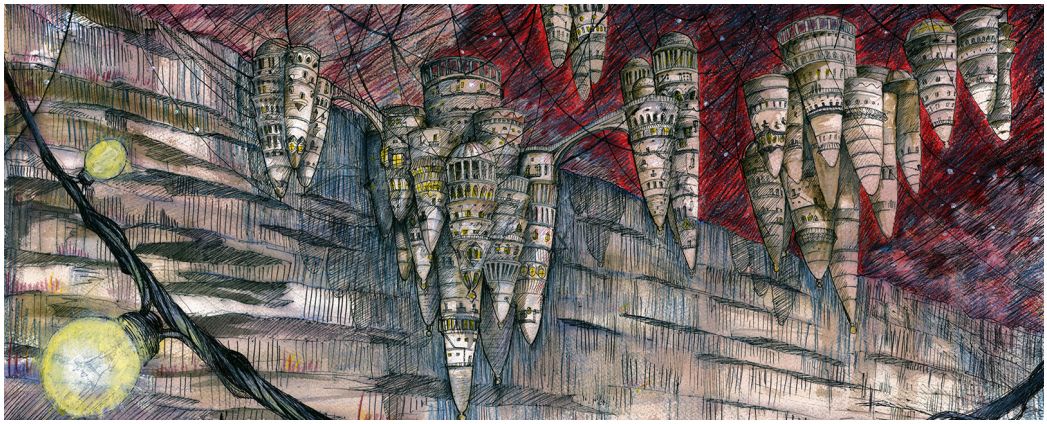
Èkphrasis and Generative AI: Analog/Digital Reflections in the Imaginary of Calvino's *Invisible Cities*

Maria Grazia Cianci
Daniele Calisi
Stefano Botta
Sara Colaceci
Michela Schiaroli

Abstract

The advent of artificial intelligence (AI) has significantly impacted numerous creative fields, including image generation. The technological boom in text-to-image generation tools has enabled the rapid creation of visual works from simple textual descriptions, democratizing access to visual production. However, this advancement raises technical, methodological, and ethical questions. The issue of art reproducibility, previously addressed by Walter Benjamin with the advent of photography and cinema, finds new dimensions today in the context of AI, challenging concepts such as originality, creativity, and copyright.
This study aims to analyze the phenomenon of generative AI for image creation through a comparison with traditional analog techniques. Drawing inspiration from Italo Calvino's 1972 work, *Invisible Cities,* the research will explore how graphic representations of imaginary cities vary according to the medium used. It will highlight the stylistic and methodological differences between human textual interpretation and the algorithmic interpretation of artificial intelligence.

Keywords
Generative AI, spatial visualization, Calvino, representation, city.



City of Ottavia, ink and pastel on paper (source: drawing by Nicola d'Addario).

doi.org/10.3280/oa-1430-c944

Introduction

The field of artificial intelligence is undergoing a true explosion across numerous domains, particularly in the area of image generation tools. This is largely due to the widespread adoption and popularization of experimental applications that allow anyone to quickly create digital artworks of all kinds by using just a few words and options. However, it is developing not without scepticism and numerous questions, including technical and methodological concerns.
The emergence and use of these technologies also bring fundamental ethical issues to the forefront, particularly concerning copyright and the reproducibility of artworks. Just as Walter Benjamin raised such questions with the advent of photography and cinema [Benjamin 2014], today, the accessibility of artificial intelligence to virtually anyone [Manovich 2018] risks erasing all the universally recognized codes traditionally used to define and codify art and the uniqueness of individual artists.
This research aims to explore the topic of text-to-image generation through artificial intelligence by comparing it with representation using traditional analogue techniques. The study takes inspiration from Italo Calvino's *Invisible Cities,* in which the author narrates the stories of fifty-five imaginary cities [Calvino 1972]
Starting from illustrations created within the *Architecture Drawing cours*e, the study proposes an analysis through the graphic interpretation of a selection of imaginary cities. The focus is placed not only on the stylistic differences in the obtained visions but, more importantly, on the mechanisms involved in the translation of words into signs, figures, and scenes. This process, when adapted to the synthesis constraints inherent in AI, also raises significant questions regarding the reduction of complex syntax into a few keywords capable of conveying the same message.
Furthermore, the research intends to highlight the relationship between cultural background and image generation. It is evident that different individuals, when interpreting the same text, may create vastly different worlds. Calvino himself, in an interview, emphasized that "the scenarios of our early years are the ones that shape our imagination" [1], as does our entire personal background. It is, therefore, interesting to analyse the extent and limitations of this statement when it comes to AI-generated images, particularly concerning the 'iconographic baggage' these tools rely on, not only in comparison to analogical artworks but also across different platforms.

State of the art: generative AI and traditional image techniques

Conceptually, AI-based image generation systems operate through advanced neural networks such as (GANs) Generative Adversarial Networks and diffusion models, capable of synthesizing complex visualizations from textual descriptions [McCormack *et al.* 2019]. This process differs from traditional artistic techniques in its ability to process vast iconographic databases, often trained on extensive visual datasets, generating images in just a few seconds [Boden 1998]. Among the currently available tools are *Midjourney, Stable Diffusion* and the most well-known, *DALL·E* by *OpenAI*, which transform textual inputs into detailed images. These systems rely on the machine's ability to interpret texts or keywords and then visually synthesize them based on statistical models [Manovich 2018].
Traditional artistic techniques, in contrast, are characterized by greater subjectivity and personal interpretation. Manual art requires a more elaborate translation process that involves the artist's cultural background, artistic training, and individual sensitivity [Grau 2002]. In this sense, human creativity is influenced by personal experience and the historical-social context of the image's creator [Cianci, Calisi 2014].
In his book *Invisible Cities,* Italo Calvino presents a narrative model in which the reader is invited to imagine alternative worlds through evocative descriptions.
According to Calvino, mental images are deeply rooted in each individual's personal experiences. This concept raises questions about how AI, lacking direct experience, can inter-

pret such descriptions, and this is one of the key factors that prompt debates on the role of artificial intelligence in creativity and artistic authorship [Zylinska 2020].
This study, therefore, aims to examine how the automatic translation of such descriptions into visual images may diverge from human representation, which is influenced by cultural background and individual experience [Floridi 2014].

Research methodology

The research, based on the comparison between illustrations created by students of the *Architectural Drawing course* and AI-generated images, aims to investigate how AI responds when prompted to create images based on textual instructions. The objective is to determine how the process of translating words into images differs between humans and artificial intelligence, highlighting the challenges and limitations of both approaches.
The research methodology developed through five sequential (and, in some cases, overlapping) phases (fig. 1). The first phase involved selecting the students' illustrations most suitable for comparison, prioritizing not only the most visually striking and graphically/artistically compelling images but also those that were most coherent with the invisible city they aimed to depict.
Secondly, the type of textual description to be used was chosen. In this phase, the text-to-image approach was twofold: both the full text excerpt from *Invisible Cities* by Italo Calvino

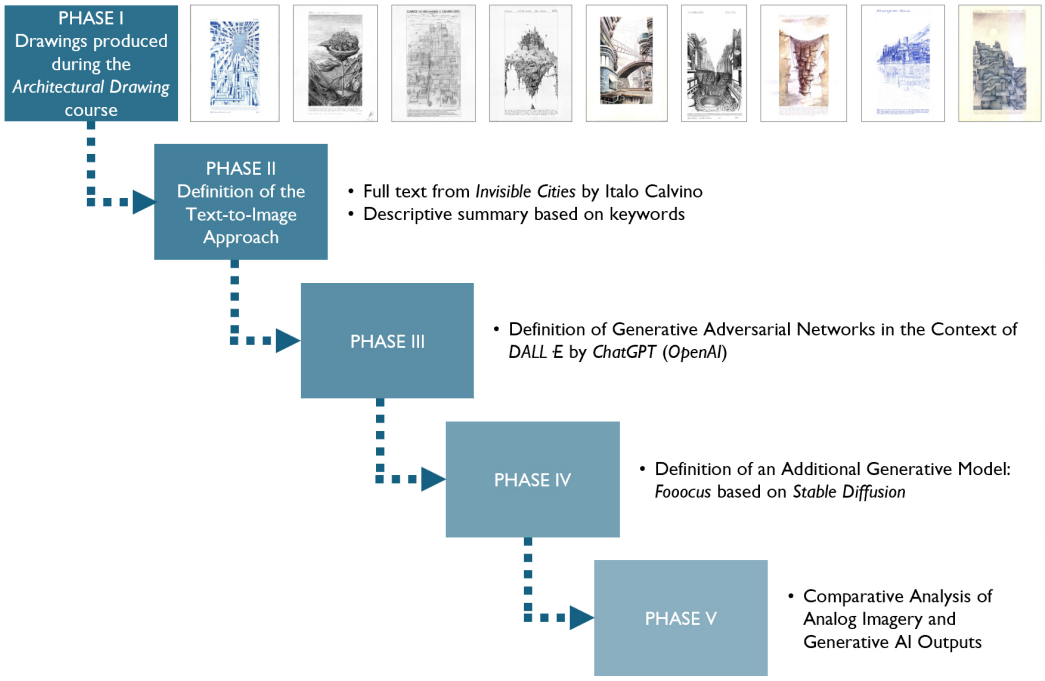


Fig. 1. Diagram of the research methodology (Source: authors' image).

and a set of carefully selected keywords believed to best express the essence of the described city were used. Next, the GAN model for the experiments was determined, with *OpenAI's DALL·E* being selected. However, to obtain additional insights, further experiments were conducted using the digital art tool *Stable Diffusion.*
The fourth phase focused on comparative analysis. The generated images were evaluated based on various criteria, including fidelity to the original text or provided keywords, aesthetic impact in terms of visual quality and composition harmony, and evocative capacity-meaning the level of suggestion and interpretative depth achieved [Mitchell 1994].

Additionally, the degree of semantic reduction occurring in the transition from descriptive language to AI-synthesized visual language was examined [Grau 2002].

Generative AI and language interpolation

The effectiveness of a generative text-to-image artificial intelligence system depends not only on the quality of the generation algorithm but also on how the model interprets and segments the text of the prompt. For this reason, understanding the reading and tokenization process is essential to optimizing prompt writing and achieving results that align with the user's intent.
Tokenization is the process of dividing text into tokens, minimal units of meaning, that can be entire words, subwords, or even individual characters. There are different tokenization strategies:
- character-based, which is more granular and useful for languages without spaces, such as Chinese. However, this comes at the cost of slower processing and greater difficulty in interpreting the semantic meaning of words in context;
- subword-based, used in advanced models (BPE – Byte-Pair Encoding, *WordPiece*), which breaks down complex words into smaller segments to better manage vocabulary, especially when dealing with rare or complex words. This reduces the number of out-of-vocabulary (OOV) words but requires higher computational performance;
- word-based, which separates text based on spaces and punctuation. While simpler to implement than other methods, it struggles with grammatical variations (e.g., eat, eating, eaten would be treated as separate, unrelated tokens);
- *sentencePiece*, which considers multiple words and characters together, allowing for

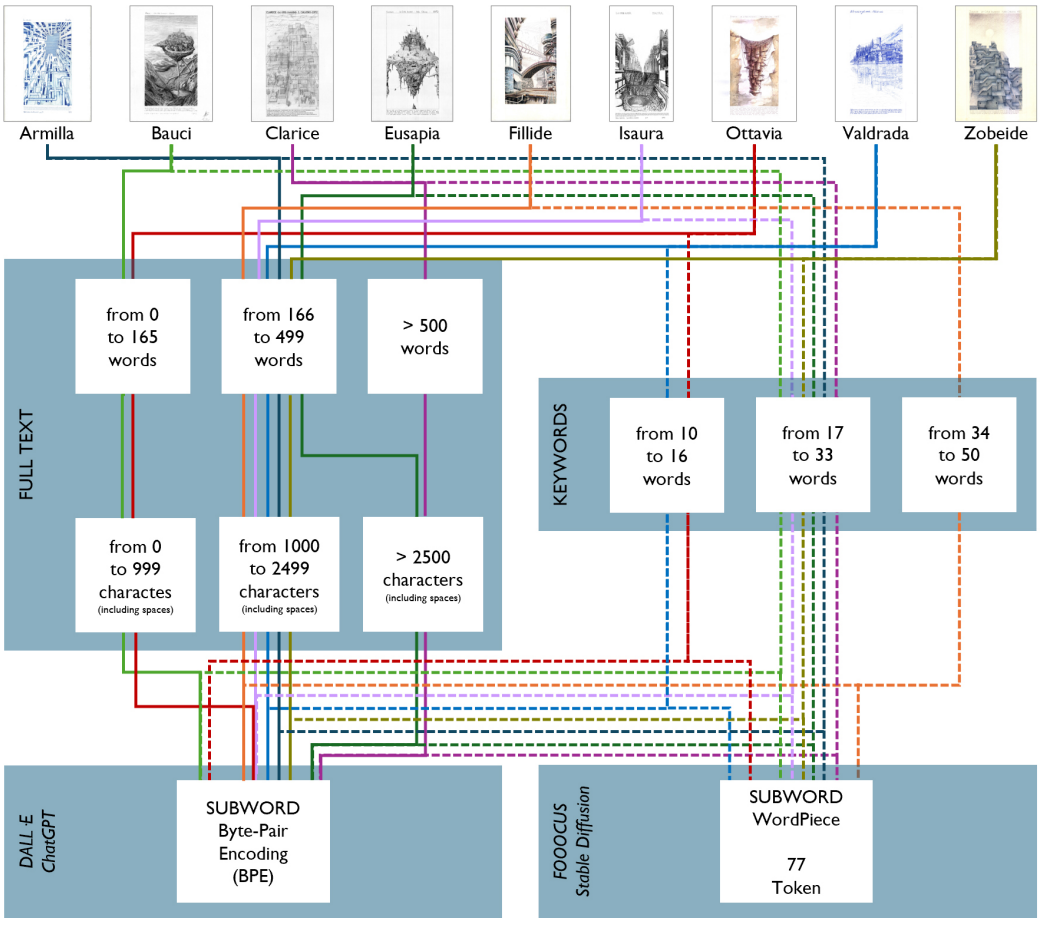


Fig. 2. Diagram of the semantic structure of language in the generative AI investigated in the research (source: authors' image).

better handling of grammatical complexity. This method is particularly useful for multilingual AI models and different alphabets, but it often results in very long tokens [Mielke 2021].
Comparing two well-known generative text-to-image AI models, *DALL·E (ChatGPT)* and *Fooocus* (based on *Stable Diffusion)*, reveals differences in how they analyse and interpret a text prompt [Jamal 2024]. *DALL·E* uses (BPE) *Byte-Pair Encoding,* a subword system optimized for contextual understanding, allowing it to process complex sentences while maintaining semantic coherence [OpenAI 2023].
Fooocus, on the other hand, relies on CLIP [OpenAI 2021], which also uses tokenization but has a strict limit of 77 tokens, reducing its ability to handle long and detailed descriptions (fig. 2). Consequently, while *DALL·E* can interpret detailed prompts with a high degree of accuracy, *Stable Diffusion* tends to prioritize keywords, often disregarding part of the information when the input is too lengthy [Ramesh e*t al.* 2022].
Moreover, the natural language in which the prompt is written is another key factor: *DALL·E* is better at handling languages other than English thanks to extensive multilingual training, whereas *Stable Diffusion* is heavily optimized for English [Zhang 2023]. To achieve the best results, it is crucial to adapt the prompt to the system being used, carefully selecting the linguistic structure and the arrangement of keywords. This may require an additional process of translation and information filtering, which can alter the nuances of meaning in the original prompt. Thus, prompt writing in generative AI is not just a creative endeavour but also a highly technical one, requiring an understanding of how the model reads and processes text to produce results that closely align with the user's intentions

Analysis and research results: from manual techniques to different generative AI platforms

The comparative analysis highlighted significant differences between the two modes of representation: analogue drawing from 'natural intelligence' and digital output from artificial intelligence. The key aspects considered include visual synthesis versus subjective interpretation, the role of cultural background, and semantic fidelity.
Regarding the first point –the comparison between visual synthesis and subjective interpretation– it is essential to emphasize that while 'manual' illustrations tend to emphasize certain details over others based on the author's personal inclinations [Cianci, Calisi 2014], such as experimenting with proportions and using a variety of different perspective methodologies, generative AI, on the other hand, consistently simplifies complex descriptions into standardized iconographic elements based on the images present in its datasets [Manovich 2002]. Despite attempts, AI was unable to fully satisfy textual requests regarding the desired representation method.
This was evident in the tests conducted with the cities of Armilla (fig. 3) and Fillide (fig. 4), which demonstrated a more developed perspective capability in 'natural intelligence'.
The greater stylistic variety demonstrated by the authors of analogue images highlights how personal experiences and cultural background influence artistic expression, whereas AI tends to produce more uniform images, suggesting the presence of a bias stemming from training data [Nochlin 1971; Zylinska 2020]. It can therefore be argued that both natural and artificial intelligence exhibit biases, but these lead to opposite outcomes in terms of graphic rendering and image representation.
This became evident in the experiment conducted on the city of Ottavia (fig. 5), where generative AI's interpretation of Italo Calvino's full text led to the complete omission of the vegetation component (which was also overlooked by the human illustrator) described in the story. Only through a careful selection of keywords was this aspect successfully restored in the generated digital image.
The fundamental difference between the two approaches emerges when comparing the semantic fidelity of their respective results. While 'natural intelligence' is capable of synthesizing an image while retaining the most significant elements of a narrative, this is not the case for AI-generated images. AI tends to prioritize explicit and synthetic descriptive

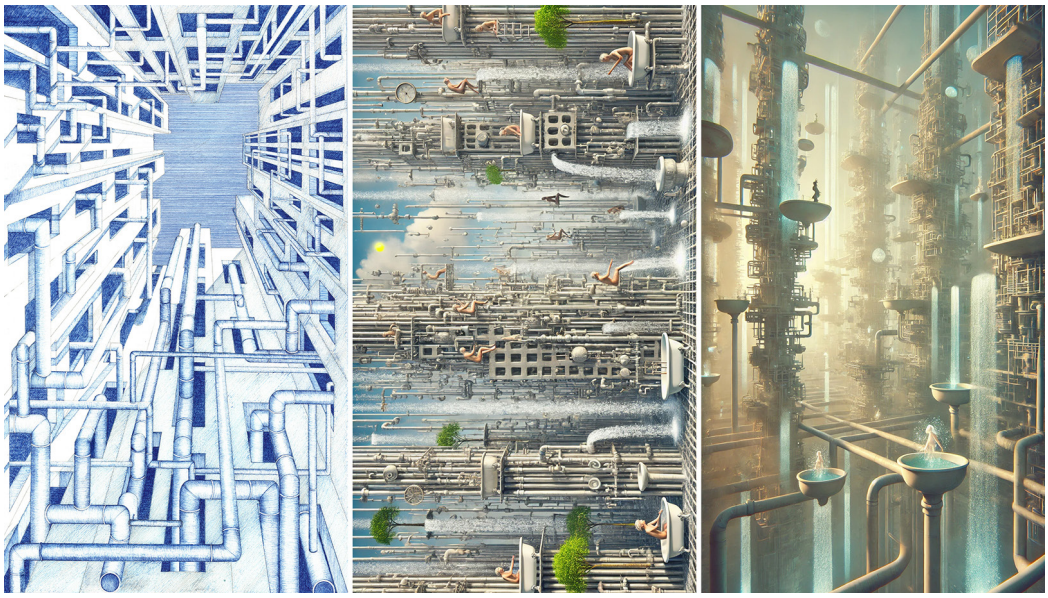


Fig. 3. City of Armilla, pastel on paper (left), image generated using the full text in *DALL·E (ChatGPT)* (center), image generated using keywords in *DALL·E (ChatGPT)* (right) (source: anonymous author of the drawing on the left, images by the authors in the centre and on the right).

Fig. 4. City of Fillide, ink and pastel on paper (left), image generated using the full text in *DALL·E (ChatGPT)* (center), image generated using keywords in *DALL·E (ChatGPT)* (right) (source: drawing by Valerio Pasquali on the left, images by the authors in the centre and on the right).

aspects, struggling to grasp more abstract or poetic concepts. This results in a reduction of Calvino's metaphors to more concrete elements [McCormack *et al.* 2019].
For example, in the case of the city of Isaura (fig. 6), although the AI was able to generate evocative images –at least in this instance– it primarily reproduced the strictly descriptive aspects of Calvino's narrative. However, the research did not appear sufficiently comprehensive when using only *DALL·E* (*ChatGPT*) (fig. 7). Therefore, it was decided to test the ability to generate images from texts using another type of *Generative Adversarial Network: Fooocus* (based on *Stable Diffusion*).
Each AI platform exhibits variations in its approach to data and the outputs it generates, depending on the model architecture and the nature of the dataset used.
However, the comparison with *Fooocus* (based on *Stable Diffusion*), despite emphasizing photorealism more than *DALL·E (ChatGPT*), did not produce satisfactory results in terms of visual synthesis, subjective interpretation, or semantic fidelity. Among the AI-generated

Fig. 5. City of Ottavia, ink and watercolour on paper (left), image generated using the full text in *DALL·E (ChatGPT)* (centre), image generated using keywords in *DALL·E (ChatGPT)* (right) (source: drawing by Giulia Gherardi on the left, images by the authors in the centre and on the right).

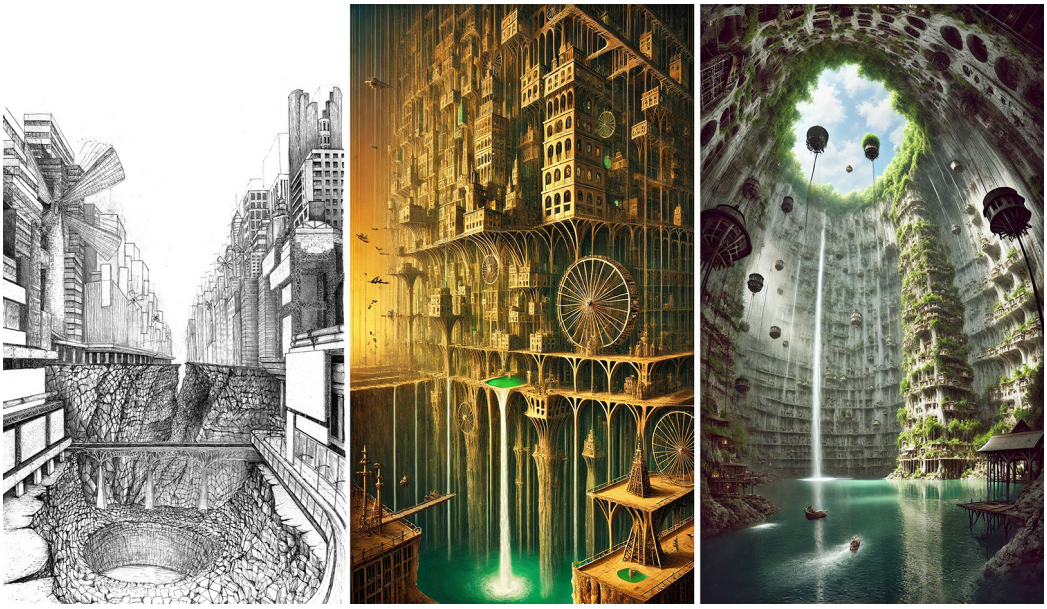


Fig. 6. City of Isaura, ink on paper (left), image generated using the full text in *DALL·E (ChatGPT)* (centre), image generated using keywords in *DALL·E (ChatGPT)* (right) (source: drawing by Beatrice Evangelisti on the left, images by the authors in the centre and on the right).

cities, exemplified by the four images shown in fig. 8, only the first and fourth appear even slightly evocative, while the central ones are basic in their descriptiveness and lack any sense of suggestion.

Conclusions

The research demonstrates that, although AI can produce visually compelling images, its process differs fundamentally from that of humans. Translating literary texts into images presents significant challenges, as artificial intelligence tends to reduce semantic complexity to statistically recurring visual patterns.
The results, however, raise important questions about the future of AI-assisted creativity across various fields, including the relationship between creativity and reproducibility, the evolving role of the artist, and ethical and copyright concerns.

Fig. 7. City of Bauci, ink on paper (left), image generated using the full text in *DALL·E (ChatGPT)* (centre), image generated using keywords in *DALL·E (ChatGPT)* (right) (source: drawing by Marco Piccoli on the left, images by the authors in the centre and on the right).

Fig. 8. Armilla (first from the left), Ottavia (second from the left), Bauci (third from the left), Isaura (fourth from the left), images generated using English keywords in *Fooocus* (based on *Stable Diffusion)* (source: authors' image).

Artificial intelligence offers new creative possibilities but also prompts crucial questions about the nature of creativity itself. Boden suggests that artificial creativity is based on combinations and variations of pre-existing elements rather than genuine innovation [Boden 1998]. The fundamental question, then, is whether a machine can ever be truly creative or if it will always remain a tool that supports human imagination.

The role of the artist may also be redefined with the increasing use of AI generation tools. In a context where anyone can create images with just a few text commands, it raises the question of the value of traditional art and human intervention in the creative process [Zylinska 2020].

Finally, the ethical debate remains unresolved. The use of AI in image generation raises concerns regarding intellectual property and copyright. AI-generated images are derived from datasets of existing works, leading to doubts about the legitimacy of using such references without the original authors' consent [Floridi 2014].

In conclusion, the analysis highlights the irreplaceable value of human interpretation in the creative act, emphasizing the risk of aesthetic homogenization due to reliance on pre-trained datasets. However, the combination of AI and human artistic intervention could open new perspectives for art and design. Future research could further explore the ongoing question of how to integrate these tools into the artistic landscape and how to regulate AI use to ensure ethically sustainable visual production.

In an era where the boundaries of art are constantly being redefined, the dialogue between artificial intelligence and human creativity stands as one of the most stimulating challenges for the worlds of culture and technology.

Note

[1] The quotation was taken from the interview titled *Italo Calvino - Un uomo invisibile,* conducted by Valerio Riva with Italo Calvino on December 8, 1974, for the series *Incontri - Fatti e personaggi del nostro tempo,* produced by RSI Radiotelevisione Svizzera and directed by Nereo Rapetti.

Reference List

Benjamin, W. (2014). *L'opera d'arte nell'epoca della sua riproducibilità tecnica.* Torino: Einaudi.

Boden, M. A. (1998). Creativity and artificial intelligence. In *Artificial intelligence,* vol. 103, n. 1-2, pp. 347-356. https://doi.org/10.1016/S0004-3702(98)00055-1.

Calvino, I. (1972). *Le città invisibili.* Torino: Einaudi.

Cianci, M. G., Calisi, D. (2014). Il mondo è un libro: visioni ispirate da "Le città invisibili" di Italo Calvino. In A. Garcia Melian (Ed.). *El Dibujo de Viaje de los Arquitectos*. Atti del XV Congreso Internacional Expresion Grafica Arquitectonica EGA2014. Las Palmas de Gran Canaria, 22-23 maggio 2014, pp. 751-759. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality.* Oxford: Oxford University Press.

Grau, O. (2002). *Virtual Art: From Illusion to Immersion.* Cambridge: MIT Press.

Jamal, S., Wimmer, H., Rebman, C. M. Jr. (2024). Perception and evaluation of text-to-image generative AI models: A comparative study of DALL-E, Google Imagen, GROK, and Stable Diffusion. In *Information Systems*, 25(2), pp. 277-292. https://doi.org/10.48009/2_iis_2024_123.

Manovich, L. (2002). *The Language of New Media.* Cambridge: MIT Press.

Manovich, L. (2018). *AI Aesthetics.* Moscow: Strelka Press.

McCormack, J., Gifford, T., Hutchings, P. (2019). Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. *Leonardo,* 52(3), pp. 285-291. https://doi.org/10.48550/arXiv.1903.02166.

Mielke, S. J., Alyafeai, Z., Salesky, E., Raffel, C., Dey, M., Gallé, M., Raja, A., Si, C., Lee, W.Y., Sagot, B., Tan, S. (2021). Between words and characters: A brief history of open-vocabulary modeling and tokenization in NLP. In *arXiv*. Cornell University. https://doi.org/10.48550/arXiv.2112.10508.

Mitchell, W. J. (1994). *The Reconfigured Eye: Visual Truth in the Post-Photographic Era.* Cambridge: MIT Press.

Nochlin, L. (1971). *Why Have There Been No Great Women Artists?.* New York: ARTnews.

OpenAI. (2021). *CLIP: Connecting text and images. OpenAI.* https://openai.com/index/clip/.

OpenAI. (2023). *Introducing ChatGPT and DALL·E. OpenAI.* https://openai.com/dall-e.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 10684-10695. https://doi.org/10.48550/arXiv.2204.06125.

Riva, V. Rapetti, F. (8 dicembre 1974). Intervista a Italo Calvino dal titolo: *Italo Calvino - Un uomo invisibile*. https://www.rsi.ch/archivi/Italo-Calvino-Un-uomo-invisibile--1643011.html.

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. (2023). Text-to-image Diffusion Models in Generative AI: A Survey. In *arXiv*.. https://doi.org/10.48550/arXiv.2303.07909.

Zylinska, J. (2020). *AI Art: Machine Visions and Warped Dreams.* Londra: Open Humanities Press.

Authors
Maria Grazia Cianci, Roma Tre University, mariagrazia.cianci@uniroma3.it
Daniele Calisi, Roma Tre University, daniele.calisi@uniroma3.it
Stefano Botta, Roma Tre University, stefano.botta@uniroma3.it
Sara Colaceci, Roma Tre University, sara.colaceci@uniroma3.it
Michela Schiaroli, Roma Tre University, michela.schiaroli@uniroma3.it

To cite this chapter: Maria Grazia Cianci, Daniele Calisi, Stefano Botta, Sara Colaceci, Michela Schiaroli (2025). Èkphrasis and Generative AI: Analog/Digital Reflections in the Imaginary of Calvino's *Invisible Cities*. 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. 3645-3664. DOI: 10.3280/oa-1430-c944.

ISBN 9788835182412