

MATERIALITÀ, INCANTAMENTO E BIAS. VERSO UN USO CONSAPEVOLE DELL'IA IN ARCHEOLOGIA

“technology at present is covert philosophy;
the point is to make it openly philosophical”
(AGRE 1997, 240)

“la falsificazione con cui ciò che è eterogeneo
e incalcolabile viene reso uguale, simile e calcolabile”
(NIETZSCHE 1975, 182)

1. INTRODUZIONE

Dopo le riflessioni sul lascito del progetto ArchAIDE (ANICHINI, GATTIGLIA 2022) e la lettura postfenomenologica delle applicazioni algoritmiche in campo archeologico (GATTIGLIA 2023), vorrei indagare se attraverso l’Intelligenza Artificiale (IA), l’archeologia sia entrata in un’era post-digitale (CASCONI 2000), ovvero se il presente archeologico, ormai caratterizzato da un’abbondanza computazionale, sia pervaso da tecnologie digitali anticipatorie e di calcolo contestuale che utilizzano un’intelligente diffusione di elaborazione computazionale per creare una rete di risorse di calcolo incorporate nel mondo materiale, per poi esplorarne le conseguenze sull’archeologia stessa. Partendo dall’idea di IA come megamacchina vorrei comprendere il paesaggio computazionale creato dall’apparato digitale e materiale e dalle pratiche di classificazione dei dati di addestramento, fino ad analizzare gli aspetti etici dell’IA e il suo uso come strumento per l’archeologia e in particolare per la narrazione archeologica. In particolare, mi interessa riflettere sugli aspetti di incantamento dell’IA: oracolari (RONCAGLIA 2023) o di determinismo incantato (CAMPOLO, CRAWFORD 2020) e sulle sue ricadute in ambito archeologico.

2. L’INTELLIGENZA ARTIFICIALE

Intelligenza Artificiale è un termine coniato nel 1956 da McCarthy (MC CARTHY *et al.* 2006) per indicare un campo interdisciplinare nato con le teorizzazioni di Alan Turing (1950). In 70 anni, l’IA ha passato cicli espansivi, come quello attuale, caratterizzati da grande fiducia nel raggiungimento degli obiettivi ed elevati finanziamenti, e depressivi, noti come inverni dell’IA, caratterizzati da sfiducia legata a un deludente sviluppo tecnologico e diminuzione dei finanziamenti. In generale, l’IA si occupa dello studio e dello sviluppo di sistemi e tecnologie in grado di emulare (NATALE 2022) alcune delle capacità umane, come il ragionamento, la percezione, l’apprendimento, il linguaggio naturale e

il riconoscimento di modelli, al fine di automatizzare compiti complessi e prendere decisioni autonome, per le quali sono necessarie grandi quantità di dati (*Big Data*) da cui apprendere, significative risorse computazionali e hardware specializzato come le unità di elaborazione grafica (GPU) e tensoriale (TPU). L'IA comprende diverse aree di ricerca e applicazione, tra cui Machine Learning (ML), che rappresenta il campo più ampio, Reti neurali artificiali (ANN), Deep Learning (DL) e IA generativa (GenAI), che sono sottoinsiemi di sottoinsiemi.

2.1 Machine Learning

Il ML comprende algoritmi che permettono di riconoscere modelli nei dati e fare previsioni. Nel ML, i dati raccolti vengono pre-processati (puliti e normalizzati) per portare tutte le *features* (le variabili che descrivono le entità del dataset), su una scala comune, per poi selezionare le più adatte a descrivere il fenomeno in esame. La selezione riduce la complessità del modello, migliora le prestazioni e diminuisce la varianza. Una bassa varianza non cattura tutte le caratteristiche nei dati di addestramento (*underfitting*), mentre una varianza elevata non generalizza correttamente su dati nuovi, cioè si adatta strettamente ai dati di addestramento, a causa della loro eccessiva specificità (*overfitting*). I modelli più comuni di ML sono la *regressione lineare*, usata per predire una variabile continua basandosi su una o più variabili indipendenti; la *regressione logistica*, che predice la probabilità che un'istanza appartenga a una determinata classe, in genere una classificazione binaria; il *Random Forest*, basato su alberi decisionali (*decision tree*) addestrati su campioni di dati casuali e usato per classificazione e regressione; e la *Support Vector Machine* (SVM), che crea uno o più iperpiani nello spazio delle *features* per separare i dati in classi. Questi sono modelli di apprendimento supervisionato (*supervised learning*), la cui scelta dipende da tipo e quantità di dati e dalla complessità del fenomeno da analizzare. L'addestramento avviene con un set di dati che insegnano al modello a relazionare i dati di input con le rispettive etichette di output, attraverso un processo di ottimizzazione che cerca di minimizzare una funzione di costo. Dopo l'addestramento, il modello viene valutato utilizzando dati di test indipendenti (*test dataset*), per stimare le prestazioni del modello su dati sconosciuti e valutare se sia in grado di generalizzare. In caso contrario, è possibile raffinarlo ulteriormente regolando i parametri, aggiungendo ulteriori *features* o utilizzando tecniche che combinano le previsioni di più modelli (*ensemble learning*).

2.2 Reti neurali artificiali (Artificial Neural Network)

Le ANN sono sistemi di algoritmi che si ispirano agli studi sulla struttura del cervello umano sviluppati negli anni '40 da McCulloch e Pitts (RONCAGLIA 2023, 81) e sul concetto di percettore (*perceptron*) introdotto da Rosenblatt, nel 1958 e ripreso da Minsky e Papert nel 1969 (MINSKY, PAPERT 2017). Le ANN sono composte da unità computazionali di base, chiamate neuroni

artificiali, caratterizzate da funzioni di attivazione, collegate tra loro da connessioni pesate e organizzate in uno strato di input, almeno uno strato intermedio o nascosto, che elabora i dati, e uno strato di output che produce il risultato. I neuroni artificiali possono accettare qualsiasi valore reale di input, non solo valori binari, assegnare un peso diverso ai valori di input e definire un valore di soglia che ne determina (o meno) l'attivazione. Il valore di soglia può essere fisso o variabile in base all'input ricevuto da altri neuroni. La variazione del valore di soglia e dei pesi dei vari input è ciò che permette al neurone di imparare. Con un valore di soglia definito arbitrariamente si crea un modello deterministico. Inserendo una funzione di attivazione probabilistica, si ottiene un modello probabilistico in cui l'attivazione del neurone avviene quando la probabilità di attivazione è superiore al 50% (RONAGLIA 2023, 83). Così facendo si introduce non linearità nella rete e si migliora il processo di addestramento, consentendole di imparare relazioni complesse. L'addestramento può essere supervisionato, con dati etichettati (*labelled*), non supervisionato (*unsupervised learning*) con dati non etichettati, semi-supervisionato (*semi-supervised learning*), con dati misti o per rinforzo (*reinforcement learning*) basato su premi e punizioni. Durante l'addestramento supervisionato viene calcolata una misura dell'errore tra output previsto e output desiderato. Questo errore viene, poi, propagato all'indietro nella rete utilizzando un algoritmo di retropropagazione (*backpropagation*), che calcola i gradienti dell'errore rispetto ai pesi della rete neurale e aggiorna i pesi in modo da ridurre gradualmente l'errore. Questo processo viene iterato modificando in maniera dinamica i pesi all'interno della rete tramite feedback di rinforzo o indebolimento fino a quando la rete non converge su una soluzione che generalizza bene su dati sconosciuti. Il feedback può essere completamente automatico, come nelle reti generative, o prevedere l'intervento umano (*human-in-the-loop*).

2.3 Reti neurali artificiali profonde (Deep Learning)

Con DL ci si riferisce alle reti neurali artificiali profonde (*deep neural networks*), cioè composte da molteplici strati nascosti. I modelli di DL non necessitano di un'ingegnerizzazione manuale delle *features*, sono particolarmente adatti quando si hanno dati complessi e non facilmente rappresentabili da algoritmi di ML e sono capaci di apprendere automaticamente rappresentazioni gerarchiche dei dati. In pratica, i primi strati della rete tendono a estrarre le caratteristiche di basso livello, ad esempio, bordi, linee e colori delle immagini, mentre i livelli più profondi combinano queste caratteristiche per formare rappresentazioni sempre più complesse, di alto livello, come oggetti o concetti. Il livello degli strati nascosti, data l'autonomia del sistema di modificare pesi e valori, risulta particolarmente opaco e spesso viene definito come una *black box*, una scatola nera che non permette di capire gli stati dei singoli neuroni (ANICHINI, GATTIGLIA 2022). Anche nel DL l'*overfitting* è

comune. Per superarlo si adottano tecniche come la riduzione del numero di parametri del modello, l'aggiunta di *dropout* (caduta casuale di alcuni nodi durante l'addestramento), la *cross-validation* (ripetizione dell'addestramento e della valutazione su diverse porzioni del dataset), l'*early stopping* (interruzione dell'addestramento quando l'errore sui dati di validazione inizia ad aumentare), la *data augmentation* (generazione di dati sintetici di addestramento attraverso la manipolazione dei dati esistenti in modo da creare variazioni realistiche dei campioni originali), e il *transfer learning* che consiste nel partire da un modello, pre-addestrato su un set di dati ampio e generale (come ImageNet), per poi perfezionarlo su un set di dati specifico, sfruttando le competenze precedentemente apprese. Queste ultime due non sono state concepite specificamente per superare l'*overfitting*, ma risultano spesso usate per questo scopo. Esistono diverse tipologie di reti neurali profonde. Le reti neurali *feedforward* (FNN), ad esempio, sono utilizzate per la classificazione e la regressione e l'informazione fluisce in una sola direzione, dallo strato di input a quello di output, senza loop. Nei modelli ricorrenti (*Recurrent Neural Network RNN*), invece, esiste un feedback ciclico, che consente di trattare dati sequenziali e di conservare memoria a lungo termine, per cui sono applicate alle serie temporali e alla generazione di testo.

Le reti *Long Short-Term Memory* (LSTM) sono RNN che mantengono e dimenticano informazioni per periodi di tempo più lunghi, mentre le reti *Gated Recurrent Unit* (GRU) sono RNN con un'architettura più semplice e un ridotto numero di parametri rispetto alle LSTM. Le reti Neurali Convoluzionali (CNN) utilizzano layer di convoluzione che, applicando un filtro (*kernel*) su piccole regioni sovrapposte dell'immagine, estraggono le caratteristiche dall'input, seguiti da layer di *pooling* per ridurre la dimensione spaziale (larghezza e altezza) dell'output e la complessità computazionale del modello, riducendo così il rischio di *overfitting*. Le CNN sono usate per l'analisi di immagini e riconoscimento di pattern al loro interno. Le *Generative Adversarial Networks* (GAN) sono composte da due reti in competizione: il generatore crea dati sintetici mentre il discriminatore cerca di distinguere tra dati reali e sintetici, con lo scopo di allenare il generatore a produrre dati sempre più realistici. Gli *Autoencoder* permettono la riduzione della dimensionalità. Sono costituiti da un'*encoder* che trasforma l'input in uno spazio latente di dimensioni ridotte e da un *decoder* che ricostruisce l'input dall'*embedding* latente. Le Reti Neurali Trasformative (*Transformers*) utilizzano meccanismi di attenzione per elaborare sequenze di token e hanno rivoluzionato il campo del *Natural Language Processing* (NLP).

2.4 AI generativa

La GenAI viene usata per la generazione di testi, immagini e video. Nei modelli testuali si parte dalla tokenizzazione di un corpus di testi, che vengono

suddivisi in piccole unità (*token*) che possono corrispondere a singole parole, morfemi, n-grammi (gruppi di n caratteri) o singoli caratteri. Attraverso una fase di *unsupervised learning*, la rete aggiusta i valori associati ai token e ai pesi dei collegamenti interni per predire il token successivo. Questa operazione genera il *Large Language Model* (LLM), nome con cui sono conosciuti questi sistemi, ovvero un modello di correlazioni statistico-probabilistiche tra i token, che associa i token a uno spazio astratto multidimensionale che esprime numericamente (*embedding*) i contesti d'uso e le relazioni del singolo token all'interno del corpus (*vettorializzazione*). Spesso, dopo una fase di *unsupervised learning*, il modello viene perfezionato attraverso *supervised* o *reinforcement learning*, oppure utilizzando il *self-supervised learning* o delle GAN. Nella fase generativa, il sistema genera risposte con un meccanismo definito *sequence-to-sequence*, in cui a partire da una serie di simboli in entrata si ottiene una sequenza di simboli in uscita. Qui interviene il *transformer*, che non esamina i token in maniera sequenziale, ma tiene conto del contesto attraverso un meccanismo di attenzione (VASWANI *et al.* 2017) che pesa i valori dei token tenendo conto dei valori degli altri token del contesto e aggiungendo altri valori (*Query*, *Value* e *Key*) al vettore dell'*embedding*. Questo metodo disambigua meglio le parole polisemiche, migliora la memoria semantica e il funzionamento con input lunghi (RONCAGLIA 2023, 89-92). In pratica, un encoder vettorializza il testo di input, mentre un decoder genera il testo di output a partire dalla rappresentazione dell'encoder. In definitiva, i testi vengono prodotti in forma predittiva. Encoder e decoder possono essere (o meno) abbinati.

La generazione di immagini, chiamata *text-to-image*, *Text-Guided Diffusion Models*, o *conditional image generation*, funziona attribuendo valori numerici a ogni pixel e creando relazioni matematiche su cui il modello impara a distinguere le immagini. Questo sistema usa una GAN nella quale il generatore produce le immagini partendo dal testo, mentre il discriminatore valuta se le immagini sono realistiche. La generazione di immagini avviene tramite diffusione (*diffusion*). A partire da un'immagine nitida viene aggiunto del rumore che perturba progressivamente l'immagine fino a farla diventare puro rumore (diffusione in avanti). Successivamente, il rumore viene eliminato fino a rendere nuovamente nitida l'immagine (diffusione all'indietro). In pratica, il generatore produce immagini casuali o rumorose e man mano che viene addestrato, impara a generare immagini sempre più realistiche. Il discriminatore funge da critico ed è responsabile di distinguere tra immagini reali e sintetiche e viene addestrato per diventare sempre più abile nel distinguere tra le due categorie. Questi due processi di addestramento si alternano in modo iterativo fino a quando il generatore non è in grado di produrre immagini che il discriminatore non riesce più a distinguere da quelle reali. Alla fine, un prompt testuale umano guida la diffusione che produce un'immagine a partire da un insieme di parole e rumore, integrando la componente visuale e quella testuale.

3. L'IA IN ARCHEOLOGIA

A dimostrazione dell'interesse che questo campo riveste per la nostra disciplina, nell'ultimo decennio, sono state prodotte 16.400 pubblicazioni su IA e archeologia, con numeri duplicati su base annuale¹. Finora, l'archeologia ha beneficiato dell'IA per analizzare grandi quantità di dati e svolgere attività altamente specializzate e dispendiose in termini di tempo, ma con la diffusione di modelli generativi come ChatGPT, DALL-E, Midjourney, Claude, etc. è facile prevederne un utilizzo sempre più ampio anche da parte di non esperti di dominio. L'IA è stata applicata soprattutto a compiti di classificazione e identificazione nello studio del paesaggio, degli oggetti, dei testi e recentemente alla robotica. Per l'identificazione di siti archeologici, ad esempio, sono state proposte CNN addestrate su dati satellitari (ORENGO *et al.* 2020), dati LiDAR con *transfer learning* (VERSCHOOF-VAN DER VAART, LAMBERS 2022) o cartografia storica (GARCIA-MOLSOSA *et al.* 2021) e anche modelli *Random Forest* con dati LiDAR (GUYOT *et al.* 2018). CNN sono state usate anche per la segmentazione semantica delle anomalie da GPR (KÜÇÜKDEMIRCI, SARRIS 2020), mentre CASINI *et al.* (2023) hanno usato un *trasformer* con immagini annotate da esperti (*human-in-the-loop*), senza transfer learning. A livello di manufatti, possiamo citare esempi di CNN con *transfer learning* per riconoscere la ceramica (GUALANDI *et al.* 2021), di GAN per il restauro virtuale della ceramica (NAVARRO *et al.* 2020) e delle monete (ZACHARIOU *et al.* 2020), mentre TSIGKAS *et al.* (2020) hanno usato una CCN pre-addestrata per l'identificazione automatica dell'arte rupestre.

La classificazione automatizzata di elementi architettonici da modelli 3D è stata, ad esempio, indagata attraverso una Dynamic Graph CNN da PIERDICCA *et al.* (2020). CNN, con *transfer learning*, sono state applicate allo studio dei resti scheletrici umani (BEWES *et al.* 2019), CNN e SVM (BYEON *et al.* (2019) e diversi tipi di CNN, alcune pre-addestrate (DOMINGUEZ RODRIGO *et al.* 2020), sono state applicate per il riconoscimento di segni di macellazione e di taglio su reperti osteologici animali. Grove e Blinkhorn (2020), invece, hanno applicato una ANN all'industria litica e RESLER *et al.* (2021) hanno sviluppato una CCN per identificare manufatti dal Paleolitico alla tarda età islamica. Applicazioni NLP sono molto sviluppate in archeologia (TENZER *et al.* 2024): ad esempio, CNN sono state applicate alla ricomposizione di frammenti di ostraca (OSTER-TAG, BEURTON-AIMAR 2020) o per tradurre tavolette cuneiformi (GUTHERZ *et al.* 2023), mentre un *trasformer* è stato utilizzato per la ricostruzione delle epigrafi greche (ASSAEL *et al.* 2022). La robotica è un campo ancora marginale, ma i recenti progetti RePAIR (<https://www.repairproject.eu/>) e il neo finanziato progetto AUTOMATA (<https://automata-eccch.eu/>), coordinato da Università

¹ Risultati della ricerca “Archaeology AND AI” escludendo “medicine and health” su Google Scholar effettuata il 5 ottobre 2024.

di Pisa, indicano che questo sarà un nuovo filone di studi applicativi. L'uso della GenAI sembra per ora limitato alla ricostruzione immersiva del paesaggio (COBB 2023) o del paleoambiente a partire dai dati pollinici (PAPERINI *et al.* 2023) o da dati entomologici e ambientali (HUMLAB n.d.).

4. COSE CHE NON VEDIAMO E COSE A CUI NON PRESTIAMO ATTENZIONE

Dietro l'IA non ci sono solo aspetti algoritmici (GATTIGLIA 2022, HUGGETT 2021), ma anche materiali. L'IA è una megamacchina (MUNFORD 1969) rappresentata da una rete intricata (HODDER 2012, 2023) composta da un insieme di approcci tecnologici che dipendono da infrastrutture industriali, data center, cavi sottomarini, catene di approvvigionamento e lavoro umano, dispositivi personali e loro componenti grezzi, segnali di trasmissione aerei, dataset e continui cicli di calcolo, etc., di cui è possibile vedere la materialità. Una materialità estrattiva, legata allo sfruttamento di risorse energetiche e minerali del pianeta, di manodopera a basso costo e dati, con una forte impronta carbonica. L'esecuzione di un solo (grande) modello di NLP, ad esempio, può produrre tanta CO₂, quanto 125 voli andata e ritorno Pechino/New York (STRUABELL *et al.* 2019), un dato impressionante se pensiamo che la potenza di calcolo per addestrare un singolo modello da parte di OpenAI aumenta di 10 volte ogni anno (OPENAI 2018). Una materialità terraformativa (GHOSH 2022) che crea un paesaggio computazionale (CRAWFORD 2023, 38) globale, archeologicamente indagabile. Una megamacchina che produce diseguaglianza (*digital e data divide*) e conserva una forte impronta coloniale.

Il termine *master-slave*, che indica sistemi di gestione, in cui il computer master coordina l'attività degli *slave* eventualmente sacrificandoli per proteggere se stesso, ad esempio, persiste nelle infrastrutture di Google e solo recentemente è stato eliminato, perché giudicato offensivo, da Python (CRAWFORD 2018, 90). Espressioni come *data mining* e metafore come *data is the new oil*, riportano alla retorica coloniale per trasformare semanticamente i dati, attraverso una visione rappresentativa (LIONELLI 2018, 93), in una risorsa naturale in attesa di essere scoperta, sfruttata e consumata (STARK, HOFFMAN 2019). Se i dati sono petrolio, l'apprendimento automatico diventa il processo di raffinazione, che, in linea con una visione neoliberista, li trasforma in capitale, non in un bene comune non-rivale, come avevamo auspicato (GATTIGLIA 2015). In questo modo, un dataset può rimanere aperto, ma il suo metavalore, ovvero il modello che viene creato, diventa di proprietà privata. La stessa creazione di un *training dataset* non è un'operazione neutra, né puramente tecnica, bensì la cristallizzazione di un mondo complesso e vario in classificazioni discrete (come, ad esempio, quelle archeologicamente discutibili di *site/off-site*), un processo che richiede scelte intrinsecamente politiche, culturali e sociali e modella i confini epistemici che governano il modo in cui opera l'IA, definendone la visione del

mondo e la sua azione sul mondo (CRAWFORD 2023, 135). La scarsità di dati a disposizione in archeologia porta spesso a usare metodi di *transfer learning*, in particolare ImageNet, per pre-allenare una rete a compiti di classificazione. ImageNet è un archivio di oltre quattordici milioni di immagini organizzate in più di ventimila categorie per essere «un’ontologia generale delle immagini» utile a «fornire addestramento critico e parametri di riferimento» agli algoritmi di oggetti e immagini (DENG *et al.* 2009). Ma non è uno strumento neutro. ImageNet è stata realizzata senza alcuna preoccupazione di carattere etico, grazie al lavoro di etichettatura di *crowdworkers* sottopagati (GERSHORN 2017) ed è basata su una struttura semantica importata da WordNet, un database di classificazioni di parole sviluppato nel 1985 dalla Princeton University, diventato una fonte essenziale per la linguistica computazionale (FELDBAUM 1998, xviii), a sua volta modellato a partire dal Brown Corpus, una lista di un milione di parole proveniente da giornali e da uno spoglio casuale di libri, compilata negli anni Sessanta e figlia di un ben determinato spazio culturale.

ImageNet considera solo i sostanzivi, sulla base dell’idea che questi siano rappresentabili per mezzo di immagini e che siano sufficienti per l’addestramento automatico degli oggetti, attraverso una tassonomia gerarchica che porta a semplificare la complessità del reale. Dalle nove categorie di alto livello (pianta, formazione geologica, oggetto naturale, sport, manufatto, fungo, persona, animale, miscellaneo), si scende fino ai livelli inferiori, dove la base linguistica della tassonomia crea delle assurdità dal punto di vista oggettuale. Sotto mela, ad esempio, si trovano immagini di burro di mela, fagottini di mela, succo di mela, ma anche di geranio odoroso (in inglese *apple geranium*) e sotto *hot* (caldo) sono elencate parole che vanno dalle auto modificate (*hot rods*), alle salse e cibi piccanti (*hot sauce, hot pots*), alla vasca a idromassaggio (*hot tubs*), alla mongolfiera (*hot-air balloons*), fino alle borse dell’acqua calda (*hot water bottles*) (CRAWFORD, PAGLEN 2020)! Questi aspetti etici e sistemici si ripropongono per la GenAI. Da un lato i modelli vengono addestrati a partire da milioni di immagini senza che venga riconosciuto il diritto d’autore, dall’altra non siamo a conoscenza dei corpora utilizzati per l’addestramento (BIRHANE *et al.* 2021), una mancanza di trasparenza che ci impedisce di capire quali bias vengano introdotti nel modello (BENDER *et al.* 2021).

5. DATI, NARRAZIONI, ETICA

Il sempre maggior utilizzo della GenAI in archeologia, a cui assisteremo nei prossimi anni, comporterà un passaggio dall’automazione delle pratiche archeologiche, come la classificazione automatica di reperti o l’identificazione di ‘siti’, all’interpretazione e alla creazione di immagini, filmati ricostruttivi o testi. Demandare più o meno parzialmente queste attività all’IA deve necessariamente condurre a un approccio critico e consapevole. D’altra parte, gli

aspetti oracolari del determinismo incantato, nei suoi gemelli metafisici utopico e distopico (IA come soluzione a ogni problema o come fonte di ogni pericolo), precludono la discussione informata, l'esame critico o il possibile rifiuto. È da qui, invece, che dobbiamo partire: dalla discussione informata tra archeologici, che analizzi gli aspetti materiali, classificatori ed etici dell'IA.

La materialità illustra gli aspetti soggettivi, politici, egemonici e autoritari (CRAWFORD 2023, 19) dell'apparato digitale e merita di essere indagata archeologicamente. In questo scenario, l'archeologia sarebbe attenta ai modi in cui la cultura si materializza, a partire dai giganteschi database, e a come sia co-creata da attori umani e dall'apparato tecnico. Per questo è, innanzitutto, necessario abbandonare una visione rappresentativa del dato archeologico, che vede i dati come elementi già esistenti, solo pronti a essere raccolti, per passare a una co-creativa (LUCAS 2023) o relazionale (LIONELLI 2018, 101). La visione rappresentativa non ci permette di comprendere né il dato, né il suo riuso, oscura le operazioni di potere ed evita la responsabilità delle conseguenze. I dati sono centrali nel processo di classificazione, che, come abbiamo visto, si configura come un atto di potere, di cui bisogna tenere in conto le interazioni, le teorie sociali e politiche che a esso soggiacciono e che da questo vengono supportate (BOWKER, STAR 1999, 319). Si tratta di un impulso colonizzatore proprio dell'IA che determina come il mondo viene misurato e definito e contemporaneamente ne nega la sua intrinseca politicità, oggettività, universalità e neutralità.

Quali rischi ci sono per l'archeologia? Il primo è quello di voler catturare le complessità semplificando eccessivamente ciò che è complesso in modo poterlo facilmente calcolare. Questo desiderio riduzionista produce un incantamento deterministico giacché assume le approssimazioni classificatorie come verità di base, attribuendo etichette fisse a una complessità fluida. Al contrario, «[i]nvece di cercare di costruire sistemi sempre nuovi in grado di raggruppare le espressioni in categorie leggibili dalla macchina, dovremmo mettere in discussione le origini delle categorie stesse, così come delle loro conseguenze sociali e politiche» (CRAWFORD 2023, 197). Ad esempio, l'IA che individua i siti parte da una classificazione sito- e antropo-centrica, che condiziona le successive interpretazioni. Anche il necessario uso dei Big Data per l'apprendimento automatico, per l'ovvio riuso di *legacy data*, comporta il rischio di reiterare idee e interpretazioni e mantenere schemi colonialistici. Con questo tipo di classificazioni e di dati, il pericolo di proporre, tramite la GenAI, ricostruzioni testuali o grafiche dannose e ricche di pregiudizi (coloniali) che «si possono radicare ed essere difficili da ribaltare» (CRAWFORD 2024, 128) risulta particolarmente alto, dal momento che questa è sempre più indirizzata a un utilizzo generalista e non solo accademico. È bene, però, tenere presente che le ricostruzioni testuali dell'IA non sono da intendersi come una vera e propria narrazione. Nelle pratiche digitali che abbiamo osservato assistiamo a un processo di accumulazione, che lavora aggiungendo un dato a un altro, simile allo storytelling, cioè alla controfigura della narrazione (HAN 2024, 68). Se

pensiamo all'accumulo dei token che creano un testo, il contesto di connessioni che istituisce il senso delle cose cede il passo «a un essere-uno-accanto-all'altro o a un esser-uno-dopo-l'altro di eventi svuotati di senso» (HAN 2024, 53). I dati procedono per addizione e accumulo; frammentano e non sono portatori di senso, mentre la creazione narrativa avviene per selezione e sintesi, e produce un continuum in cui il senso transita (HAN 2024, 10). L'IA emula forme di pensiero, ma non interpreta, cioè non ci permette di capire come gli umani guardassero le cose e agissero nel passato. Per fare questo è necessario sviluppare teorie, ossia delle forme di narrazione che sviluppino contesti concettuali in grado di rendere le cose intellegibili, ancor di più in questo momento di accelerazione digitale. Parafrasando Han (2024, 90), come archeologi dobbiamo avere «più coraggio per la filosofia, per la teoria, il che significa [avere] più coraggio per la prassi narrativa [...]» se vogliamo mantenere il controllo sul racconto archeologico. Tanto più perché la ricerca archeologica non si svolge nel vuoto, senza alcuna responsabilità per le idee che diffondono, al contrario ha una forte propensione pubblica e non può esimersi dalle questioni etiche. Dobbiamo, quindi, essere consapevoli di come, oltre a perpetrare narrazioni fortemente compromesse dal discorso coloniale, «le questioni più complesse del contesto, del condizionamento, delle relazioni e dei fattori disciplinari sono difficilmente conciliabili con gli attuali approcci disciplinari dell'informatica o con le ambizioni commerciali del settore tecnologico» (CRAWFORD 2023, 195) e rischiano di mantenere consolidati apparati di potere accademico e accrescere il *digital divide* tra la ricerca occidentale e quella del Sud globale. La privatizzazione del metavalore prodotto dai dati, ci deve spingere a riflettere senza pregiudizi sulla disseminazione aperta dei dati delle nostre ricerche o sulla necessità di assicurarsi un uso non commerciale degli stessi, per garantire un uso pubblico del metavalore prodotto.

6. ARCHEOLOGIA POST-DIGITALE?

Il concetto di post-digitale racchiude in sé sia l'idea di uno stato in cui la perturbazione causata dalla digitalizzazione è già avvenuta, in cui la tecnologia è normalizzata e non percepita come disturbante, sia il rifiuto dell'innovazione tecnologica e del tecno-positivismo, che associato alla mescolanza di analogico e digitale, digitale e mondo reale, diventa il motore dell'innovazione (CRAMER 2014). Un decennio fa, a questa stessa domanda Jeremy Huggett (2015) rispose negativamente. Considerando solamente la normalizzazione e la fine della perturbazione dovuta al digitale, le vedeva vicine (in alcuni settori), ma non ancora avvenute. A mio avviso, non sbagliheremmo a definire l'archeologia post-digitale, se tenessimo in considerazione le applicazioni digitali escludendo l'IA. Nessuno trova disturbanti GIS, modelli 3D, analisi spaziali, etc. In questi campi, il tecno-positivismo è superato e la commistione tra mondo reale, analogico e digitale è quotidiana. L'introduzione dell'IA ha, però, scompaginato le carte.

Gli aspetti perturbativi non sono ancora superati, troppo recente e ancora in evoluzione la sua introduzione, e in molti ricercatori gli aspetti oracolari del determinismo incantato sono ancora radicati. Per entrare in un periodo di normalizzazione, dobbiamo coltivare maggior disincanto e scetticismo verso l'IA, in modo da umanizzarla e incorporarla nel mondo reale. Forse, però, «[i]n un'era post-digitale, la domanda su se qualcosa sia digitale o meno non è più davvero importante, proprio come l'ubiquità della stampa, poco dopo Gutenberg, ha reso obsoleti tutti i dibattiti sulla 'rivoluzione della stampa'» (CRAMER 2014).

7. CONCLUSIONI

In conclusione, l'IA sta iniziando a plasmare la ricerca archeologica, anche grazie ai finanziamenti che riceve, ed è, quindi, necessario stare attenti a quali direzioni intraprendere, valutando, assieme ai benefici, i possibili costi finanziari, ambientali e umani legati alla creazione di modelli di IA. Le risorse necessarie per addestrare e ottimizzare i modelli rischiano di aumentare le disuguaglianze, mentre i bias incorporati nei dataset utilizzati meritano di essere attentamente esaminati. Dovrebbe essere dedicato un tempo significativo all'assemblaggio dei dataset, anziché assumere enormi quantità di dati da fonti comode o facilmente accessibili. Ciò comporta un attento e completo lavoro di documentazione che includa obiettivi, valori e motivazioni alla base della raccolta, selezione e assemblaggio dei dati e della creazione di un determinato modello, e di analisi dei potenziali utenti, ponendo una particolare attenzione verso coloro che potrebbero essere negativamente influenzati dagli errori o dall'abuso del modello (BENDER *et al.* 2023). Questi approcci richiedono tempo e sono più preziosi quando applicati all'inizio del processo piuttosto che come analisi *post-hoc* dei rischi. La ricerca dell'IA in campo archeologico, essendo interessata a dati profondamente umani e alla creazione di narrazioni con cui gli esseri umani interagiscono in modi immediati e vividi, dovrebbe mettere al centro le persone che rischiano di essere danneggiate dalla tecnologia risultante, con una visione ampia sui possibili modi in cui l'IA può influenzare le persone. Questo significa dedicare tempo per coinvolgere gli interessati all'inizio del processo di progettazione, per esplorare più percorsi possibili verso obiettivi a lungo termine, per porre attenzione a possibili scenari di *dual use* e, infine, per allocare sforzi di ricerca per la mitigazione dei danni in tali casi.*

* Questo articolo è stato realizzato all'interno del progetto FAIR FAIR – Future Artificial Intelligence Research finanziato dall'Unione Europea – NextGenerationEU, nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR).

GABRIELE GATTIGLIA

Laboratorio MAPPA, Dipartimento di Civiltà e Forme del Sapere
Università di Pisa
gabriele.gattiglia@unipi.it

BIBLIOGRAFIA

- AGRE P.E. 1997, *Computation and Human Experience*, Cambridge, Cambridge University Press.
- ANICHINI F., GATTIGLIA G. 2022, *Reflecting on artificial intelligence and archaeology: The ArchAIDE perspective*, «Post-Classical Archaeologies», 12, 69-86.
- ASSAEL Y., SOMMERSCHIELD T., SHILLINGFORD B., BORDBAR M., PAVLOPOULOS J., CHATZI-PANAGIOTOU M., ANDROUTSOPoulos I., PRAG J., DE FREITAS N. 2022, *Restoring and attributing ancient texts using deep neural networks*, «Nature», 603, 280-283.
- BENDER E.M., GEBRU T., McMILLAN-MAJOR A., SHMITCHELL M. 2021, *On the dangers of stochastic parrots: Can language models be too big?*, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, Association for Computing Machinery, New York, 610-623.
- BEWES J., LOW A., MORPHETT A., PATE F.D., HENNEBERG M. 2019, *Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls*, «Journal of Forensic and Legal Medicine», 62, 40-43.
- BIRHANE A., PRABHU V. U., KAHUMBWE E. 2021, *Multimodal datasets: Misogyny, pornography, and malignant stereotypes*, «arXiv», 2110.01963.
- BOWKER G.C., STAR S.L., *Sorting Things Out: Classification and its Consequences*, Cambridge, The MIT Press.
- BYEON W., DOMÍNGUEZ-RODRIGO M., ARAMPATZIS G., BAQUEDANO E., YRAVEDRA Y., MATÉ-GONZÁLEZ M.A., KOUOMOUTSAKOS P. 2019, *Automated identification and deep classification of cut marks on bones and its paleoanthropological implications*, «Journal of Computational Science», 32, 36-43.
- CAMPOLI A., CRAWFORD K. 2020, *Enchanted determinism. Power without responsibility in Artificial Intelligence*, «Engaging Science, Technology and Society», 6, 1-19.
- CASCONÉ K. 2000, *The aesthetics of failure: 'Post-Digital' tendencies in contemporary computer music*, «Computer Music Journal», 24, 4, 12-18.
- CASINI L., MARCHETTI N., MONTANUCCI A., ORR V., ROCSETTI M. 2023, *A human-AI collaboration workflow for archaeological sites detection*, «Scientific Reports», 13, 8699.
- COBB P. 2023, *Large language models and generative AI, Oh my!: Archaeology in the time of ChatGPT, Midjourney, and beyond*, «Advances in Archaeological Practice», 11, 363-369.
- CRAMER F. 2014, *What is 'Post-Digital'?*, «APRJA», 3, 1.
- CRAWFORD K. 2023, *Né intelligente, né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino.
- CRAWFORD K., PAGLEN T. 2020, *Excavating AI. The Politics of Images in Machine Learning Training Sets* (<https://excavating.ai/> [ultimo accesso 5 marzo 2024]).
- DENG J., DONG W., SOCHER R., LI L.-J., KAI LI, FEI-FEI LI 2009, *ImageNet: A large scale hierarchical image database*, in *2009 IEEE Conference on Computer Vision and Pattern Recognition (Miami 2009)*, 248-255.
- DOMÍNGUEZ-RODRIGO M., CIFUENTES-ALCOBENDAS G., JIMÉNEZ-GARCÍA B., ABELLÁN N., PIZARRO-MONZO M., ORGANISTA E., BAQUEDANO E. 2020, *Artificial intelligence provides greater accuracy in the classification of modern and ancient bone surface modifications*, «Scientific Reports», 10, 18862.
- FELDBAUM C. (ed.) 1998, *WorldNet: An Electronic Lexical Database*, Cambridge, The MIT Press.
- GARCIA-MOLSOSA A., ORENGO H.A., LAWRENCE D., PHILIP G., HOPPER K., PETRIE C.A. 2021, *Potential of deep learning segmentation for the extraction of archaeological features from historical map series*, «Archaeological Prospection», 28, 2, 187-199.
- GATTIGLIA G. 2015, *Think big about data: Archaeology and the Big Data challenge*, «Archäologische Informationen», 38, 113-124.

- GATTIGLIA G. 2022, *A postphenomenological perspective on digital and algorithmic archaeology*, «Archeologia e Calcolatori», 33.2, 319-334 (<https://doi.org/10.19282/ac.33.2.2022.17>).
- GERSHORN D. 2017, *The data that transformed AI research and possibly the world*, «Quartz», 26 July (<https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world> [ultimo accesso 5 marzo 2024]).
- GHOSH A. 2022, *La maledizione della noce moscata: parabole per un pianeta in crisi*, Milano, Neri Pozza.
- GUALANDI M.L., GATTIGLIA G., ANICHINI F. 2021, *An open system for collection and automatic recognition of pottery through neural network algorithms*, «Heritage», 4, 140-159.
- GROVE M., BLINKHORN J. 2020, *Neural networks differentiate between Middle and Later Stone Age lithic assemblages in eastern Africa*, «PLoS ONE», 15, 8, e0237528.
- GUYOT A., HUBERT-MOY L., LORHO T. 2018, *Detecting Neolithic burial mounds from LiDAR-derived elevation data using a multi-scale approach and Machine Learning techniques*, «Remote Sensing», 10, 2, 225.
- GUTHERZ G., GORDIN S., SÁENZ L., LEVY O., BERANT J. 2023, *Translating Akkadian to English with neural machine translation*, «PNAS Nexus», 2, 5, pgad096.
- HODDER I. 2012, *Entangled: An Archaeology of the Relationships between Humans and Things*, London, Wiley-Blackwell.
- HODDER I. 2023, *Entangled: A New Archaeology of the Relationships between Humans and Things*, London, Wiley-Blackwell.
- HUGGETT J. 2015, *A manifesto for an introspective digital archaeology*, «Open Archaeology», 1(1), 86-95 (<https://doi.org/10.1515/opar-2015-0002>).
- HUGGETT J. 2021, *Algorithmic agency and autonomy in archaeological practice*, «Open Archaeology», 7, 417-434.
- HUMLAB n.d., *SEAD Conservation Paleobiology* (<https://showcase.humlab.umu.se/infravisead> [ultimo accesso 27 marzo 2024]).
- KÜÇÜKDEMIRCI M., SARRIS A. 2022, *GPR data processing and interpretation based on artificial intelligence approaches: Future perspectives for archaeological prospection*, «Remote Sensing», 14, 3377.
- LUCAS G. 2023, *Archeological Situations. Archeological Theory from the Inside Out*, London, Routledge.
- LIONELLI S. 2018, *La ricerca scientifica nell'era dei Big Data*, Milano, Meltemi.
- MCCARTHY J., MINSKY M., ROCHESTER N., SHANNON C.E. 2006, *A proposal for Dartmouth Summer Research Project on Artificial Intelligence*, «AI Magazine», 27, 4, 12.
- MINSKY M., PAPERT S. 2017, *Perceptrons: An Introduction to Computational Geometry*, Cambridge, The MIT Press.
- MUNFORD L. 1969, *Il mito della macchina*, Milano, Il Saggiatore.
- NATALE S. 2022, *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, Torino, Einaudi.
- NAVARRO P., CINTAS C., LUCENA M., FUERTES J.M., SEGURA R., DELRIEUX C., GONZÁLEZ-JOSÉ R. 2022, *Reconstruction of Iberian ceramic potteries using generative adversarial networks*, «Scientific Reports», 12, 10644.
- NIETZSCHE F. 1975, *Frammenti postumi (1884-1885)*, vol. 7, tomo 3, Milano, Adelphi.
- OPENAI 2018, *AI and Compute*, «OpenAI» (<https://openai.com/research/ai-and-compute> [ultimo accesso 1 marzo 2024]).
- ORENGO H.A., CONESA F.C., GARCIA-MOLSOSA A., LOBO A., GREEN A.S., MADELLA M., PETRIE C.A. 2020, *Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data*, «PNAS», 117, 31, 18240-18250.

- OSTERTAG C., BEURTON-AIMAR M. 2020, *Matching ostraca fragments using a Siamese neural network*, «Pattern Recognition Letters», 131, 336-340.
- PAPERINI E., SCIUTO C., GATTIGLIA G. 2023, *Paleoenvironmental reconstruction from archaeological pollen data: Could AI tell stories about what the landscape looked like?*, in *Conference Abstract of 43rd Conference of the Association for Environmental Archaeology (AEA)* (Tarragona 2023).
- PIERDICCA R., PAOLANTI M., MATRONE F., MARTINI M., MORBIDONI C., MALINVERNI E.S., FRONTONI E., LINGUA A.M. 2020, *Point cloud semantic segmentation using a deep learning framework for Cultural Heritage*, «Remote Sensing», 12, 6, 1005.
- RESLER A., YESHURUN R., NATALIO F., GIRYES, R. 2021, *A deep-learning model for predictive archaeology and archaeological community detection*, «Humanities and Social Sciences Communications», 8, 1, 1-10.
- RONCAGLIA G. 2023, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*, Roma-Bari, Laterza.
- STARK L., HOFFMANN A.L., *Data is the new what? Popular metaphor and professional ethics in emerging data culture*, «Journal of Cultural Analytics», 4, 1.
- STRUPELL E., GANESH A., MCCALLUM A. 2019, *Energy and policy considerations for Deep Learning in NLP*, «ArXiv», 1906.02243.
- TENZER M., PISTILLI G., BRANDSEN A., SHENFIELD A. 2024, *Debating AI in Archaeology: Applications, implications, and ethical considerations*, «Internet Archaeology», 67 (<https://doi.org/10.11141/ia.67.8>).
- TURING A.M. 1950, *Computing machinery and intelligence*, «Mind», 59, 433-460.
- TSIGKAS G., SFIKAS G., PASIALIS A., VLACHOPOULOS A., NIKOU C. 2020, *Markerless detection of ancient rock carvings in the wild: Rock art in Vathy, Astypalaia*, «Pattern Recognition Letters», 135, 337-345.
- VASWANI A., SHAZEEB N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A.N., POLOSUKHIN I. 2017, *Attention is all you need*, in I. GUYON, U. VON LUXBURG, S. BENGIO, H. WALLECH, R. FERGUS, S. VISHWANATHAN, R. GARNETT (eds.), *Proceedings in Advances in Neural Information Processing Systems (NeurIPS)*, 30, NIPS Foundation, 5998-6008.
- VERSCHOOF-VAN DER VAART W.B., LAMBERS K. 2022, *Applying automated object detection in archaeological practice: A case study from the southern Netherlands*, «Archaeological Prospection», 29, 1, 15-31.
- ZACHARIOU M., DIMITRIOU N., ARANDJELOVIĆ O. 2020, *Visual reconstruction of ancient coins using cycle-consistent generative adversarial networks*, «Science», 2, 3, 52.

ABSTRACT

The rise of AI in archaeology merges algorithmic and material aspects, forming a complex mega-machine reliant on industrial infrastructure, labour, and resources. This materiality shapes a global landscape marked by inequalities and colonial legacies. Creating training datasets involves political and cultural choices, shaping AI's worldview and biases. AI's increasing role shifts archaeology from automation to interpretation, urging critical awareness to avoid perpetuating biases. Archaeologists must engage in philosophical reflection and narrative praxis to retain control over archaeological discourse, mindful of ethical responsibilities and potential harm. Ultimately, AI's role in archaeology demands careful consideration of its impacts, including financial, environmental, and human costs. AI research in archaeology should prioritise people, embracing a broad perspective on how AI can affect human lives while remaining vigilant against potential risks and ethical considerations.