èkphrasis

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

Parole che modellano spazi: IA generativa e rappresentazioni 3D architettoniche

Romina Nespeca
Renato Angeloni
Laura Coppetta

Abstract

L'Intelligenza Artificiale generativa ha rivoluzionato la relazione tra parola e immagine, proponendosi come strumento in grado di generare rappresentazioni grafiche a partire da descrizioni testuali. Il presente studio indaga le potenzialità di questo moderno processo ecfrastico ai fini della generazione di ricostruzioni tridimensionali in ambito architettonico, focalizzandosi su tre casi di studio: gli ordini, come definiti da Vitruvio nel *De Architectura*, alcuni dipinti rinascimentali raffiguranti spazi costruiti e, infine, un edificio monumentale esistente. Ricorrendo a diversi strumenti per la generazione di testi, immagini e modelli 3D si sono evidenziate le potenzialità, ma anche le numerose criticità relative alla fedeltà semantica e geometrica delle rappresentazioni generate. In un contesto in cui la relazione tra parola e immagine assume nuove forme, l'IA emerge dunque come una tecnologia in grado di arricchire e potenzialmente rivoluzionare i modi della rappresentazione grafica, aprendo a nuove prospettive metodologiche e culturali nel settore disciplinare del Disegno.

Parole chiave
Text-to-image-to-3D, image-to-3D, text-to-3D, IA generativa, *èkphrasis*.

Utilizzo di *prompt* testuale *DALL-E 3*: Parole che modellano Spazi: IA generativa e rappresentazioni architettoniche (elaborazione degli autori).

doi.org/10.3280/oa-1430-c916

Introduzione

L'Intelligenza Artificiale (IA) si sta imponendo come strumento in grado di generare rappresentazioni visive a partire da descrizioni testuali [Arzomand *et al*. 2024], trovando un'interessante affinità con l'antica pratica retorica dell'*èkphrasis*. Questa, definita da Ermogene di Tarso nel II secolo d.C., consiste in un discorso descrittivo capace di evocare immagini vivide attraverso le parole, enfatizzando dettagli che arricchiscono la comprensione del soggetto rappresentato. Intrisa di *enárgeia*, la capacità di generare immagini potenti nella mente del fruitore, questa pratica trova oggi una nuova declinazione nei metodi di generazione di rappresentazioni tramite sistemi di IA come le reti neurali profonde, le *Generative Adversarial Networks* (GAN) e i modelli di diffusione [Wu *et al*. 2023], quali i modelli GPT (*Generative Pretrained Transformer*) e BERT (*Bidirectional Encoder Representations from Transformers*), capaci di interpretare testi scritti [Gozalo-Brizuela, Garrido-Merchan 2023]. Grazie a queste tecnologie, strumenti come *MidJourney, DALL-E* e *Stable Diffusion* permettono di generare immagini a partire da descrizioni testuali (*text-to-image*), inaugurando una nuova modalità di rappresentazione che pone la disciplina del Disegno al crocevia tra parola e immagine [Zhang *et al*. 2024]. Come presentato da studi recenti, le immagini generate tramite tali strumenti si caratterizzano per complessità e dettaglio, ne sono un esempio le rappresentazioni di spazi urbani in grado di offrire nuove prospettive per l'identità urbana e i modelli di design [De Natale 2024]. Attraverso l'uso di tecnologie come il *Natural Language Processing* (NLP), l'IA non solo crea, ma agisce anche come mediatore tra opera e pubblico, generando descrizioni dettagliate che combinano dati tecnici e interpretazioni estetiche [D'Uva 2024]. Questo processo crea un circolo virtuoso in cui l'IA interpreta e arricchisce il contenuto, rendendo fondamentale l'analisi linguistica e l'addestramento delle reti per affrontare le complessità semantiche dei testi e produrre immagini realistiche [Pellicio *et al*. 2023]. Questa evoluzione richiama le descrizioni architettoniche medievali, dove il testo guidava la costruzione fisica [Rasetti 2023]. La convergenza tra l'antica arte della descrizione verbale e i contenuti generati tramite IA solleva però nuovi interrogativi, in merito alle implicazioni metodologiche e culturali di questa collaborazione tra linguaggio umano e capacità generativa dell'IA. Questioni come l'autorialità e la creatività emergono con forza, evidenziando la tensione tra il controllo umano e l'automazione algoritmica. Sebbene il confine tra creazione umana e artificiale si stia assottigliando, la generazione di immagini e contenuti visivi da parte dell'IA rimane strettamente legata alle descrizioni fornite dall'utente [Battini 2023]. In questo contesto, il presente articolo intende esplorare l'utilizzo dell'IA nella modellazione 3D generata da input testuali e immagini, un ambito ancora poco approfondito. Attraverso un approccio interdisciplinare, che collega le radici teoriche dell'*èkphrasis* alle applicazioni più recenti nella rappresentazione digitale [Battini, Martínez Chao 2024; Trizio *et al*. 2024], questa ricerca indaga come l'accuratezza della descrizione verbale possa diventare il fulcro di una nuova *poiesis* condivisa tra uomo e macchina [Condorelli *et al.* 2023; Loddo *et al*. 2024]. L'analisi proposta, radicata nei fondamenti della disciplina del Disegno, invita a riflettere su un panorama in cui il rapporto tra parola e immagine assume nuove e suggestive declinazioni, contribuendo alla creazione di rappresentazioni capaci di evocare e innovare.
È possibile, dunque, che l'IA acquisisca quella forza espressiva e comunicativa che Ermogene attribuiva all'*èkphrasis*, rispondendo così alle sfide di una realtà in continua evoluzione?

Metodologia e strumenti

Grazie alle potenzialità dell'IA, l'*èkphrasis* - intesa come descrizione verbale capace di evocare la rappresentazione visiva di un oggetto o di un'idea - assume una nuova dimensione. L'interazione tra linguaggio, testo e immagine si arricchisce, permettendo la creazione di rappresentazioni tridimensionali attraverso l'elaborazione mediata dall'IA. In particolare, questo studio esplora come processi quali il *text-to-3D* e l'*image-to-3D* possano essere applicati nell'ambito della rappresentazione architettonica focalizzandosi su casi studio appartenenti a tre tipologie emblematiche:

– gli ordini architettonici: dorico, ionico e corinzio, come descritti da Vitruvio;
– alcuni dipinti rinascimentali di architetture: *La Città Ideale di Urbino*, *Lo Sposalizio della Vergine* e *La Scuola di Atene* di Raffaello;
– un edificio monumentale esistente: la Chiesa di Santa Maria della Piazza ad Ancona.

Tali casi di studio, con le loro diverse peculiarità, costituiscono un terreno fertile per indagare le potenzialità e i limiti dell'IA nel tradurre descrizioni testuali o visive in rappresentazioni tridimensionali. In particolare, la metodologia adottata si basa sull'utilizzo di *ChatGPT* per l'elaborazione di descrizioni testuali, *DALL-E* per la generazione di immagini bidimensionali sintetiche e di differenti piattaforme per la generazione di modelli tridimensionali.

Per meglio valutare e confrontare l'efficacia della generazione di modelli 3D, sono state prese in esame quattro differenti piattaforme che offrono sia la modalità *text-to-3D* che *image-to-3D*: *Rodin* di *Hyper3D*, *3D AI Studio*, *Meshy AI* e *Tripo AI*. Ognuna di queste utilizza approcci diversi per l'elaborazione dei dati di input e per la creazione delle geometrie tridimensionali. Con questi strumenti sono stati, quindi, definiti tre flussi di lavoro, ognuno dei quali ha portato alla creazione di una rappresentazione 3D seguendo un procedimento distinto.

Nel primo flusso (*text-to-image to* 3D) si è partiti da descrizioni testuali dettagliate, redatte da Vitruvio nel primo caso studio e da Wikipedia nel secondo. Tali descrizioni sono state quindi utilizzate da ChatGPT per generare immagini sintetiche attraverso DALL-E. Le immagini così ottenute sono state successivamente caricate sui quattro tools, che hanno estratto da queste le informazioni necessarie per creare modelli tridimensionali.

Nel secondo approccio (*image-to-*3D), si è partiti da immagini esistenti: le rappresentazioni degli ordini vitruviani realizzati da Daniele Barbaro, e le opere pittoriche individuate come casi di studio: *La Città Ideale di Urbino*, *Lo Sposalizio della Vergine* e *La Scuola di Atene di Raffaello*. Come nel precedente flusso di lavoro, tali immagini sono state poi elaborati tramite i tools per la generazione dei rispettivi modelli 3D.

Infine, nel terzo flusso (*text-to-*3D) sono stati generati tramite *ChatGPT* prompt testuali utilizzati per creare modelli tridimensionali.

Per il terzo caso studio, relativo a un edificio esistente, si è cercato di spingersi ancora oltre. In una prima fase, sono state elaborate singole immagini fotografiche e disegni CAD, ottenendone modelli tridimensionali. Successivamente, si è testata l'applicazione del metodo con immagini multiple. Un primo tentativo ha visto il caricamento di due sezioni, della pianta e del prospetto principale, nel secondo si sono invece utilizzate diverse viste fotografiche della chiesa. Tali differenti approcci hanno permesso di esplorare le potenzialità del software nella ricostruzione tridimensionale a partire da input eterogenei, come disegni ortografici e immagini prospettiche.

Gli strumenti utilizzati svolgono un ruolo cruciale nell'intero processo. *ChatGPT*, grazie alla sua capacità di comprendere e generare linguaggio naturale, è fondamentale per la traduzione di descrizioni complesse in prompt testuali adatti alla generazione di immagini. I *tools* per la generazione di modelli 3D, d'altra parte, rappresentano un avanzamento tecnologico significativo, permettendo la conversione di prompt testuali e di immagini bidimensionali in rappresentazioni tridimensionali. Tuttavia, nonostante le potenzialità di questi strumenti, emergono ancora sfide significative da superare, rappresentando quello della generazione di modelli 3D tramite IA un ambito ancora poco esplorato e in fase di sviluppo.

Discussione e risultati

I risultati delle sperimentazioni hanno evidenziato diverse criticità e potenzialità nei flussi di lavoro testati. Nel primo caso studio, relativo al *De Architectura* di Vitruvio, sono emerse difficoltà significative nella generazione del capitello dorico: nonostante le descrizioni testuali dettagliate presenti nel trattato, *ChatGPT* tende a generare immagini di capitelli ionici invece che dorici, influenzando di conseguenza la creazione dei modelli 3D, che risultano coerenti con l'immagine generata ma non con la descrizione originale. Al contrario, i capitelli ionici e corinzi non hanno presentato particolari problemi di generazione (fig. 1).

L'analisi comparativa dei diversi strumenti di generazione di modelli 3D ha evidenziato risultati contrastanti nell'elaborazione dei disegni di Daniele Matteo Alvise Barbaro. *Rodin*

Fig. 1. *Text-to-image-to-3D* del primo caso studio: gli ordini architettonici.

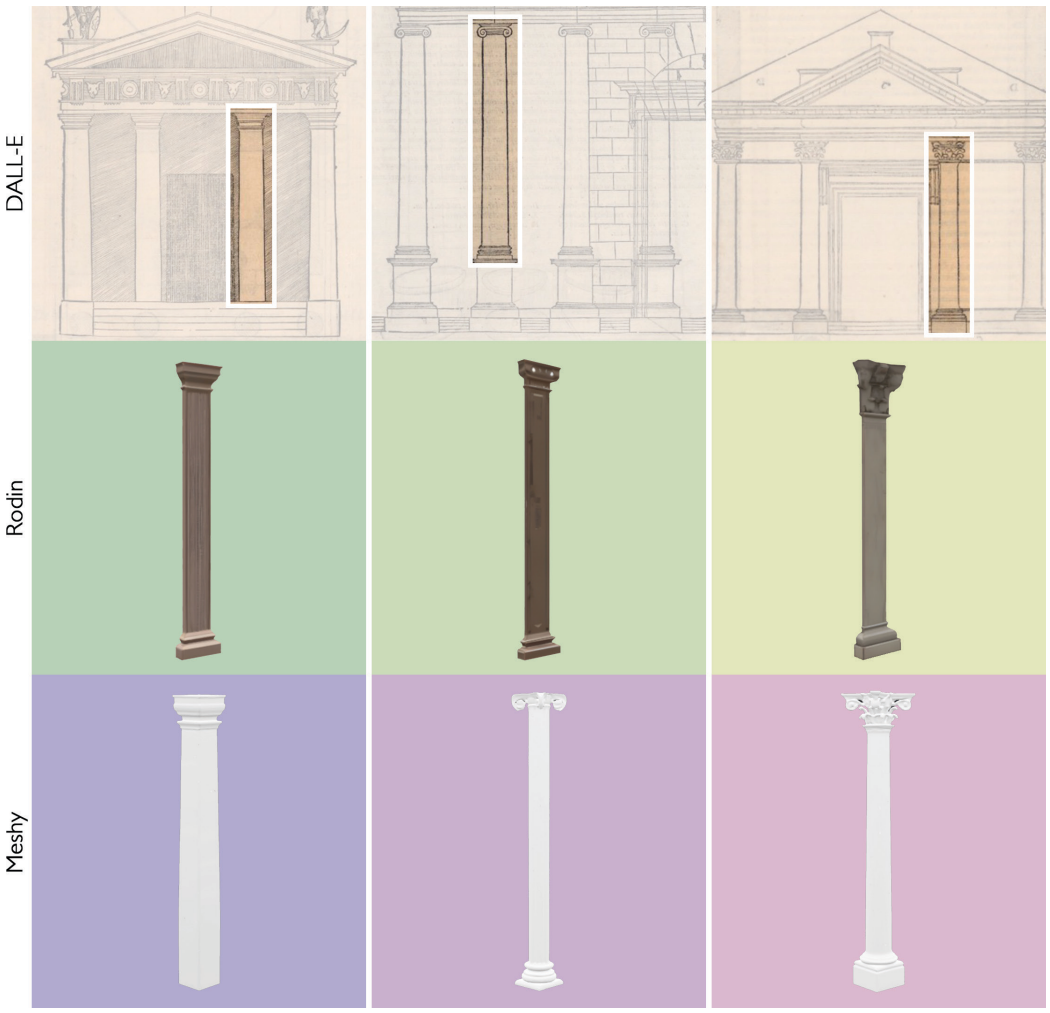

Fig. 2. *Image-to -3*D del primo caso studio: gli ordini architettonici.

e *Meshy* sono riusciti a generare un modello tridimensionale a partire da rappresentazioni bidimensionali, sebbene con alcune criticità. In particolare, *Rodin* ha erroneamente classificato le colonne come paraste, mentre *Meshy*, per l'ordine dorico, ha prodotto un pilastro con una sezione quadrata, discostandosi così dalla forma originale. Al contrario, *3D AI Studio* e *Tripo AI* non sono stati in grado di generare un modello 3D partendo dai disegni del Barbaro (fig. 2).

Quando si è utilizzata la descrizione testuale diretta per generare i modelli 3D, i risultati variano notevolmente a seconda dello strumento utilizzato. *Rodin* ha prodotto colonne basse e tozze, con capitelli dorici più lontani dalla descrizione vitruviana rispetto a quelli ionici e corinzi. Tripo è risultato lo strumento più accurato nell'interpretazione della richiesta, anche se non in grado di riprodurre propriamente una colonna dorica. Invece, *Meshy* e *3D AI Studio* tendono a generare esclusivamente colonne ioniche, mostrando una preferenza sistematica per questo ordine architettonico, indipendentemente dalle specifiche testuali fornite (fig. 3).

Nel secondo caso studio, basato sui dipinti di architetture, i modelli 3D generati da *Rodin* direttamente dai dipinti originali si sono rivelati insoddisfacenti, con errori significativi nella comprensione, sia delle architetture, sia delle figure, come dimostrato in particolare nel caso della *Scuola di Atene*. Al contrario, *3D AI Studio* è riuscito a capire che la *Scuola di Atene* rappresenta un ambiente interno, ricostruendo approssimativamente la volumetria della stanza (fig. 4). In entrambe le piattaforme, si sono ottenuti dei risultati migliori isolando manualmente gli elementi architettonici, come nel caso del tempio della *Città Ideale* e de *Lo Sposalizio della Vergine* (fig. 5).

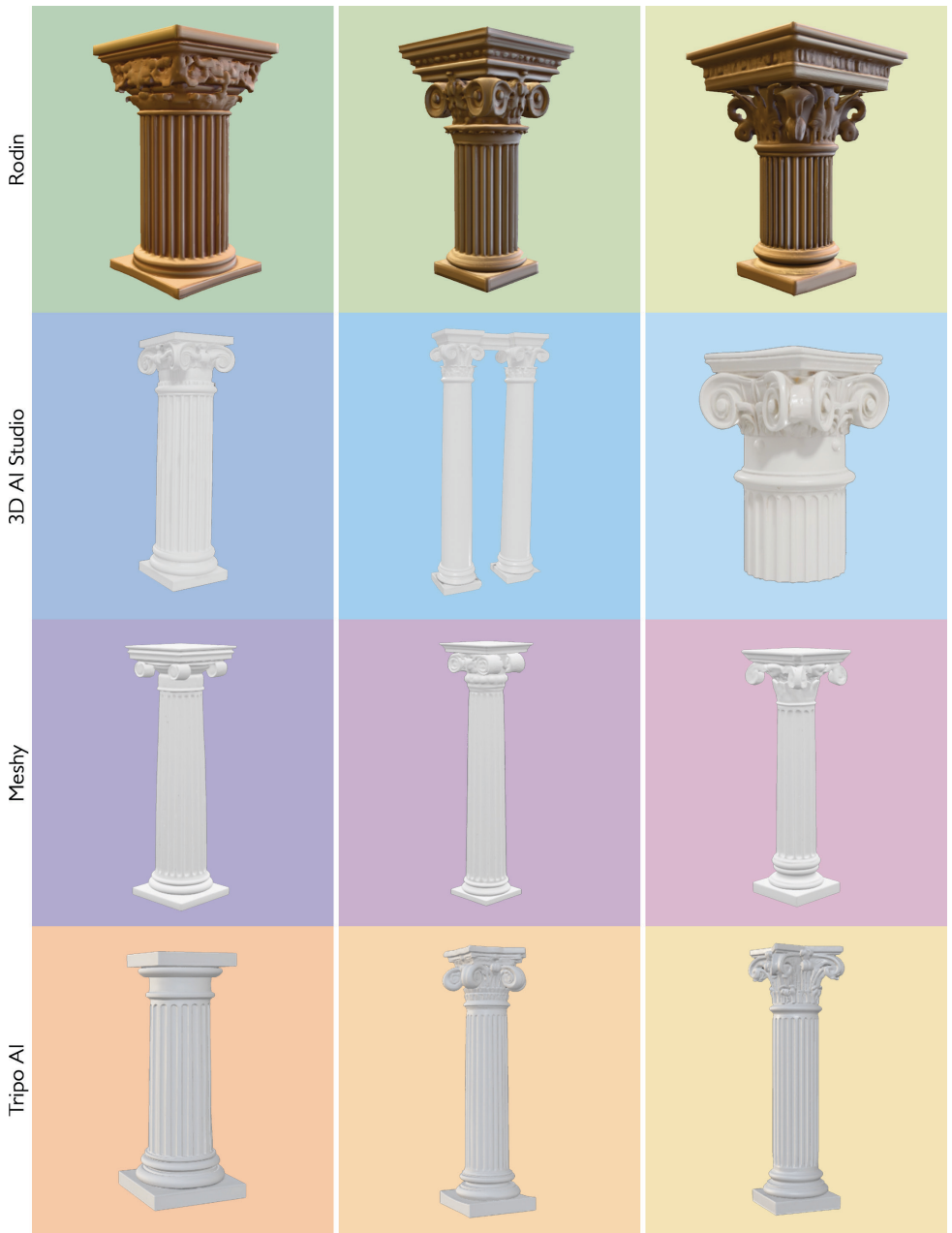

Fig. 3. *Text-to*-3D del primo caso studio: gli ordini architettonici.

Inoltre, utilizzando immagini generate da *ChatGPT* a partire da descrizioni testuali di *Wikipedia*, i modelli 3D prodotti da *Rodin* hanno mostrato una maggiore aderenza agli elementi originali nei casi del tempio del *La Città Ideale* e de *Lo Sposalizio della Vergine*, mentre *La Scuola di Atene* ha continuato a generare risultati insoddisfacenti. Invece, *3D AI Studio* riconosce anche in questo caso che la ricostruzione de *La Scuola di Atene* rappresenta uno spazio interno, mentre ne *La Città Ideale* riesce a ricostruire il tempio, ma introduce elementi che non sono presenti nell'input bidimensionale (fig. 6).
Infine, la generazione diretta di modelli 3D a partire dalle descrizioni testuali dei dipinti non ha prodotto modelli fedeli agli originali, con nessuna delle due piattaforme (fig. 7).
Nel terzo caso studio, relativo alla generazione di modelli 3D di Santa Maria della Piazza di Ancona, i risultati più fedeli sono stati ottenuti caricando immagini fotografiche scattate da diverse angolazioni, che hanno permesso a *Rodin* di ricostruire con maggiore accuratezza l'architettura esistente (figg. 8, 9).

Fig. 4. *Image-to-*3D del secondo caso studio: dipinti di architetture. Immagini senza segmentazione.

Partendo da singoli disegni CAD (pianta, sezione, prospetto), i risultati sono stati disomogenei: il prospetto si avvicina all'originale, mentre la pianta e la sezione richiedono un'interpretazione significativa da parte dell'utente (fig. 10). Utilizzando più rappresentazioni in proiezione ortogonale, il modello risultante è apparso incoerente rispetto ai disegni stessi (fig. 11).
I risultati del lavoro evidenziano aspetti positivi e criticità, mettendo in luce il ruolo dell'intervento umano nella metodologia, come nella segmentazione dell'immagine per favorire il riconoscimento, quindi la ricostruzione tridimensionale di alcuni elementi. L'uso di strumenti di AI, come *ChatGPT* e piattaforme per la generazione di modelli 3D, rappresenta un approccio innovativo nel *Digital Heritage*. Tuttavia, la natura aleatoria dei processi generativi rende difficile prevedere i risultati, soprattutto in casi complessi come quelli basati su descrizioni testuali o dipinti.
Sebbene tecniche come l'uso di immagini ritagliate o fotografie multiple abbiano migliorato la precisione, permangono limiti significativi, soprattutto nei dettagli architettonici. La necessità di tecniche di *fine-tuning* e modelli addestrati su contesti specifici emerge come una possibile soluzione, ma richiede risorse notevoli e una stretta collaborazione interdisciplinare.

Conclusioni

Il presente contributo ha dimostrato come dall'intreccio tra l'antica pratica dell'*èkphrasis* e le potenzialità dell'Intelligenza Artificiale Generativa possa scaturire un nuovo rapporto tra parola e immagine, essere umano e macchina. Le sperimentazioni condotte evidenziano come l'IA sia in grado di tradurre descrizioni verbali in immagini e modelli tridimensionali, tuttavia, questa collaborazione non è esente da sfide. I risultati delle sperimentazioni, seppur promettenti, mettono in evidenza alcuni limiti. Nei casi studio analizzati, il processo di *text-to-*3D ha evidenziato criticità nella fedeltà semantica e geometrica rispetto agli *input* testuali.

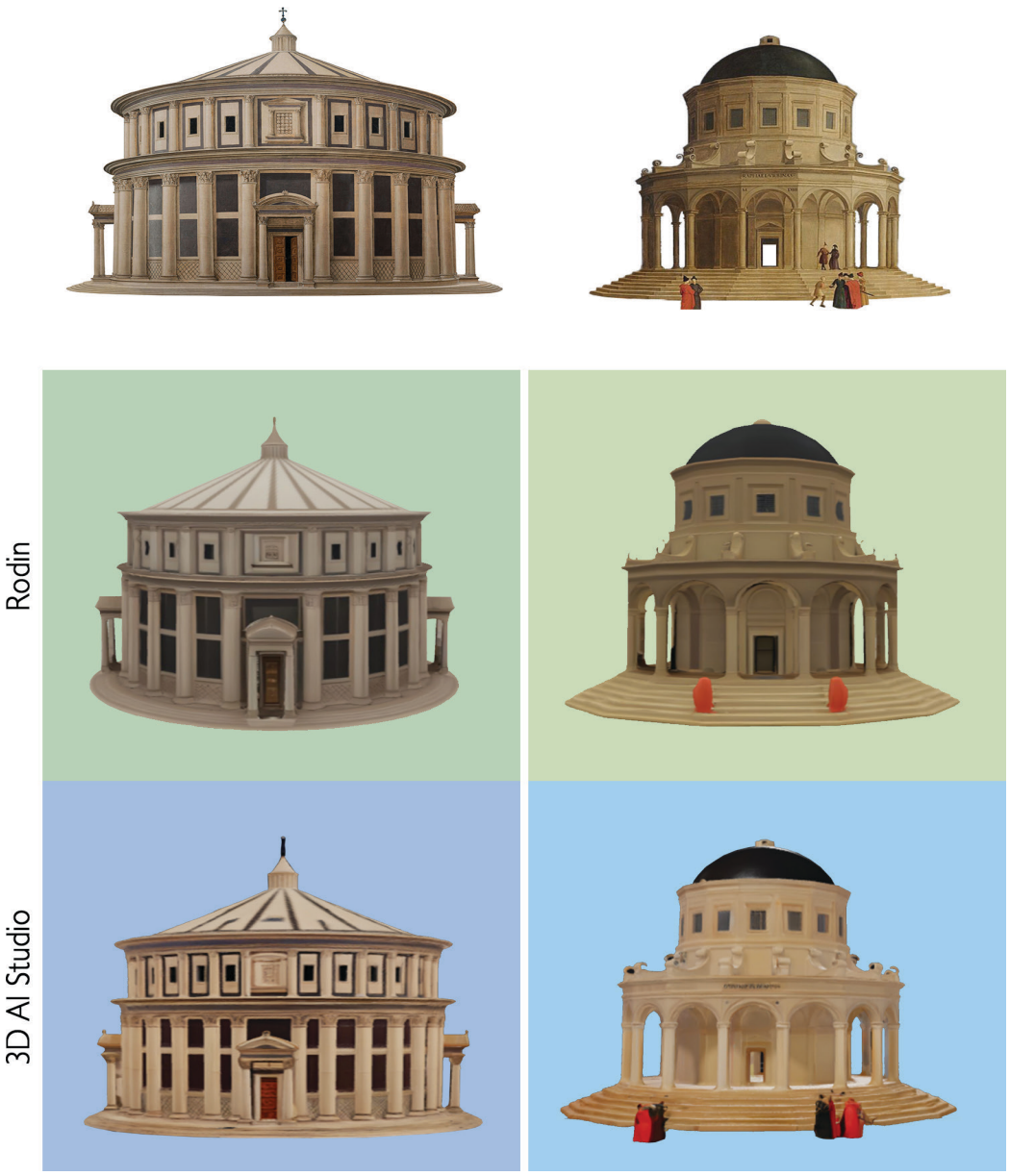

Fig. 5. *Image-to-*3D del secondo caso studio: dipinti di architetture. Immagini segmentate.

Ad esempio, l'incapacità degli strumenti utilizzati di rappresentare un capitello dorico o di ricostruire tridimensionalmente architetture dipinte, suggerisce che il potenziale di tali sistemi sia ancora in evoluzione. Nonostante tali criticità, l'IA generativa rappresenta un'opportunità rivoluzionaria per migliorare i processi analitici e creativi.

In conclusione, questo studio invita a considerare l'IA non come un sostituto, ma come un co-creatore, capace di potenziare l'immaginazione umana e di ampliare i confini della rappresentazione architettonica e artistica. La fase di co-creazione emerge chiaramente nella metodologia adottata. Ad esempio, per ottenere risultati accurati, è stato necessario segmentare manualmente le immagini dei dipinti, consentendo all'IA di ricostruire correttamente gli oggetti rappresentati.

Allo stesso modo, la generazione di un capitello dorico corretto, non raggiunta tramite la descrizione vitruviana, è possibile fornendo un riferimento esplicito ad un edificio esistente come il Partenone ("Genera un capitello dorico come quello del Partenone"), dimostrando l'importanza di indicazioni contestuali e precise per orientare l'*output* generativo.

Fig. 6. *Text-to-Image-to-*3D del secondo caso studio: dipinti di architetture.

Fig. 7. *Text-to-3D* del secondo caso studio: dipinti di architetture.

Questa osservazione suggerisce che per migliorare l'affidabilità dei modelli generativi sia necessaria una maggiore specializzazione, sia per interpretare riferimenti a oggetti esistenti, sia per affrontare la rappresentazione di immagini ortografiche, un ambito in cui l'IA, addestrata principalmente su dataset fotografici, presenta ancora delle lacune. In questo contesto, una prospettiva futura potrebbe riguardare l'addestramento di reti neurali su dataset contenenti proiezioni ortogonali, aumentando così la capacità dell'IA di gestire contesti grafici più complessi e specifici.
Ulteriori ricerche potrebbero concentrarsi sulla capacità di integrare conoscenze interdisciplinari, affinando la capacità dell'IA di generare modelli 3D più fedeli, contestualizzati e utilizzabili in ambiti più applicativi, come il HBIM e altri sistemi informativi. In questo scenario, l'intervento umano rimarrà essenziale per guidare, correggere e ottimizzare il processo, sfruttando l'IA come strumento di co-creazione piuttosto che un semplice automatismo per la rappresentazione e l'analisi dei beni culturali.

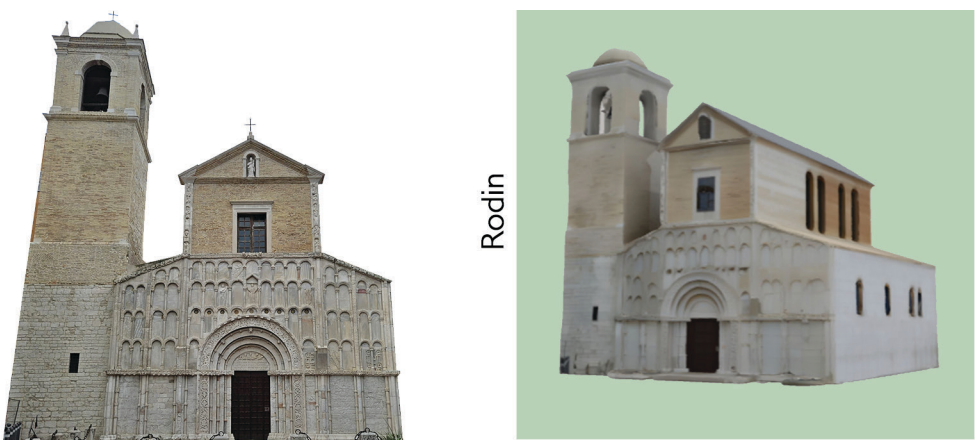

Fig. 8. *Image-to-3D* del terzo caso studio: opera architettonica. Fotografia singola.

Fig. 9. *Image-to-3D* del terzo caso studio: opera architettonica. Fotografie multiple.

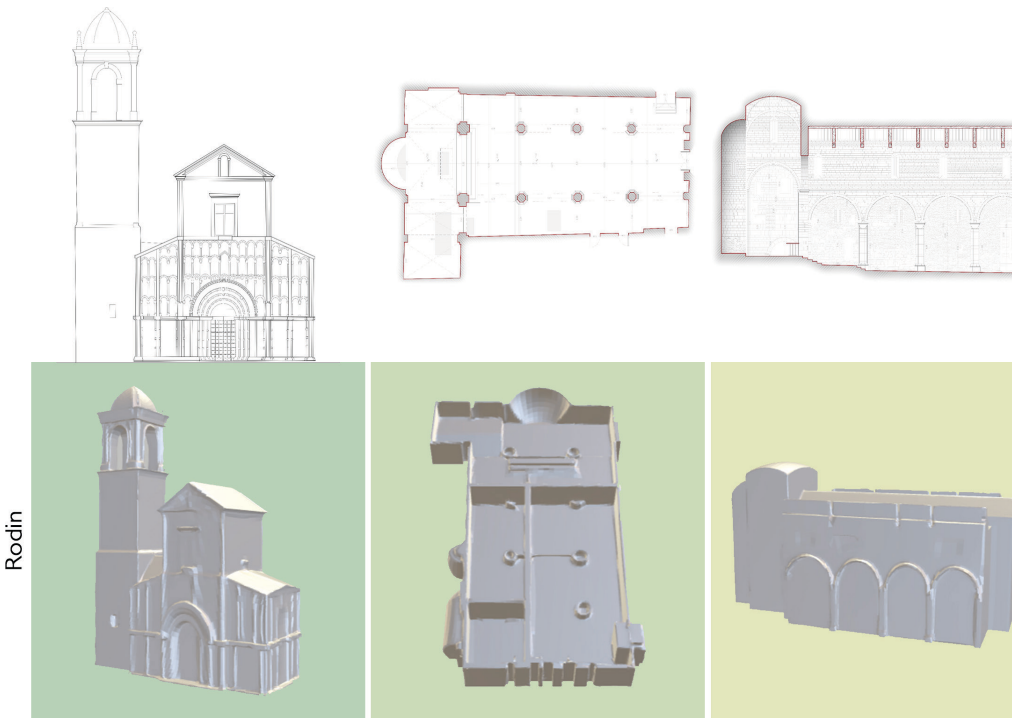

Fig. 10. *Image-to-3D* del terzo caso studio: opera architettonica. Disegno ortografico.

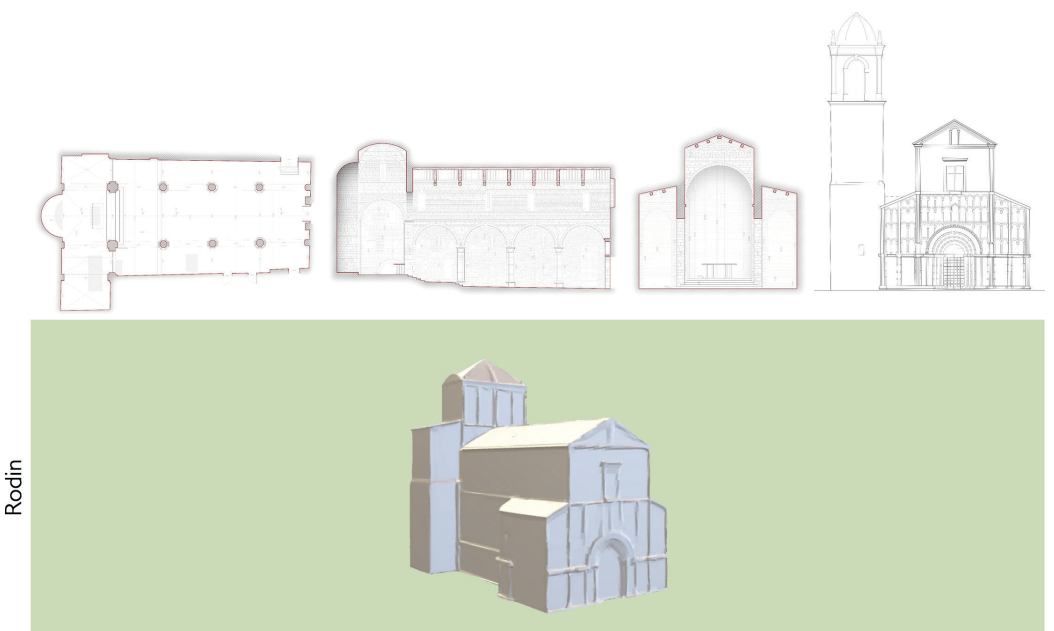

Fig. 11. *Image-to-3D* del terzo caso studio: opera architettonica. Disegni ortografici multipli.

Ringraziamenti e crediti

A valere sul Piano Nazionale di Ripresa e Resilienza, Missione 1 "Digitalizzazione, innovazione, competitività, cultura e turismo", Componente 3 "Turismo e cultura 4.0", MISURA 3 "Industrie culturali e creative", INVESTIMENTO 3.3 "Capacity building per gli operatori della cultura per gestire la transizione digitale e verde", Sub-investimento 3.3.1 "Interventi per migliorare l'ecosistema in cui operano i settori culturali e creativi, incoraggiando la cooperazione tra operatori culturali e organizzazioni e facilitando upskill e reskill", finanziato dall'Unione Europea – NextGenerationEU – Avviso di cui al Decreto Direttoriale del 9 giugno 2023, n. 149. Progetto SkillsHub "Servizi per Culture&Creativity" Codice progetto PNRRI-20230003160497/2, Codice CUP C31B23000430004, ammesso a finanziamento con Decreto Direttoriale n. 737 del 7 dicembre 2023.

Riferimenti bibliografici

Arzomand, K., Rustell, M., Kalganova, T. (2024). From ruins to reconstruction: Harnessing text-to-image AI for restoring historical architectures. *Challenge Journal of Structural Mechanics,* 10(2), 69-85. https://doi.org/10.20528/cjsmec.2024.02.004.

Battini, C. (2023). Intelligenza artificiale tra scienza e creatività. Casi studio nelle arti visive. In M. Cannella, A. Garozzo, S. Morena (A cura di). *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 2380-2393. https://doi.org/10.3280/oa-1016-c411.

Battini, C., Martínez Chao, T. E. (2024). Progettazione e IA. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (A cura di). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 61-76. https://doi.org/10.3280/oa-1180-c472.

Condorelli, F., Luigini, A., Nicastro, G., Tramelli, B. (2023). Disegno e intelligenza artificiale. Enunciati teorici e prassi sperimentale per una poiesi condivisa. In M. Cannella, A. Garozzo, S. Morena (A cura di), *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 2623-2640. https://doi.org/10.3280/oa-1016-c426.

De Natale, I. (2024). La misura dell'identità urbana con l'IA generativa. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (A cura di). *Misura / Dismisura.* Atti del 45° Convegno Inter- nazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 2769-2780. https://doi.org/10.3280/oa-1180-c611.

D'Uva, D. (2024). AI-Enhanced Facade Design: Exploring the Synergy of Generative Models and Architectural Creativity. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (A cura di). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 355-362. https://doi.org/10.3280/oa-1180-c487.

Gozalo-Brizuela, R., Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*. ArXiv. http://arxiv.org/abs/2301.04655.

Loddo, F., Osello, A., Rimella, N., Rodriguez, D. P., Ugliotti, F. M., Ven tura, G. M. (2024). Approccio semantico alla rappresentazione: verso una collaborazione Uomo-AI per la misura della dismisura. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (A cura di), *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 3135-3154, srl. https://doi.org/10.3280/oa-1180-c630.

Pellicio, A., Saccucci, M., & Miele, V. (2023). AI Text-To-Image for the Representation of Treaties Texts. The Case Study of Le Vite by Vasari. In M. Cannella, A. Garozzo, S. Morena (A cura di). *Transizioni. Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione*. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 1824-183. https://doi.org/10.3280/oa-1016-c380.

Rasetti, G. (2023). Disegnare l'invisibile, il paesaggio. Esperimenti con intelligenza artificiale text to image. In M. Cannella, A. Garozzo, S. Morena (A cura di). *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 1954-1969. https://doi.org/10.3280/oa-1016-c387.

Trizio, I., Marra, A., Savini, F., Giallonardo, M., Cordisco, A., Saccucci, M. (2024). Misura o dismisura? Considerazioni e confronti tra NeRF e fotogrammetria digitale. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (A cura di). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 2113-2132. https://doi.org/10.3280/oa-1180-c578.

Wu, J., Gan, W., Chen, Z., Wan, S., Lin, H. (2023). *AI-Generated Content (AIGC): A Survey.* ArXiv. http://arxiv.org/abs/2304.06632.

Zhang, C., Zhang, C., Zhang, M., Kweon, I. S., & Kim, J. (2024). T*ext-to-image Diffusion Models in Generative AI: A Survey.* ArXiv. http://arxiv.org/abs/2303.07909.

Autori
Romina Nespeca, Università Politecnica delle Marche, r.nespeca@univpm.it
Renato Angeloni, Università Politecnica delle Marche, r.angeloni@univpm.it
Laura Coppetta, Università Politecnica delle Marche, l.coppetta@pm.univpm.it

Per citare questo capitolo: Romina Nespeca, Renato Angeloni, Laura Coppetta (2025). Parole che modellano spazi: IA generativa e rappresentazioni 3D architettoniche. 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. 3097-3120. DOI: 10.3280/oa-1430-c916.

 ISBN 9788835182412

èkphrasis

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

Words Shaping Spaces: Generative AI and Architectural 3D Representations

Romina Nespeca
Renato Angeloni
Laura Coppetta

Abstract

Generative Artificial Intelligence has revolutionized the relationship between words and images, positioning itself as a tool capable of generating graphic representations from textual descriptions. This study explores the potential of this modern ekphrastic process for the generation of three-dimensional reconstructions in the architectural field, focusing on three case studies: the orders as defined by Vitruvius in *De Architectura*, selected Renaissance paintings depicting built spaces, and an existing monumental building. By employing various tools for generating texts, images, and 3D models, the research highlights both the potential and the numerous critical issues related to the semantic and geometric fidelity of the generated representations. In a context where the relationship between words and images takes on new forms, AI thus emerges as a technology capable of enriching and potentially revolutionizing graphic representation methods, opening up new methodological and cultural perspectives in the disciplinary field of Drawing.

Keywords
Text-to-image-to-3D, image-to-3D, text-to-3D, generative AI, *èkphrasis*.

Use of textual prompts in *DALL-E* 3: Words Shaping Spaces: Generative AI and Architectural Representations (elaboration by the authors).

doi.org/10.3280/oa-1430-c916

Introduction

Artificial Intelligence (AI) is establishing itself as a tool capable of generating visual representations from textual descriptions [Arzomand *et al.* 2024], finding an intriguing affinity with the ancient rhetorical practice of *èkphrasis*. Defined by Hermogenes of Tarsus in the 2nd century AD, *èkphrasis* consists of a descriptive discourse capable of evoking vivid images through words, emphasizing details that enhance the understanding of the represented subject. Imbued with *enárgeia*, the ability to generate powerful mental images in the viewer's mind, this practice finds a modern counterpart in AI-driven generative representation methods, including deep neural networks, Generative Adversarial Networks (GANs), and diffusion models [Wu *et al*. 2023], such as Generative Pretrained Transformer (GPT) and Bidirectional Encoder Representations from Transformers (BERT), which can interpret written texts [Gozalo-Brizuela and Garrido-Merchan 2023].
Thanks to these technologies, tools like *MidJourney*, *DALL-E*, and *Stable Diffusion* enable the generation of images from textual descriptions (text-to-image), inaugurating a new mode of representation that positions the discipline of Drawing at the intersection of word and image [Zhang *et al*. 2024]. As recent studies have shown, images generated through such tools exhibit remarkable complexity and detail, as exemplified by representations of urban spaces that offer new perspectives on urban identity and design models [De Natale 2024]. Through technologies such as Natural Language Processing (NLP), AI not only creates but also acts as a mediator between the artwork and the audience, generating detailed descriptions that combine technical data with aesthetic interpretations [D'Uva 2024]. This process establishes a virtuous cycle in which AI interprets and enriches content, making linguistic analysis and network training essential for addressing the semantic complexities of texts and producing realistic images [Pellicio *et al*. 2023].
This evolution recalls medieval architectural descriptions, where text guided physical construction [Rasetti 2023]. However, the convergence of the ancient art of verbal description and AI-generated content raises new questions regarding the methodological and cultural implications of this collaboration between human language and AI's generative capabilities. Issues such as authorship and creativity emerge prominently, highlighting the tension between human control and algorithmic automation. Although the boundary between human and artificial creation is becoming increasingly blurred, AI-generated images and visual content remain closely tied to the descriptions provided by users [Battini 2023].
In this context, the present article aims to explore the use of AI in 3D modeling generated from textual and image-based inputs, a field still largely unexplored. Through an interdisciplinary approach that connects the theoretical roots of *èkphrasis* with the latest applications in digital representation [Battini and Martínez Chao 2024; Trizio *et al*. 2024], this research investigates how the precision of verbal description can become the cornerstone of a new *poiesis* shared between humans and machines [Condorelli *et al*. 2023; Loddo *et al*. 2024]. The proposed analysis, grounded in the fundamental principles of the discipline of Drawing, invites reflection on a landscape in which the relationship between word and image takes on new and evocative forms, contributing to the creation of representations that both evoke and innovate.
Could AI, then, acquire the expressive and communicative power that Hermogenes attributed to *èkphrasis*, thereby addressing the challenges of an ever-evolving reality?

Methodology and tools

Thanks to the potential of AI, *èkphrasis*, understood as a verbal description capable of evoking the visual representation of an object or an idea, takes on a new dimension. The interaction between language, text, and image is enriched, enabling the creation of three-dimensional representations through AI-mediated processing. Specifically, this study explores how processes such as text-to-3D and image-to-3D can be applied in the field of architectural representation, focusing on case studies belonging to three emblematic typologies:
– architectural orders: Doric, Ionic, and Corinthian, as described by Vitruvius;

– Renaissance paintings of architectural subjects: *The Ideal City of Urbino*, *The Marriage of the Virgin*, and *The School of Athens* by Raffaello;
– an existing monumental building: the Church of Santa Maria della Piazza in Ancona.

These case studies, with their distinct characteristics, provide fertile ground for investigating the potential and limitations of AI in translating textual or visual descriptions into three-dimensional representations.

The adopted methodology is based on the use of *ChatGPT* for textual description processing, *DALL-E* for generating synthetic two-dimensional images, and different platforms for generating three-dimensional models.

To better assess and compare the effectiveness of 3D model generation, four different platforms offering both text-to-3D and image-to-3D models were examined: *Rodin* by *Hyper3D*, *3D AI Studio*, *Meshy AI*, and *Tripo AI*. Each of these platforms employs different approaches for processing input data and creating three-dimensional geometries.

Using these tools, three workflows were defined, each leading to the creation of a 3D representation following a distinct process.

In the first workflow (text-to-image-to-3D), detailed textual descriptions were used as a starting point, Vitruvius' writings for the first case study and *Wikipedia* for the second. These descriptions were processed by *ChatGPT* to generate synthetic images through *DALL-E*. The resulting images were then uploaded to the four tools, which extracted the necessary information to create three-dimensional models.

The second approach (image-to-3D) started from existing images: representations of Vitruvian orders created by Daniele Barbaro and the selected paintings, *The Ideal City of Urbino*, *The Marriage of the Virgin*, and *The School of Athens* by Raffaello. As in the previous workflow, these images were then processed through the tools to generate their corresponding 3D models.

Finally, in the third workflow (tex- to-3D), textual prompts were generated via *ChatGPT* and used to create three-dimensional models.

For the third case study, involving an existing building, the methodology was pushed further. In the initial phase, individual photographic images and CAD drawings were processed to obtain three-dimensional models. Subsequently, the method was tested with multiple images. A first attempt involved uploading two sections, a plan, and the main elevation, while a second attempt used multiple photographic views of the church. These different approaches allowed for an exploration of the software's capabilities in reconstructing three-dimensional models from heterogeneous inputs, such as orthographic drawings and perspective images.

The tools used play a crucial role in the entire process. *ChatGPT*, with its ability to understand and generate natural language, is essential for translating complex descriptions into textual prompts suitable for image generation. The 3D model generation tools, on the other hand, represent a significant technological advancement, enabling the conversion of textual prompts and two-dimensional images into three-dimensional representations. However, despite the potential of these tools, significant challenges remain to be overcome, as the field of AI-generated 3D modeling is still relatively unexplored and in development.

Discussion and results

The experimental results highlighted various critical issues and potentials within the tested workflows. In the first case study, related to *De Architectura* by Vitruvius, significant difficulties emerged in generating the Doric capital: despite the detailed textual descriptions in the treatise, *ChatGPT* tends to generate images of Ionic capitals instead of Doric ones. Consequently, the resulting 3D models align with the generated image rather than the original description. In contrast, the Ionic and Corinthian capitals did not present significant generation issues (fig. 1).

The comparative analysis of different 3D model generation tools revealed contrasting results in processing Daniele Matteo Alvise Barbaro's drawings. *Rodin* and *Meshy* successfully generated three-dimensional models from two-dimensional representations, albeit with some inaccuracies. Specifically, *Rodin* mistakenly classified the columns as pilasters, while

Fig. 1. Text-to-Image-to-3D of the first case study: classical orders.

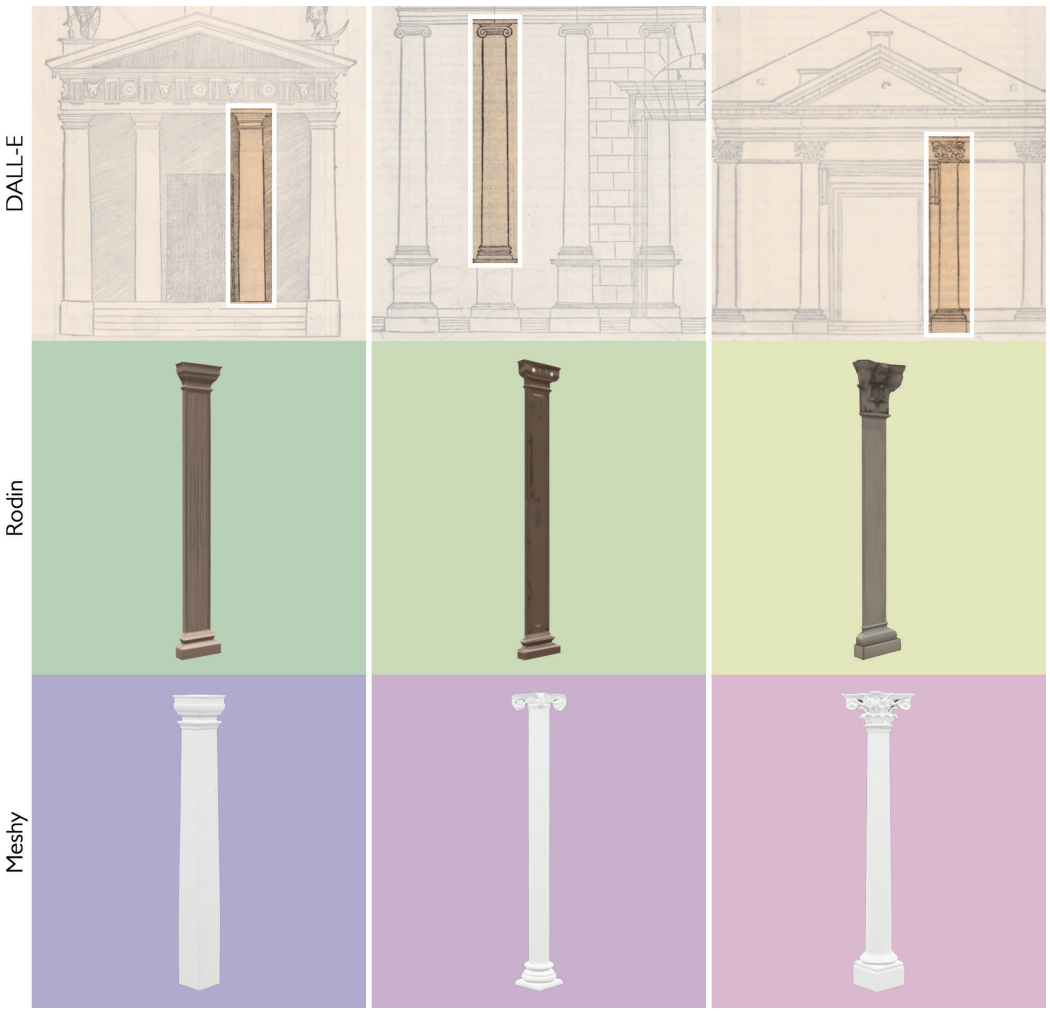

Fig. 2. Image-to-3D of the first case study: classical orders.

Meshy, in the case of the Doric order, produced a pillar with a square section, deviating from the original shape. Conversely, *3D AI Studio* and *Tripo AI* were unable to generate a 3D model from Barbaro's drawings (fig. 2).
When textual descriptions were used directly to generate 3D models, the results varied significantly depending on the tool used. *Rodin* produced short and stocky columns, with Doric capitals that deviated more from Vitruvian descriptions than Ionic and Corinthian ones. *Tripo* proved to be the most accurate tool in interpreting the request, although it still failed to properly reproduce a Doric column. Meanwhile, *Meshy* and *3D AI Studio* systematically generated Ionic columns, showing a persistent preference for this architectural order regardless of the textual input provided (fig. 3).
In the second case study, based on paintings of architectural subjects, the 3D models generated by *Rodin* directly from the original paintings proved unsatisfactory, with significant errors in understanding both architectural elements and figures, particularly in the case of *The School of Athens*. Conversely, *3D AI Studio* was able to recognize that *The School of Athens* represents an interior space, approximately reconstructing the room's volumetry (fig. 4). In both platforms, better results were obtained by manually isolating architectural elements, such as the temple in *The Ideal City* and *The Marriage of the Virgin* (fig. 5).
Additionally, using images generated by *ChatGPT* from *Wikipedia* textual descriptions, the 3D models produced by *Rodin* showed greater adherence to the original elements in the cases of *The Ideal City*'s temple and *The Marriage of the Virgin*, while *The School of Athens* continued to yield unsatisfactory results. On the other hand, *3D AI Studio*, in this case as

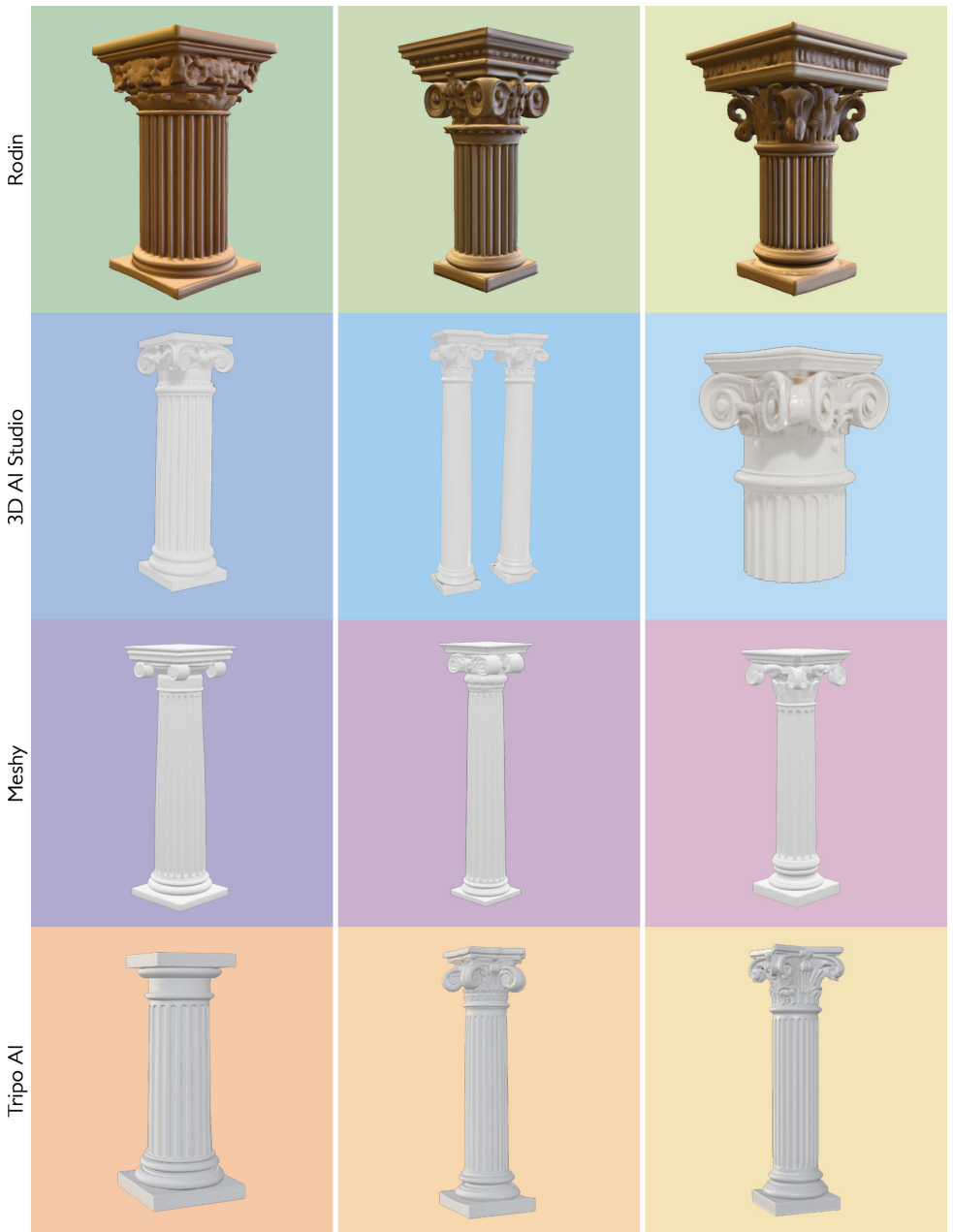

Fig. 3. Text-to-3D of the first case study: classical orders.

well, correctly recognized that *The School of Athens* represents an interior space, while in *The Ideal City,* it was able to reconstruct the temple but introduced elements that were not present in the two-dimensional input (fig. 6).
Finally, the direct generation of 3D models from textual descriptions of the paintings did not produce faithful reproductions of the originals in either of the two platforms (fig. 7).
In the third case study, concerning the generation of 3D models of Santa Maria della Piazza in Ancona (AN), the most accurate results were obtained by uploading photographic images taken from different angles, allowing *Rodin* to reconstruct the existing architecture with greater precision (figs. 8, 9).
Starting from individual CAD drawings (plan, section, elevation), the results were inconsistent: the elevation closely resembled the original, while the plan and section required significant user interpretation (fig. 10). Using multiple orthogonal projections, the resulting model appeared inconsistent with the original drawings (fig. 11).

Fig. 4. Image-to-3D of the second case study: architectural paintings. Images without segmentation.

The results of this study highlight both strengths and limitations, emphasizing the role of human intervention in the methodology, such as image segmentation to facilitate recognition and, consequently, the three-dimensional reconstruction of certain elements. The use of AI tools, such as *ChatGPT* and platforms for 3D model generation, represents an innovative approach in Digital Heritage. However, the random nature of generative processes makes it difficult to predict outcomes, particularly in complex cases based on textual descriptions or paintings.
Although techniques such as cropped images or multiple photographs have improved accuracy, significant limitations remain, especially in architectural details. The need for fine-tuning techniques and models trained on specific contexts emerges as a potential solution, but it requires substantial resources and close interdisciplinary collaboration.

Conclusions

This study demonstrated how the intersection between the ancient practice of *èkphrasis* and the potential of Generative Artificial Intelligence can foster a new relationship between word and image, human and machine. The conducted experiments highlight AI's ability to translate verbal descriptions into images and three-dimensional models; however, this collaboration is not without challenges. While the experimental results are promising, they also reveal certain limitations. In the analyzed case studies, the text-to-3D process exhibited critical issues concerning semantic and geometric fidelity in relation to the textual inputs. For example, the inability of the tested tools to accurately represent a Doric capital or to reconstruct painted architectural elements in three dimensions suggests that the potential of such systems is still evolving. Despite these challenges, generative AI presents a revolutionary opportunity to enhance analytical and creative processes.

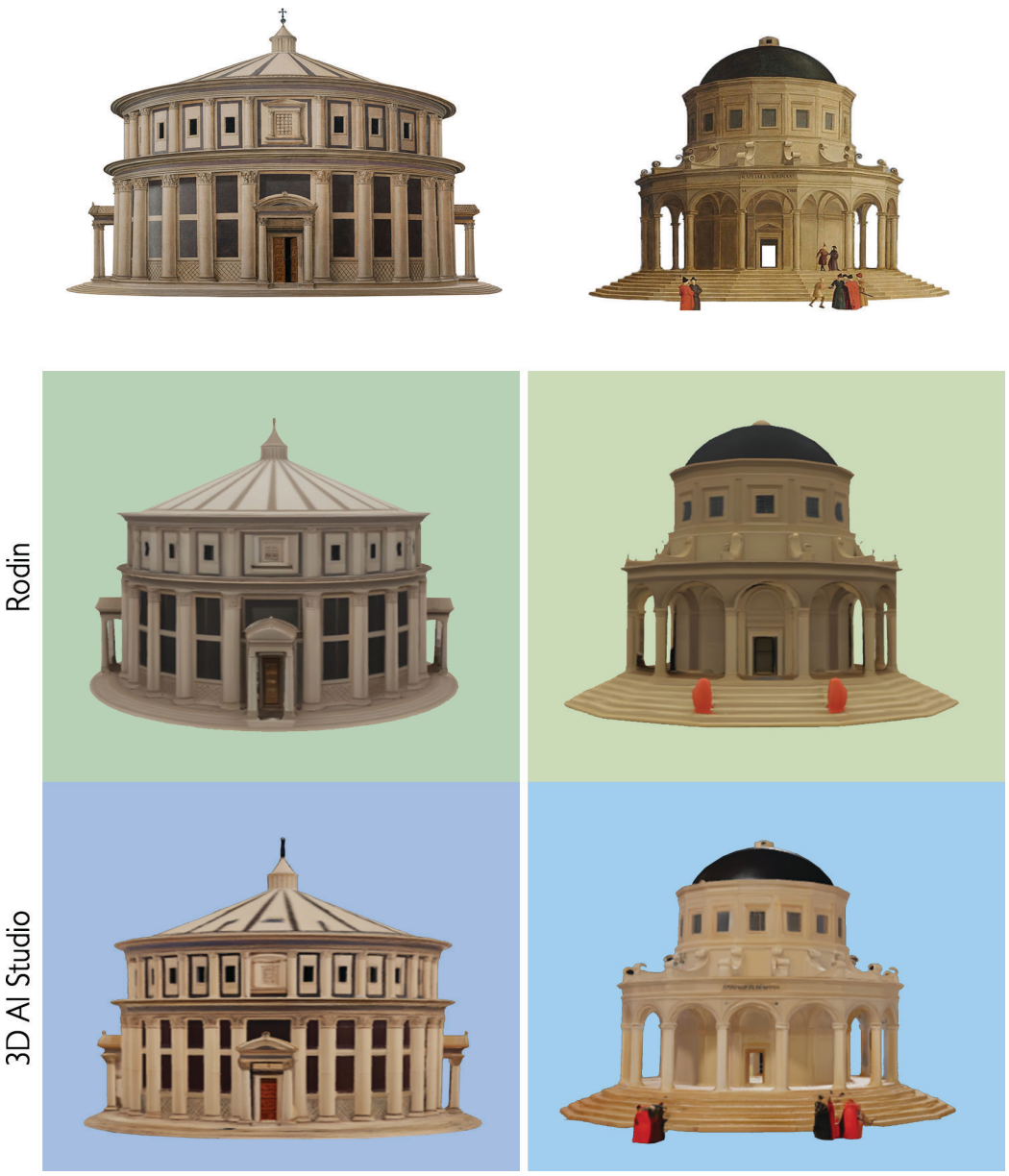

Fig. 5. Imag- to-3D of the second case study: architectural paintings. Segmented images.

In conclusion, this study encourages viewing AI not as a substitute but as a co-creator, capable of augmenting human imagination and expanding the boundaries of architectural and artistic representation. The co-creation phase clearly emerges in the adopted methodology. For instance, to achieve accurate results, it was necessary to manually segment paintings, allowing the AI to correctly reconstruct the depicted objects.
Similarly, generating a proper Doric capital, unsuccessful when relying solely on Vitruvius' description, became possible when providing an explicit reference to an existing building, such as the Parthenon ("Generate a Doric capital like the one from the Parthenon").
This demonstrates the importance of contextual and precise instructions in guiding generative outputs.
This observation suggests that improving the reliability of generative models requires greater specialization, both in interpreting references to existing objects and in handling orthographic representations, an area where AI, primarily trained on photographic datasets, still exhibits gaps. In this context, a future perspective could involve training

Fig. 6. Text to Image to 3D of the second case study: architectural paintings.

Fig. 7. Text-to-3D of the second case study: architectural paintings.

neural networks on datasets containing orthogonal projections, thereby increasing AI's ability to manage more complex and specific graphic contexts.
Further research could focus on integrating interdisciplinary knowledge, refining AI's capability to generate more faithful and contextually relevant 3D models that can be applied in practical domains such as HBIM and other information systems. In this evolving scenario, human intervention will remain essential for guiding, correcting, and optimizing the process, leveraging AI as a tool for co-creation rather than a mere automation mechanism for the representation and analysis of cultural heritage.

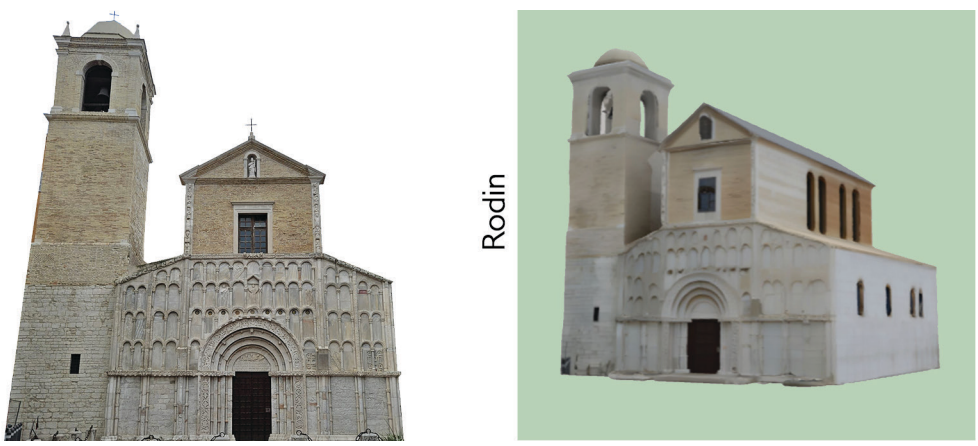

Fig. 8. Image-to-3D of the third case study: architectural building. Single photo.

Fig. 9. Image-to-3D of the third case study: architectural building. Multiple photos.

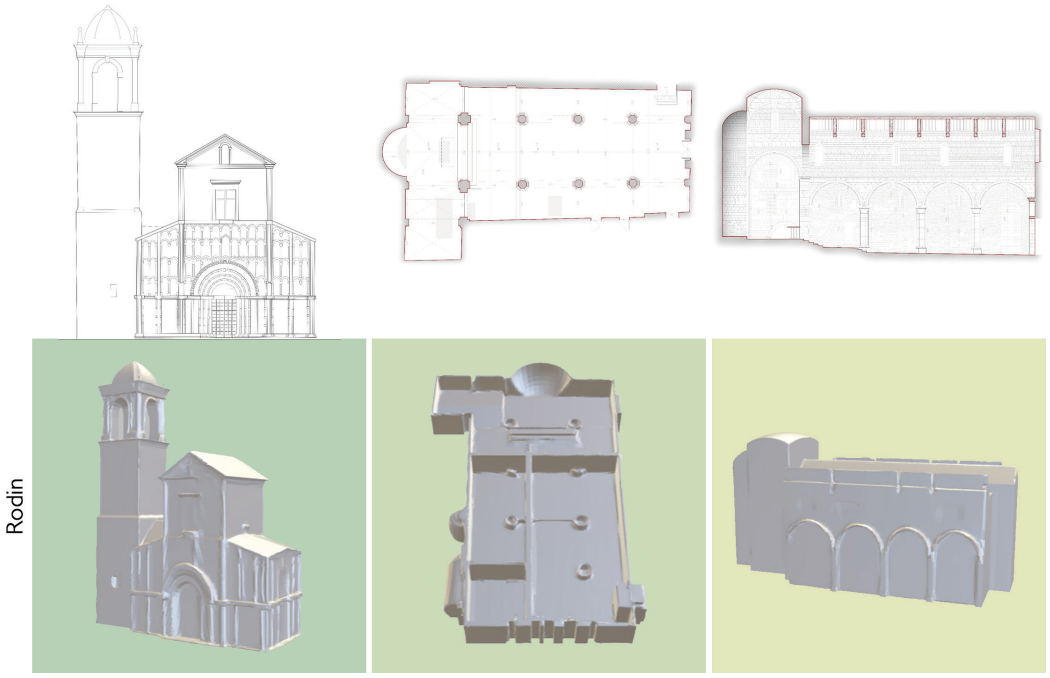

Fig. 10. Image-to-3D of the third case study: architectural building. Orthographic drawing.

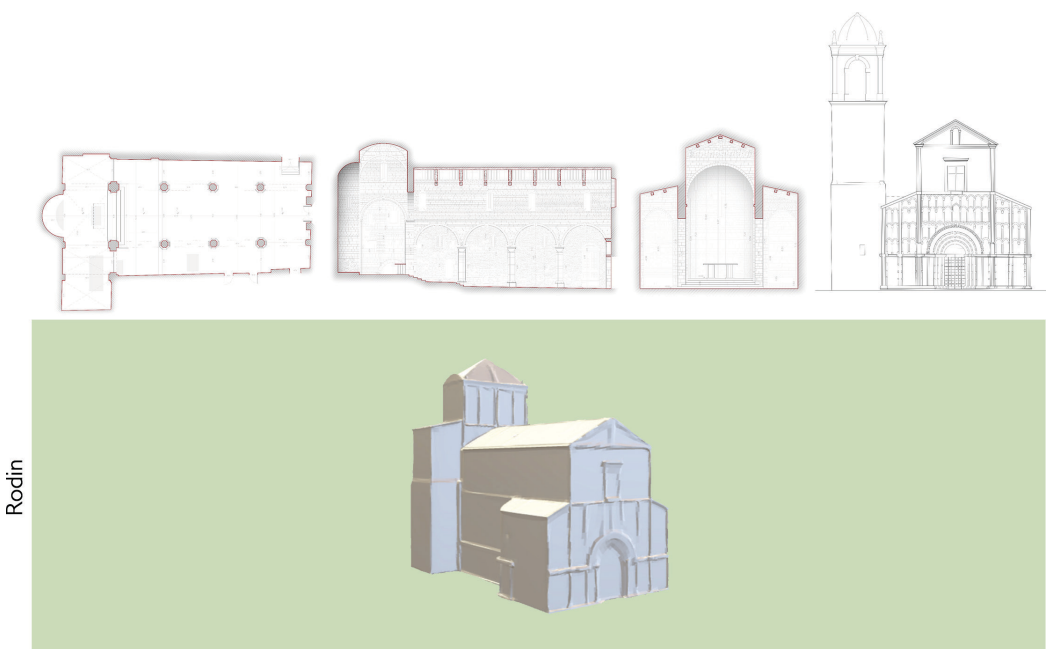

Fig. 11. Image-to-3D of the third case study: architectural building. Multiple orthographic drawings.

Aknowledgment and credits

The research was funded by *Piano Nazionale di Ripresa e Resilienza, Missione 1 "Digitalizzazione, innovazione, competitività, cultura e turismo", Componente 3 "Turismo e cultura 4.0", MISURA 3 "Industrie culturali e creative", INVESTIMENTO 3.3 "Capacity building per gli operatori della cultura per gestire la transizione digitale e verde", Sub-investimento 3.3.1 "Interventi per migliorare l'ecosistema in cui operano i settori culturali e creativi, incoraggiando la cooperazione tra operatori culturali e organizzazioni e facilitando upskill e reskill", funded by the European Union – NextGenerationEU – Avviso di cui al Decreto Direttoriale del 9 giugno 2023, n. 149. Progetto SkillsHub "Servizi per Culture&Creativity" Codice progetto PNRRI-202300031 60497/2, Codice CUP C31B23000430004, ammesso a finanziamento con Decreto Direttoriale n. 737 del 7 dicembre 2023.*

Reference List

Arzomand, K., Rustell, M., Kalganova, T. (2024). From ruins to reconstruction: Harnessing text-to-image AI for restoring historical architectures. *Challenge Journal of Structural Mechanics,* 10(2), 69-85. https://doi.org/10.20528/cjsmec.2024.02.004.

Battini, C. (2023). Intelligenza artificiale tra scienza e creatività. Casi studio nelle arti visive. In M. Cannella, A. Garozzo, S. Morena (Eds.). *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 2380-2393. https://doi.org/10.3280/oa-1016-c411.

Battini, C., Martínez Chao, T. E. (2024). Progettazione e IA. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (Eds.). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 61-76. https://doi.org/10.3280/oa-1180-c472.

Condorelli, F., Luigini, A., Nicastro, G., Tramelli, B. (2023). Disegno e intelligenza artificiale. Enunciati teorici e prassi sperimentale per una poiesi condivisa. In M. Cannella, A. Garozzo, S. Morena (Eds.). *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 2623-2640. https://doi.org/10.3280/oa-1016-c426.

De Natale, I. (2024). La misura dell'identità urbana con l'IA generativa. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (Eds.). *Misura / Dismisura.* Atti del 45° Convegno Inter- nazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 2769-2780. https://doi.org/10.3280/oa-1180-c611.

D'Uva, D. (2024). AI-Enhanced Facade Design: Exploring the Synergy of Generative Models and Architectural Creativity. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (Eds.). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 355-362. https://doi.org/10.3280/oa-1180-c487.

Gozalo-Brizuela, R., Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models.* ArXiv. http://arxiv.org/abs/2301.04655.

Loddo, F., Osello, A., Rimella, N., Rodriguez, D. P., Ugliotti, F. M., Ven tura, G. M. (2024). Approccio semantico alla rappresentazione: verso una collaborazione Uomo-AI per la misura della dismisura. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (Eds.). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 3135-3154, srl. https://doi.org/10.3280/oa-1180-c630.

Pellicio, A., Saccucci, M., & Miele, V. (2023). AI Text-To-Image for the Representation of Treaties Texts. The Case Study of Le Vite by Vasari. In M. Cannella, A. Garozzo, S. Morena (Eds.). *Transizioni. Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione*. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 1824-183. https://doi.org/10.3280/oa-1016-c380.

Rasetti, G. (2023). Disegnare l'invisibile, il paesaggio. Esperimenti con intelligenza artificiale text to image. In M. Cannella, A. Garozzo, S. Morena (Eds.). *Transizioni.* Atti del 44° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Palermo, 14-16 settembre 2023. Milano: FrancoAngeli, pp. 1954-1969. https://doi.org/10.3280/oa-1016-c387.

Trizio, I., Marra, A., Savini, F., Giallonardo, M., Cordisco, A., Saccucci, M. (2024). Misura o dismisura? Considerazioni e confronti tra NeRF e fotogrammetria digitale. In F. Bergamo, A. Calandriello, M. Ciammaichella, I. Friso, F. Gay, G. Liva, C. Monteleone (Eds.). *Misura / Dismisura.* Atti del 45° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Padova, Venezia 12-14 settembre 2024. Milano: FrancoAngeli, pp. 2113-2132. https://doi.org/10.3280/oa-1180-c578.

Wu, J., Gan, W., Chen, Z., Wan, S., Lin, H. (2023). *AI-Generated Content (AIGC): A Survey.* ArXiv. http://arxiv.org/abs/2304.06632.

Zhang, C., Zhang, C., Zhang, M., Kweon, I. S., & Kim, J. (2024). T*ext-to-image Diffusion Models in Generative AI: A Survey.* ArXiv. http://arxiv.org/abs/2303.07909.

Authors
Romina Nespeca, Università Politecnica delle Marche, r.nespeca@univpm.it
Renato Angeloni, Università Politecnica delle Marche, r.angeloni@univpm.it
Laura Coppetta, Università Politecnica delle Marche, l.coppetta@pm.univpm.it

To cite this chapter: Romina Nespeca, Renato Angeloni, Laura Coppetta (2025). Words Shaping Spaces: Generative AI and Architectural 3D Representations. 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. 3097-3120. DOI: 10.3280/oa-1430-c916.

 ISBN 9788835182412