

L'IA PER LA GENERAZIONE DI IMMAGINI ARCHEOLOGICHE

1. INTRODUZIONE

Da qualche anno l'intelligenza artificiale è entrata a fare parte della quotidianità non solo dei ricercatori ma anche di tutti i cittadini e le previsioni vedono questo fenomeno in crescita. Se le prime applicazioni dell'IA, qualche decennio fa, erano riservate agli addetti ai lavori, oggi il grande pubblico ha iniziato a fruire in maniera talvolta consapevole di questa tipologia di algoritmi. Le due principali tipologie di IA generalmente utilizzate sono la generazione di testo e la generazione di immagini. In questo articolo si intende esaminare quest'ultima. Pur non volendo cadere in tecnicismi non necessari, pare fondamentale cercare di capire meglio come funziona la generazione di immagini artificiali dal punto di vista tecnico (ELASRI *et al.* 2022). La generazione di contenuti (testuali, musicali, fotografici e video) ha alla base del suo funzionamento la rete neurale artificiale del deep learning, branca del machine learning. Questa tipologia di algoritmo, la quale riproduce idealmente la struttura della mente umana, basa la sua conoscenza su una grande quantità di dati (datasets).

I principali sistemi generativi sono i Diffusion Models e i Generative Adversarial Networks (BANH, STROBEL 2023). Il primo è un modello generativo che, come si è accennato, viene allenato attraverso grandi banche dati. Le immagini del database vengono destrutturate attraverso l'inserimento di rumore (noise) per poi essere ricostituite con differenti variazioni nel prodotto finale (YANG *et al.* 2023, 6, fig. 2). Il secondo sistema di generazione immagini IA, sviluppato per la prima volta nel 2014, è il Generative Adversarial Network (GAN). In questo caso l'algoritmo è dotato di due reti differenti contraddittorie tra loro (una rete generativa e una rete discriminatoria), una che genera esempi falsi e l'altra che ha il compito di confutarla (GOODFELLOW *et al.* 2014). Gli sviluppatori di questi sistemi, soprattutto quelli ad alti livelli, sono principalmente privati, per cui, ad oggi, risulta difficoltoso avere piena conoscenza della tipologia e della quantità dei dati impiegati. Questa tipologia di software ha iniziato ad essere velocemente utilizzata in tutto il mondo dalla più disparata tipologia di utenti, grazie non ultimo alla loro generale gratuità. Tra gli utilizzi vi è la creazione di contenuti d'intrattenimento (si pensi ai content creators sui social media come TikTok ed Instagram), ma anche usi a scopo commerciale e pubblicitario.

Se l'IA è entrata a far parte del mondo dell'archeologia in molti modi già dal secolo scorso (a partire dal GIS fino al Remote Sensing per l'individuazione

di siti archeologici (ARGYROU, AGAPIOU 2022) o al riconoscimento di frammenti di ceramiche attraverso l'applicazione ArchAIDE), il suo utilizzo per la generazione di immagini risulta essere ancora acerbo.

2. LA GENERAZIONE DI IMMAGINI

2.1 *Midjourney e Playground AI*

Ad oggi sono estremamente numerosi, e costantemente in aumento, le applicazioni di generazioni immagini con IA. Per questo studio sono state prese in considerazione due tra le web-app più famose e utilizzate attualmente: Midjourney e Playground AI. Sia Midjourney che Playground AI sono stati sviluppati in piccoli laboratori privati, con sede ufficiale a San Francisco. Nella pratica le immagini vengono generate a partire da alcune frasi scritte (per cui il processo è definito text-to-image), ma esiste, come si approfondirà in seguito, la possibilità di inserire proprie immagini da cui generare l'illustrazione finale (per cui il processo è definito image-to-image). Entrambe le applicazioni sono estremamente intuitive, anche se presentano qualche piccola differenza di interfaccia. In entrambi i casi le immagini, una volta generate, vengono pubblicate all'interno dell'app non risultando di proprietà dell'utente che ha sviluppato la loro generazione.

Midjourney, che si appoggia alla piattaforma di messaggistica istantanea Discord, presenta una scheda di lavoro che riproduce a tutti gli effetti le sembianze di una chat. Per attivare il processo generativo è necessario scrivere “/imagine” all'interno dell'apposita barra di testo, situata nella parte inferiore dello schermo e automaticamente comparirà la scritta prompt, alla cui destra si potrà digitare il testo che si ha in mente (ad es. “genera l'immagine di una mela che cade da un albero”). Una volta premuto il tasto “invio” il bot della chat provvederà, in un tempo variabile tra i 10 e i 70 secondi, a restituire quattro varianti della sua interpretazione del testo. La scrittura del prompt va ponderata, selezionando le parole che meglio rappresentano la scena di nostro interesse. Per non appesantire il prompt andrebbero utilizzate tra le 10 e le 30 parole, impiegarne un numero maggiore vorrebbe dire sovraccaricare il sistema di informazioni che non verrebbero processate adeguatamente, mentre un numero inferiore lascerebbero eccessiva libertà al software, portando alla realizzazione di un'immagine molto lontana da quella pensata.

Playground AI presenta, a differenza di Midjourney, una vera e propria schermata di lavoro, divisa in uno spazio centrale, dove è presente la casella del prompt e dove si sviluppano le figure, e due colonne ai lati con alcuni parametri cui sono applicabili varie modifiche. Il processo da seguire per ottenere le immagini è però lo stesso di Midjourney. Si segnala, tuttavia, che il design di questo software è in costante cambiamento. Negli ultimi mesi Playground

AI ha reso possibile il suo utilizzo tramite la piattaforma Discord, seguendo l'esempio di Midjourney.

Una funzionalità sviluppata meglio in Playground AI rispetto a Midjourney è quella dell'image-to-text. Il software permette infatti all'utente di stabilire la "forza" dell'immagine di partenza, cioè in percentuale quanto l'immagine finale dovrà somigliare a quella iniziale, anche se purtroppo non vi è la possibilità di segnalare quali elementi nello specifico vanno mantenuti e quali no. Per quanto riguarda le correzioni, entrambe le applicazioni permettono di modificare, seppur in maniera limitata, le immagini ottenute. Le raffigurazioni generate da Midjourney, selezionate singolarmente, possono essere modificate attraverso due modalità. La prima prevede la variazione dell'intera immagine, mentre la seconda la modifica di una selezione, effettuata dall'utente, della figura. In entrambi i casi l'alterazione da effettuare viene stabilita dall'IA e non vi è la possibilità da parte del fruitore di suggerire un cambiamento. Playground AI permette anche di selezionare manualmente un oggetto della figura per poi cancellarlo; è anche possibile selezionare una zona specifica della scena per effettuare delle modifiche attraverso l'inserimento di un prompt, tentando una correzione più puntuale.

3. LE RICOSTRUZIONI DEL CONTESTO DELL'ABBAZIA CISTERCENSE DI SAN GALGANO

Per testare a 360 gradi lo strumento si è deciso di portare avanti la sperimentazione della generazione di immagini con IA nel contesto dell'Abbazia cistercense di San Galgano. L'abbazia di San Galgano si trova nell'Alta Val di Merse, nel territorio del Comune di Chiusdino (SI). Le ricerche archeologiche, portate avanti dal 2019 dal LIAAM (Laboratorio di Informatica Applicata all'Archeologia Medievale), con direzione scientifica del Prof. Marco Valenti, hanno permesso di comprendere meglio le fasi medievali e post medievali della chiesa e di alcuni ambienti della cittadella monastica cistercense (BERTOLDI, NARDINI 2022; MENGHINI, PALMAS 2022).

La conoscenza attuale del complesso, nonostante sia ancora incompleta, permette di giudicare con occhio critico e minuzioso le immagini generate dall'IA. Tra i primi tentativi vi è la ricostruzione della fase di abbandono del refettorio monastico (riferibile alla fine del XIV secolo), scavato durante le campagne del 2021 e del 2022. Le ricostruzioni sono state ottenute attraverso l'utilizzo di variazioni al generico prompt "refettorio cistercense abbandonato" riguardo la tipologia dei materiali di costruzione, delle finestre e degli elementi architettonici. La sperimentazione sul testo ha mostrato dei forti limiti sull'alterazione puntuale di alcuni elementi. Un errore in entrambe le immagini è infatti il materiale utilizzato per il pavimento. Grazie al dato di scavo si può affermare che il rivestimento pavimentale fosse in legno e

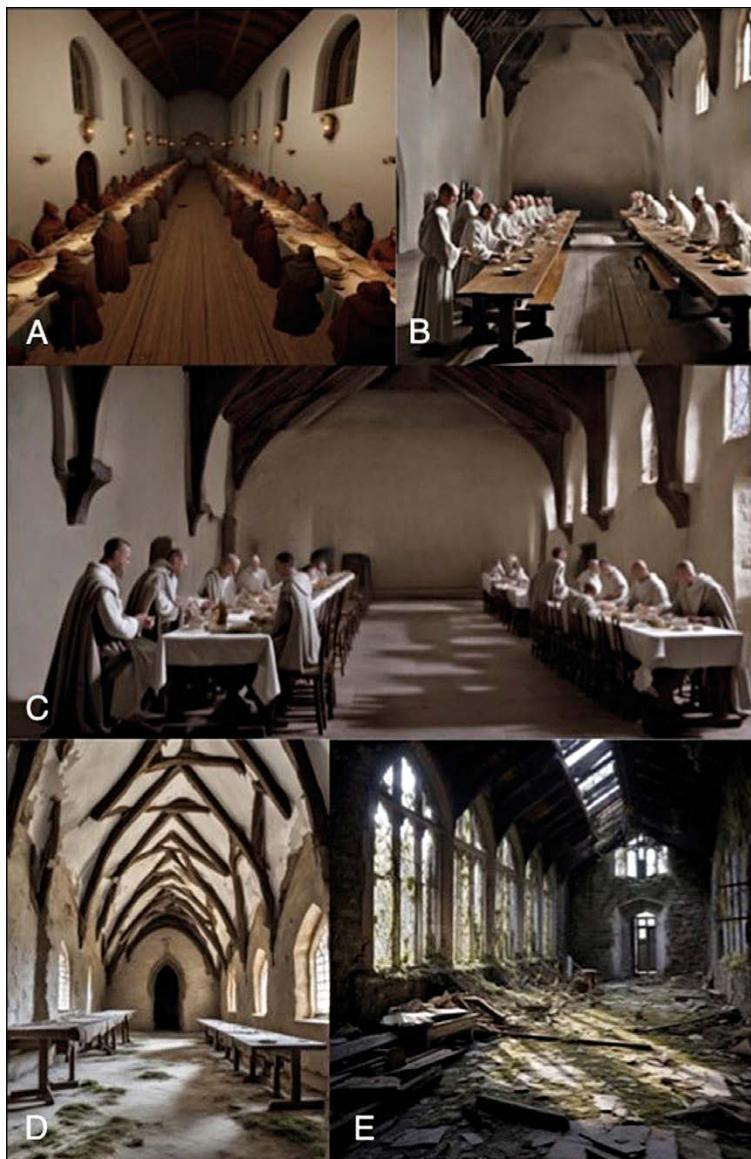


Fig. 1 – A) Midjourney, ricostruzione della fase di vita del refettorio a partire da una ricostruzione 3D. I due principali difetti sono costituiti dall'eccessiva lunghezza dell'ambiente e dal colore delle tuniche dei monaci; B) Playground, ricostruzione della fase di vita del refettorio; C) Playground, ricostruzione della prima fase di abbandono del refettorio; D) Playground, ricostruzione della seconda fase di abbandono del refettorio; E) Midjourney, ricostruzione di una fase avanzata di abbandono del refettorio.

non in pietra. Questo materiale viene utilizzato erroneamente nella Fig. 1E anche nella costituzione dei muri perimetrali, che probabilmente erano costituiti da mattoni rivestiti di intonaco, come si può vedere nella Fig. 1D.

Sulla forma rettangolare dell'ambiente l'IA è allenata e comprende il significato di refettorio. Sono maggiori le difficoltà quando si tratta di rappresentare elementi specifici di un contesto come quello di San Galgano. Alcuni dettagli, tuttavia, non possono essere confermati né confutati. Si pensi all'altezza, la dimensione e la forma delle finestre, la forma del tetto, l'altezza dei muri, la forma della porta, la tipologia di tavoli. È evidente poi che non si possa determinare nel dettaglio neanche il processo post de-posizionale che ha portato al crollo del tetto del refettorio. Di conseguenza il risultato dato dall'IA si può ritenere non completamente corretto ma abbastanza soddisfacente.

Si è poi tentato di ottenere una ricostruzione della fase di vita del refettorio, in utilizzo dagli inizi del XIII secolo alla fine del XIV secolo. La Fig. 1A, creata con Midjourney, è stata generata da un'immagine (image-to-image), tratta da una restituzione 3D dell'ambiente (cortesia di Stefano Bertoldi e Manuele Putti), che è stata presa dal software come base, aggiungendo la presenza dei monaci. La lunghezza eccessiva dell'ambiente è certamente un'imprecisione, ma un vero e proprio errore è il colore della tunica dei monaci. È infatti noto che i religiosi cistercensi indossassero una tunica di colore bianco e non di colore marrone. Questo inconveniente, cui Midjourney non è stato in grado di ovviare, è stato risolto con l'utilizzo di Playground AI, come si può vedere nelle Figg. 1B e 1C.

Per la creazione di queste due figure è stata utilizzata la generazione text-to-image, che ha portato a due risultati abbastanza soddisfacenti. La Fig. 1B risulta quella maggiormente corretta, in quanto rispetto alla 1C presenta il pavimento in legno e una tipologia di seduta più plausibile. Sia nella Fig. 1B che nella 1C si è tentato di inserire una porta sulla parete di fondo, purtroppo con scarso successo. Un'ulteriore imperfezione che si può riscontrare nella Fig. 1B sono i tavoli leggermente decentrati, soprattutto quello di sinistra, che si trova eccessivamente spostato verso il centro dell'ambiente. È evidente che si tratti di un dettaglio, che, tuttavia, non si è stati in grado di modificare.

Rimanendo sulle ricostruzioni basate sui dati di scavo, nella Fig. 4D si è tentato di ricostruire, partendo da una foto scattata durante lo scavo (Fig. 4C), la fase post medievale della chiesa abbaziale. Durante lo scavo sono state infatti individuate alcune buche di palo, riferibili alla fase di abbandono del XVII secolo, che formavano una recinzione, dimostrando l'utilizzo di tale frazione della chiesa in questo periodo come recinto per gli animali. L'immagine, ottenuta con variazioni al prompt “aggiungi dei bovini e una recinzione in legno”, non presenta grossi errori e mostra un risultato accettabile. Tuttavia,



Fig. 2 – A) Playground, un'enciclopedia sulle specie acquatiche; B) Midjourney, il trasporto merci dalla costa toscana all'Abbazia di San Galgano; C) Playground, scena di pesca di fiume da parte di un monaco dell'Abbazia.

appare ancora una volta il limite dell'IA che in questo caso non è stata in grado di inserire un elemento semplice come la palizzata.

Oltre alle ricostruzioni degli ambienti scavati, che richiedono un maggiore rigore nella ricostruzione, un ulteriore utilizzo di questo tipo di IA è possibile per la rappresentazione di momenti quotidiani all'interno dell'Abbazia, di dinamiche di scambio commerciale o nella ricostruzione del paesaggio antico della Val di Merse. Quest'ultimo, plasmato nel medioevo dall'Abbazia di San Galgano, aveva alla base un complesso ed organizzato sistema di sfruttamento delle risorse del territorio.

Nella Fig. 2C si è tentato di ricostruire una scena di pesca al fiume con protagonista un monaco. In questo caso il dato archeologico non è in grado di confermare o confutare l'ipotesi rappresentata. Il dato che si ha, grazie allo scavo del refettorio e delle cucine, è l'alta percentuale di resti ittici, di acqua salata e di acqua dolce, che permettono di chiarire in parte la dieta all'interno della cittadella monastica. Si segnala però che nella raffigurazione sono presenti alcuni elementi bizzarri, che non è stato possibile modificare, come il volto e la mano sinistra del monaco.



Fig. 3 – A) Playground, scena di vita quotidiana all'interno dello *scriptorium*; B) Playground, ipotesi ricostruttiva del paesaggio della Val di Merse nel Medioevo; C) Playground, scena di vita quotidiana all'interno dello *scriptorium*.

Sulla stessa tematica è stata pensata la Fig. 2B. L'idea iniziale era quella di rappresentare un carico di pesce che dalla riviera maremmana – da cui parte della materia prima partiva – giungesse all'Abbazia. Ad un primo sguardo, nella sua totalità, l'immagine si può ritenere soddisfacente. Tuttavia, soffermandosi ad analizzare il carro si nota il carico poco realistico costituito da frutti arancioni non bene identificati. I molteplici tentativi di ricreazione di barili contenenti il pesce non hanno avuto, come in alcuni casi precedentemente trattati, una buona esecuzione. Bisogna pensare quindi che in questi casi il software non presenta dati simili da cui originare l'immagine.

Proseguendo nella sperimentazione riguardante l'archeologia del paesaggio, si è voluta tentare la rappresentazione di una grancia all'Abbazia (Fig. 3B). La gestione del territorio attraverso le grance è una caratteristica ben nota dell'ordine cistercense. Nel caso specifico dell'Abbazia di San Galgano si hanno numerosi documenti cartografici e d'archivio che la citano come proprietaria di mulini e grance nel territorio dell'attuale Val di Merse.

Anche in questo caso l'intento è di riprodurre idealmente il paesaggio e non ricreare fedelmente la morfologia del territorio e la struttura architettonica della grancia. Nella sua totalità la figura permette di visualizzare come dovesse apparire una struttura di questo genere nel Medioevo: caratterizzata da più corpi di fabbrica – magazzini e luoghi di lavorazione – isolata nella campagna e circondata da terreni agricoli di diversa natura. Altri utilizzi simili sono visibili nelle Fig. 2A – la riproduzione di un'encyclopedia sulle specie acquatiche – la 3A e la 3C – che rappresentano due momenti di vita all'interno, rispettivamente, dello scriptorium e del chiostro.

4. CONCLUSIONI

Dopo una prima sperimentazione si può affermare che sono due le categorie di immagini archeologiche generabili con l'IA, la prima è la ricostruzione di ambienti effettivamente scavati, mentre la seconda è la riproduzione di scenari legati alla vita del monastero, di cui non vi è – o non vi può essere – un riscontro archeologico. Nel primo gruppo rientrano le ricostruzioni delle fasi di vita e di abbandono del refettorio e la fase di abbandono della chiesa abbaziale, mentre nella seconda categoria rientrano le scene di pesca, di trasporto di vivande, di scrittura di manoscritti e quelle del paesaggio. Se nella prima categoria è necessario un buon livello di accuratezza, nel secondo le scene rappresentate permettono una libertà di sperimentazione maggiore, non potendo essere confutate dal dato archeologico.

Tra i vantaggi si annovera l'economia dello strumento, sia in termini di costi che di tempi. Quest'ultimi come si è detto sono molto brevi, in media sessanta secondi per la generazione delle immagini e qualche minuto per la scelta del prompt. Anche i costi sono molto bassi. Playground AI presenta

una versione gratuita che permette di generare 500 immagini e attuare poche modifiche al mese, mentre si può acquistare per dodici euro al mese la versione a pagamento, che permette di attuare modifiche illimitate. Midjourney invece non presenta una versione gratuita, ma l'abbonamento mensile costa solo otto euro.



Fig. 4 – A) Foto dell’area di scavo del transetto sud dell’Abbazia, foto di partenza per la ricostruzione; B) uno dei tentativi non riusciti di restituzione della fase post medievale dell’Abbazia; C) foto dell’area di scavo del transetto sud dell’Abbazia, foto di partenza per la ricostruzione; D) ricostruzione della fase post medievale dell’Abbazia; E) foto aerea dell’Abbazia di San Galgano; F) uno dei tentativi non riusciti di ricostruzione della cittadella monastica nel periodo medievale.

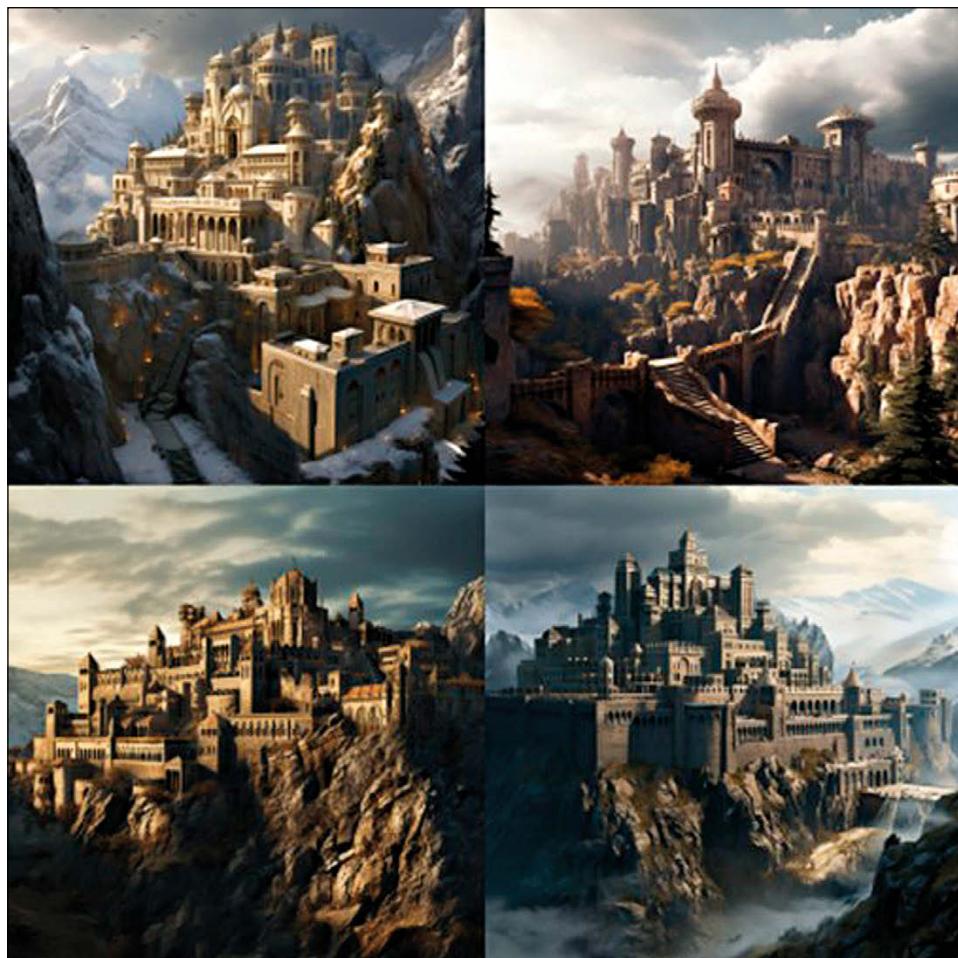


Fig. 5 – Il risultato ottenuto dal software Midjourney al prompt “cittadella monastica medievale”

Al momento si può quindi affermare che entrambi siano uno strumento utile per generare ricostruzioni ipotetiche anche durante le fasi di scavo, permettendo di visualizzare, non senza qualche inesattezza, le ipotesi in progress. D'altra parte, questa tipologia di IA presenta numerosi limiti, i quali risultano particolarmente evidenti quando sia necessario inserire dettagli puntuali e derivanti dal dato archeologico. Alcune parole vengono comprese dal software mentre altre no, senza però che si evidenzi uno schema di scelta del lessico da seguire per un'esecuzione ottimale del processo.

Le due principali conseguenze dell'incomprensione del lessico utilizzato sono il mancato inserimento dell'elemento richiesto (Fig. 4D) o la creazione di un'immagine estremamente astratta (Fig. 4B, Fig. 4F). La causa dei limiti di questo strumento sono certamente le banche dati. Come si è detto all'inizio, per funzionare i generatori di immagini traggono spunto e idee da grandi archivi, che al momento difettano di dettagliate immagini di contesti archeologici e complessi architettonici antichi.

Per risolvere il problema sarebbe quindi necessaria la creazione di una banca dati con foto di elementi architettonici, archeologici e decorativi (archi, finestre, cornici, mensole, capitelli, pavimenti etc.) per cui ad una ipotetica e generica richiesta "cittadella monastica cistercense" il software possa rispondere attingendo a particolari complessi architettonici, ancora esistenti, e non a fantasiosi scenari utopici (Fig. 5). Solo attraverso il lavoro congiunto di archeologi ed informatici specializzati in questo settore si potrà arrivare ad ottenere una tipologia di IA generativa in grado di restituire risultati più fedeli e, si spera, utilizzabili a tutti.

GILIA GHIGLIA

Ricercatrice indipendente
giuliaghiglia98@gmail.com

BIBLIOGRAFIA

- ARGYROU A., AGAPIOU A. 2022, *A review of Artificial Intelligence and remote sensing for archaeological research*, Special Issue Remote Sensing, Archaeology and Heritage Research: Researching the Past from Satellite, Aerial and Terrestrial Methods, «Remote Sensing», 14, 23, 1-23.
- BANH L., STROBEL G. 2023, *Generative artificial intelligence*, «Electronic Markets», 33, 1, 62-79.
- BEN-ROMDHANE H., FRANCIS D., CHERIF C., PAVLOPOULOS K., GHEDIRA H., GRIFFITHS S. 2023, *Detecting and predicting archaeological sites using remote sensing and Machine Learning – Application to the Saruq Al-Hadid Site, Dubai, UAE*, «Geosciences», 13, 2, 179, 1-34.
- BERTOLDI S., NARDINI A. 2022, *Lo scavo all'interno della chiesa di San Galgano (Chiusdino - SI): dalla fondazione ai riusi postmedievali*, in M. MILANESE (ed.), IX Congresso Nazionale di Archeologia Medievale (Alghero 2022), Sesto Fiorentino (FI), All’Insegna del Giglio, 190-195.
- ELASRI M., ELHARROUSS O., AL-MAADEED S., TAIRI H. 2022, *Image generation: A Review*, «Neural Processing Letters», 1-39.
- GOODFELLOW I.J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., Bengio Y. 2014, *Generative Adversarial Nets*, arXiv:1406.2261, 1-9.
- MENGHINI C., PALMAS C. 2022, *Il refettorio e la cucina di San Galgano (Chiusdino - SI): analisi sui materiali della campagna di scavo 2021*, in M. MILANESE (ed.), IX Congresso Nazionale di Archeologia Medievale (Alghero 2022), Sesto Fiorentino (FI), All’Insegna del Giglio, 158-163.
- YANG L., ZHANG Z., SONG Y., HONG S., XU R., ZHAO Y., ZHANG W., CUI B., YANG M.H. 2023, *Diffusion Models: A comprehensive survey of methods and applications*, «ACM Computing Surveys», 56, 4, 105, 1-51.

ABSTRACT

Artificial Intelligence (AI) has become an integral part of daily life, with applications expanding rapidly. This article focuses on AI's ability to generate images for archeological purposes, delving into the technical aspects of image generation using deep learning and machine learning technologies. Midjourney and Playground AI, two popular web applications for image generation, are examined in detail, highlighting their functionality and interface differences. The study applies these tools to reconstruct scenarios from the Cistercian Abbey of San Galgano, assessing their accuracy and limitations. While AI-generated reconstructions offer rapid and cost-effective visualization, they often lack precision, especially in incorporating specific archaeological details. The article underscores the need for improved databases containing detailed architectural and archaeological elements to enhance AI-generated reconstructions' accuracy and relevance.