



A JOURNAL OF THE
SOCIAL IMAGINARY

Immaginario *Black Box*: prospettive mitiche e critiche tra innovazione bellica ed esplorazione artistica

Lorenzo Aimo

lorenzo.aimo@unimi.it

DIPAFILO | *Università degli Studi di Milano*

DAR | *Alma Mater Studiorum Università di Bologna*



Luca Brunet

luca.brunet@unimi.it

DIPAFILO | *Università degli Studi di Milano*

DFE | *Università di Torino*

Abstract

Black Box Imaginary: Mythical and Critical Perspectives Between Military Innovation and Artistic Exploration

Within the imaginaries of AI, automation occupies a central position and receives varying nuanced connotations. In both technical-scientific and popular discourses, this notion takes on the rhetorical and visually metaphorical form of the black box, implying the inability to fully visualize and understand what happens inside AI systems due to a complex and intricate layering of neural networks and algorithms. The rhetoric of the black box shapes both the justification and the critique of the use of AI, by showing or obscuring implications regarding accountability, values, and ethics. The present contribution explores how the myth of the black box and various corresponding imaginaries manifest across diverse case studies in two distinct domains, revealing their political, social, and semiotic implications. Through the analysis of the representation of AI by military websites (such as Palantir and Gospel) and of artworks on AI (like Paglen's 2019 and Akten's 2019 works), this paper endeavors to showcase how contrasting perspectives – mythical and critical – characterize the metaphor of the black box within the imaginary of AI.

Keywords

Artificial Intelligence | Imaginary | Black Box | Myth | Opacity



1. Introduzione¹

Volgendo uno sguardo al passato, guardando a uno degli esempi più famosi di sistemi (apparentemente) intelligenti – tuttora figura presente nel contemporaneo – è possibile comprendere il percorso parallelo e congiunto che caratterizza l'Intelligenza Artificiale (IA) e la figura della *black box*. Progettato nel 1770 da Wolfgang von Kempelen, il Turco Meccanico rappresenta un'icona (Alexander, 2008) la cui immagine e il derivante significato riadattato vengono rievocati dal programma omonimo di Amazon, quello dei Mechanical Turk, ossia l'insieme di lavoratori anonimi che contribuiscono al processo di costruzione di una moderna IA. Nella "congegnata illusione" (Crawford, 2021: 79) di Von Kempelen, veniva offerta alla vista del pubblico una figura meccanica antropomorfa a grandezza naturale in grado di giocare a scacchi; tuttavia, all'interno del mobile a cui sedeva l'automa, letteralmente inscatolato in esso, si nascondeva un essere umano che, pilotando l'automa, risultava essere il vero artefice della performance nonchè dell'inganno². Rimanendo celata all'occhio esterno, la persona inscatolata, coincidendo con il funzionamento della macchina, garantiva la simulazione dell'intelligenza umana da parte di un automa.

Oltre a ricordare quanto l'umano sia costitutivo della costruzione e del funzionamento dei sistemi di IA attuali, l'esempio storico risulta evocativo di una dinamica che, nella dicotomia tra esterno-visibile e interno-invisibile, ha assunto diverse forme e livelli nell'ambito dei sistemi di IA: ciò che viene metaforicamente indicato come *black box*.

All'interno dello spettro di immaginari socio-tecnici (Jasanoff, 2015)³ che circondano l'IA, la figura metaforica della *black box* appare cruciale, in quanto si inserisce in un sostrato narrativo strettamente legato all'idea cardine e onnipresente dell'automazione – da cui derivano le figure narrative caratteristiche di questi



¹ Il presente contributo è stato pensato nella sua interezza da entrambi gli autori. Nello specifico, introduzione e conclusione sono state elaborate e scritte in collaborazione; Lorenzo Aimo è responsabile dei paragrafi 3 e 5; Luca Brunet è responsabile dei paragrafi 2 e 4, sebbene durante la stesura si sia sempre tenuto un confronto tra gli autori.

² Risulta interessante per uno studio sull'immaginario delle IA notare come uno dei primi casi di automa intelligente consistesse in realtà in un vero e proprio inganno, basato sull'occultamento di un meccanismo all'interno di un artefatto, e sulla resa visibile di una superficie significativa che viene usata per mentire; si tratta cioè di un meccanismo fondamentalmente semiotico (cfr. Paolucci, 2021; Eco, 1975). Sul rapporto tra IA e inganno si veda Natale (2022).

³ Jasanoff definisce tali immaginari come "collectively held and performed visions of desirable futures (or of resistance against the undesirable)" (2015: 19) che vengono sorretti e sorreggono il progresso scientifico-tecnologico. La storia degli sviluppi dell'Intelligenza Artificiale, caratterizzata da fasi alterne, paure ed entusiasmi, dimostra come tale dicotomia sugli esiti del futuro ne abbia condizionato tanto la progettazione e costruzione effettiva - sia in senso economico e materiale che epistemologico e concettuale - quanto l'utilizzo e la percezione comune.



immaginari⁴. In ragione della forza concettuale e cognitiva caratteristica delle metafore (Lakoff e Johnson, 1981), la metafora della *black box* adempie al compito di significare un limite, quello tra visibile e conoscibile, che scaturisce dalla tecnologia stessa o dal fenomeno contestualmente osservato, il cui meccanismo di funzionamento rimane sia celato allo sguardo umano, sia incomprensibile.

La metafora della *black box* non appartiene infatti unicamente all'ambito discorsivo degli artefatti tecnologici, ma a quello dei sistemi più in generale. Lo dimostra l'utilizzo che, dagli albori della disciplina, ne fece Wiener nell'ambito della cibernetica, definendo la *black box* come "a piece of apparatus [...] which performs a definite operation [...] but for which we do not necessarily have any information of the structure of which this operation is performed" (Wiener, 1961: X-XI). L'ambito della cibernetica fu, oltre al contesto teorico e scientifico di origine dell'IA, il nucleo discorsivo in cui la figura della *black box* si afferma e da cui si propaga interdisciplinariamente. Il termine "black box" nasce dallo stretto legame tra ambito militare e cibernetica: come dimostra Galison (1994), esso risale al periodo bellico del secondo conflitto mondiale, ossia a quando nel 1940, all'interno della missione Tizard, un magnetron venne trasportato in una scatola nominata appunto "black box" dal Regno Unito agli Stati Uniti (Von Hilgers, 2011: 47). Al tempo, le ricerche di Wiener che poi confluirono nella sua teoria cibernetica riguardavano proprio lo sviluppo di tecnologie militari (Galison, 1994)⁵.

Da una specifica tecnologia militare, la *black box* divenne all'interno della cibernetica contemporaneamente "a mathematical theory, a device, and a metaphor" (Petrick, 2020: 577) caratterizzante la sua epistemologia. Il modello della *black box* fu fondante del primo modello matematico neurale di McCulloch e Pitts (1943), con cui si cercava di spiegare il funzionamento dei processi cerebrali umani e che costituisce il punto di partenza per i seguenti sviluppi dell'IA.

Da concetto prettamente macchinico, la *black box* è stata presto utilizzata anche in campo psicologico e cognitivo⁶: per via dell'osmosi tra gli studi sulla cibernetica e sulla psicologia del tempo⁷, sulla scia della ricerca sull'IA in quanto simulazione del cervello umano, la stessa metafora è stata usata anche nel contesto delle teorie cognitive comportamentiste per riferirsi al funzionamento del sistema mente cervello⁸. Ma non solo, fu infatti proiettata, oltre che sull'umano, sul mondo stesso: secondo Ross Ashby "we do in fact work, in our daily lives, much more with Black

⁴ Per una ricognizione degli immaginari della narrativa letteraria quanto cinematografica in cui l'IA prende forma antropomorfa in varie figure si veda Cave, Dihal e Dillon (2020); per una prospettiva che va oltre il contesto occidentale, Cave e Dihal (2023).

⁵ Come spiega Galison, infatti, Wiener introdusse nelle sue teorie il termine "black box" e "feedback" proprio a partire dalla progettazione e lo sviluppo di un "anticraft AA predictor", una tecnologia in grado di anticipare e prevedere i movimenti di un aereo nemico per colpirlo.

⁶ Per gli intrecci e le influenze reciproche che caratterizzarono la storia delle teorie psicologiche e cibernetiche si veda Tilak, Glassman, Kuznetcova e Pelfrey, 2022.

⁷ Assieme a due suoi colleghi, lo stesso Wiener applicò il modello della *black box* al comportamento umano. A riguardo si veda N. Hayles in Hui (2024).

⁸ Per una panoramica specifica sull'utilizzo della metafora della *black box* nelle teorie comportamentiste si veda Friedenberg & Gordon (2006: 85-88), e più recentemente in chiave critica Gigerenzer (2019).

Boxes than we are apt to think [...] the real objects are in fact all Black Boxes” e la teoria della *black box* riguarda “what information comes from the object” (1957: 110) – questa prospettiva caratterizzò la cibernetica di secondo ordine che ne espansero le implicazioni in relazione alle scoperte relative ai processi neurofisiologici, nel tentativo di applicare i modelli cibernetici a esperienze vive, a fenomeni che non hanno natura meccanica né una struttura chiusa (Tilak et al., 2022). La metafora della *black box*, dunque, proietta un’ombra di inscruetabilità sul meccanismo interno di funzionamento di un sistema, rendendo pertinente ciò che emerge come visibile, ovvero il risultato e l’informazione ivi contenuta.

Attualmente, tale metafora risulta ancora più pregnante dati i contemporanei sistemi di IA *machine deep learning* che stanno, progressivamente, inserendosi in vari ambiti del quotidiano: dal riconoscimento facciale, di persone o di oggetti, ai recenti sviluppi di IA generativa. In questo contesto, la metafora della *black box* assume la forma tramite cui ci si riferisce all’inscruetabilità dovuta alla complessa stratificazione di reti neurali, algoritmi e non solo. Infatti, anche la componente di partenza, i dati su cui il modello IA viene allenato, può risultare problematica e oscura, questione che ha portato diversi studiosi a indicarne le criticità, ad esempio la possibilità di avere a che fare con “dataset biased”, ricchi di pregiudizi, dovuti soprattutto alla massiva rappresentazione di alcune identità e alla sottorappresentazione o esclusione di altre identità dai dati selezionati⁹.

Come si tenterà di dimostrare, all’interno dell’immaginario dell’IA, la metafora della *black box* si articola in un modo tale per cui si produce una valorizzazione dell’automazione insita al funzionamento dell’IA. Tale valorizzazione può essere positiva, per cui l’idea di un funzionamento automatico non sondabile assume la sinonimia di assenza di intervento umano, intesa come baluardo di obiettività e neutralità; oppure negativa, in cui assume il significato di un’impossibilità di controllo totale da parte dell’umano. Da tale valorizzazione bivalente derivano dunque due assi narrativi secondo cui si strutturano determinate porzioni dell’immaginario dell’IA, spesso in una contrapposizione polarizzata che ricorda quella tra “apocalittici e integrati” (Eco, 1964). All’interno di queste narrazioni, la metafora della *black box* detiene un potere retorico che viene sfruttato o contestato secondo due prospettive: la prima che, seguendo il senso barthesiano del termine¹⁰, è possibile definire *mitica*, e la seconda che vi si contrappone, definibile *critica*. Partendo da un assunto che interpreta a livello semiotico il concetto di immaginario, attraverso l’analisi dei casi studio si vuole rendere conto di come l’immaginario dell’IA che scaturisce dalla metafora della *black box* istituisca discorsi e pratiche in cui tale tecnologia viene considerata in modi diversi.



⁹ Si veda a questo proposito Cave e Dihal (2020) sulle implicazioni razziali e di genere all’interno dell’immaginario e Voto (2022) per quanto riguarda il rapporto tra IA e identità di genere, con l’analisi di un caso studio artistico che problematizza proprio il lavoro invisibile di categorizzazione dei dati, in maniera simile ai casi che presentiamo qui nell’ultimo paragrafo.

¹⁰ Ogni riferimento al mito è fatto a partire dal concetto elaborato da Roland Barthes (1974), cfr. infra 4.

2. Immaginario: una prospettiva semiotica

La *black box* possiede uno statuto ambivalente: essa appare da un lato come una metafora utile a rendere conto del funzionamento (anche se inscrutabile) della tecnologia, ma dall'altro, proprio in quanto oggetto dal contenuto oscuro, indefinito, e in grado di applicarsi a diversi fenomeni, si presta a essere oggetto caratteristico dello spazio dell'immaginario; infatti, il concetto di immaginario non si oppone in maniera esclusiva a quello di realtà. Piuttosto la dimensione dell'immaginario svolge un continuo processo di re-interpretazione e traduzione della realtà, e circolarmente la realtà stessa riprende e re-interpreta i prodotti dell'immaginario. Questo assunto teorico si rivela particolarmente fecondo per comprendere nello specifico i diversi immaginari dell'IA¹¹, dal momento che, come sostengono Stefano e Paolo Bory, "gli immaginari dell'AI non sono mere rappresentazioni finzionali di una realtà sociale futura, piuttosto si tratta di forme creative che agiscono nell'istituzione concreta del pensiero tecnologico, socio-culturale e storico-filosofico del presente" (Bory e Bory, 2015: 83).

Da una prospettiva semiotica è possibile considerare l'immaginario come una dimensione di senso che svolge una funzione di "mediazione tra essere umano e ambiente" (Ferraro, 2011: 92), che dipende quindi dalla facoltà dell'immaginazione¹², e che possiede una natura collettiva, culturale e semiotica (Lorusso, 2011)¹³. Su questo punto la tradizione semiotica di ispirazione peirciana, attraverso la ripresa dello schematismo kantiano, si avvicina all'idea di Castoriadis sul rapporto tra "immaginazione radicale" e "immaginario sociale istituyente", pensato come rapporto fondamentale per l'esistenza del "reale", esperibile individualmente grazie all'immaginazione a partire da significati collettivi dell'immaginario¹⁴.

Proprio come l'immaginazione, l'immaginario costituisce una dimensione di senso dove si mescolano e si esplorano congiuntamente aspetti reali e virtuali dell'esperienza (Volli, 2011). In questo modo l'immaginario viene inteso come un'istanza di mediazione che contribuisce a plasmare i discorsi e parallelamente viene plasmato dai discorsi che circolano nella cultura, si tratta cioè di un sistema autopoietico, in termini semiotici un'enciclopedia parziale.

Seguendo Volli, l'immaginario può essere pensato come un "sistema semantico [...] di relazioni [...] che si può considerare seriamente solo nei suoi tagli locali" (Volli, 2011: 34), la località, in questo caso, è rappresentata dai contesti e dalla posizionalità specifica dei due casi studio presi in esame. Al contempo, un immaginario funziona



¹¹ Su questo punto si è espresso in maniera analoga Guido Ferraro: "in prospettiva semiotica uno degli aspetti più significativi potrebbe consistere proprio nel fatto che la linea d'evoluzione dell'IA presenta un'inconsueta ambivalenza per la quale, a ben vedere, tanto l'immaginario fantascientifico si ispira agli avanzamenti tecnologici quanto questi ultimi prendono spunto, sempre più decisamente, dalle proposte dell'immaginario." (Ferraro, 2023: 66)

¹² Non a caso l'immaginazione è il nucleo fondamentale per il funzionamento della percezione *tout court* (cfr. Paolucci, 2021).

¹³ Più approfonditamente, Lorusso sostiene che l'immaginario possiede la funzione di costruire comunità: "sharing imaginaries is essential to create a community link." (Lorusso, 2011: 114).

¹⁴ Si veda a questo proposito l'articolo di Pavan (2022).

anche come “meccanismo di esclusione” (Volli, 2011: 33), che seleziona alcune caratteristiche come pertinenti, altre no, alcune come euforiche, altre disforiche; in questo modo un immaginario socio-tecnicamente inteso si costruisce anche e soprattutto all’insegna di visioni di futuri desiderabili o di resistenza contro l’ indesiderabile (Jasanoff, 2015). Questa prospettiva permette, quindi, di considerare le tendenze contrastanti interne all’immaginario, che possono manifestare punti di vista più ideologici-mitici o più critici, impegnandosi nell’affermare un determinato asse narrativo-retorico dell’immaginario o al contrario, nel tentare di de-costruirlo¹⁵.

Di conseguenza, enunciare intorno a un certo tema-oggetto, come l’IA nei casi presi in esame, significa necessariamente prendere posizione in relazione a esso e ai discorsi che già circolano nell’enciclopedia (Paolucci, 2020), eventualmente nel tentativo di portare a una loro decostruzione o ridefinizione. Tale presa di posizione ha sempre anche una valenza politica, un significato relativo alle diverse economie (politiche) mobilitate nell’uso dei dispositivi tecnologici (Eugeni, 2021: 259-262), ovvero da una serie di logiche di distribuzione e trasformazione di risorse e di lavoro.

L’immaginario, dunque, più che essere specchio finzionale della realtà, svolge un ruolo fondamentale nell’istituzione, nell’interpretazione e nella creazione della realtà in quanto tale, e quindi degli artefatti, delle relazioni sociali, delle stesse istituzioni politiche ed economiche. Come scrive Castoriadis: “Il soggetto non può incominciare ad abbozzare gli elementi del reale [...] se non muovendo da (e obbedendo a) schemi immaginari che gli sono propri” (1975: 185).

Proprio per questo la mediazione degli immaginari si annida anche nei contesti più “reali”, al punto che è impossibile porre una soglia a priori che separa lo spazio reale da quello immaginario. Piuttosto, come le analisi qui proposte tenteranno di dimostrare, un aspetto importante riguarda la capacità dell’immaginario di riflettere su sé stesso, mettendosi in mostra in quanto immaginario, e la capacità opposta di passare inosservato, come qualcosa di “naturale”. In questo modo un immaginario che riflette su sé stesso può realizzare operazioni critiche, meta-semiotiche, mentre un immaginario che esprime valori ideologici e mitici tende a nascondersi e ad evitare il meta-discorso.

3. Antro *black box*: opacità, profondità e automazione

Come già accennato, dal punto di vista semantico, la metafora della *black box* coinvolge le dicotomie tra interno ed esterno, visibile e invisibile. Attraverso una specifica indicazione cromatica, l’interno costituisce l’elemento invisibile per via di

¹⁵ La prospettiva semiotica si interessa da tempo a tali problematiche, a inclusione dell’intelligenza artificiale; come scrive Massimo Leone: “Una filosofia della comunicazione digitale orientata alla semiotica mira a leggere le tecnologie del senso nel lungo periodo della storia dei sistemi semiotici umani, per rivelare le ideologie implicite che stanno alla base della creazione di nuovi dispositivi, processi e artefatti di significato. L’intelligenza artificiale non fa eccezione, poiché il suo sviluppo è solitamente sostenuto da preconcetti specifici su cosa sia l’intelligenza, su come dovrebbe funzionare e su quali tipi di risultati dovrebbe generare nel mondo.” (Leone, 2023: 35)





un'opacità che in vari discorsi riguardo gli algoritmi e l'IA viene letta in termini di *opacità*. Un'opacità interna che risulta da una stratificazione di reti neurali che si rende profonda: come chiarisce De Seta riguardo l'utilizzo della metafora della profondità che oscilla tra "structural description and metaphorical explanation, technical aspiration and mystifying imaginary" (2023: 261) nelle ricerche del *machine learning* e specificamente del *deep learning*, nel cui specifico caso viene sfruttata la "semantic mystique to promote products and services" (262) – come dimostra Leone (2024), infatti, la figura della *black box*, con la sua aura di mistero, consiste in un dispositivo retorico volto alla promozione di queste tecnologie. All'interno del relativo discorso scientifico, tale profondità si connota secondo tre diverse concezioni: l'accuratezza del modello dovuta alla sua struttura complessa; come idea di rappresentazione della stratificazione interna di una rete neurale e come impenetrabilità, riferita all'impossibilità di comprendere e interpretare le operazioni algoritmiche ivi contenute (Arnold, Tinton, 2021, in De Seta, 2023).

L'opacità della *black box* è inoltre scalare, cioè si applica a diversi livelli che coinvolgono l'IA, come mostra Burrell (2016): a partire dal proprio funzionamento interno e dal modo in cui gli algoritmi operano, al linguaggio di programmazione che rimane inaccessibile alla maggioranza della popolazione e, infine, più in generale, alle forme di protezione e segreto aziendale che non rendono accessibili i propri codici e modelli. Di fatto, dunque, come scrivono Roberge e Seyfert: "there is not one box, but multiple boxes. The opacity of algorithms is more precisely expressed in different forms of opacity, all of which, in specific ways, are contingent on the inbetweenness of a plethora of actors, both human and non-human" (2016: 2).

La metafora della *black box*, dunque, si applica a più livelli e, in ragione della sua sfaccettatura semantica, viene veicolata in diversi apparati discorsivi a seconda delle sue possibili declinazioni. Tanto che, data la contemporanea automazione algoritmica stratificata che coinvolge pervasivamente gli oggetti e le piattaforme digitali della vita quotidiana, senza che questi processi possano essere compresi scalarmente nell'operatività decisionale che li contraddistingue, la società contemporanea è stata definita nei termini di "black box society" (Pasquale, 2015), a cui corrisponde una necessità di trasparenza correlativamente articolata¹⁶.

Ciò è il risultato di una tendenza per cui all'aumentare della densità e dell'opacità dei processi computazionali si produce uno scollamento tra la *possibilità di vedere* e la *possibilità di conoscere* (Chun, 2011: 17). Come spiega Latour, questo processo rientra nelle logiche di lavoro scientifiche e tecniche: nel momento in cui una tecnologia si dimostra efficiente, il punto focale non diviene più il suo funzionamento interno, che diviene "opaque and obscure" (Latour, 1999: 304), ma il suo risultato. In tal modo, l'efficienza diviene un meccanismo retorico e persuasivo che retroattivamente *mitizza* la tecnologia: come illustra Andreas Hepp, "this blackboxing is part of technology's process of materialization: a technology as black box is rarely questioned but taken

¹⁶ Si veda Crawford e Ananny (2018) per un riepilogo riguardo le sfaccettature semantiche della "trasparenza" relativamente all'opacità algoritmica e i limiti che ne contraddistinguono la funzionalità in quanto strategia.

for granted as something 'quasi-natural' which is 'just there' and 'just works'" (2019: 78).

L'efficienza delle attuali tecnologie di IA deriva dalle potenzialità algoritmiche su cui poggiano. L'attuale definizione e concezione degli algoritmi proviene dai processi di meccanizzazione e computazionalizzazione che, all'interno della cornice contestuale della razionalizzazione moderna di politica, economia e scienza (Daston, 2022), hanno stabilizzato il termine nell'indicare "any process of systematic calculation, that is a process that could be carried out automatically" (Chabert, 1999: 1-2), la cui caratteristica è l'efficacia (Knuth, 1997: 4-6)¹⁷. Gli algoritmi sono istruzioni operative automatizzate per la risoluzione di problemi che ne condizionano la progettazione: nella loro programmazione, gli algoritmi sono progettati come "functionally *automatic*, to act when triggered without any regular human intervention or oversight" (Gillespie, 2014: 170). L'automazione diviene dunque *black boxed*, nel momento in cui, data l'impossibilità di interpretarne il funzionamento, ne si problematizzano solo i risultati.

Nella costellazione di discorsi che provengono e costituiscono l'immaginario dell'IA, spesso con fini retorici legati a interessi economici, di tipo utilitaristico (Santangelo, 2023), l'idea di automazione¹⁸ che le tecnologie portano con sé può assumere connotazioni diverse, euforiche o disforiche a seconda dei contesti. All'interno del discorso scientifico, in accordo con la logica della «mechanical objectivity» (Galison e Daston, 2007: 115-183), l'assenza di intervento umano diviene baluardo di obiettività e accuratezza e gli algoritmi, nella loro automazione inumana, divengono "stabilizers of trust, practical and symbolic assurances that their evaluations are fair and accurate, free from subjectivity, error or attempted influence" (Gillespie, 2014: 179). La mitologia dell'efficienza dell'automazione viene, dunque, euforicamente valorizzata come promessa di obiettività e accuratezza. Sempre all'interno dei discorsi scientifici e anche non scientifici, l'idea di un'automazione *black boxed* "is closely tied to conspiracy theory sentiments in which AI is an occult power that cannot be studied, known, or politically controlled" (Pasquinelli e Joler, 2021: 1265), dando luogo a immaginari distopici in cui la macchina soprassiede all'umano in cui l'automazione viene disforicamente valorizzata¹⁹. Particolare testimone della valorizzazione disforica è la lettera pubblicata a marzo 2023²⁰ firmata da Musk e Mustaque (CEO di Stability AI) fra gli altri, in cui viene richiesta dai firmatari una pausa dell'avanzamento della ricerca sull'IA in nome di una serie di ragioni che esprimono le paure e le preoccupazioni riguardo a una perdita di controllo dovuta a un'automazione in grado di prevalere sull'umano.

¹⁷ Per un'approfondita ricostruzione archeologica del termine e del concetto di algoritmo si veda Pasquinelli 2023, pp. 34-38.

¹⁸ Anche l'automazione stessa consiste in un mito. Per una ricognizione sulla genealogia e le implicazioni a livello dell'immaginario dell'automazione si veda Munn (2022).

¹⁹ Si rimanda a Cave, Dihal e Dillon (2020) per la panoramica a riguardo, come a Bory e Bory (2015) specificamente per il cinema.

²⁰ "Pause Giant AI Experiments: An Open Letter", 23 marzo 2023, sul sito di Future Of Life Institute, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.



Lo stretto legame esistente a livello metaforico e immaginativo tra *black box* e automazione dimostra quanto quest'ultima, intesa come promessa di efficienza, si caratterizzi di un'opacità che coinvolge diversi livelli e strati della tecnologia e dell'infrastruttura in cui essa agisce come prodotto e motore.

Al contempo, al tempo presente, diversi approcci e utilizzi della tecnologia dell'IA possono essere letti attraverso le lenti di questo immaginario, ossia come pratiche che si costituiscono a partire dall'immaginario stesso e che, a ritroso, lo costituiscono, in modo reciproco. Da un lato, l'attualizzazione della prospettiva intesa come *mitica* dell'immaginario della *black box*, cioè che naturalizza l'utilizzo della tecnologia in ragione della sua automazione efficiente; dall'altro, l'attualizzazione della prospettiva che, in contrasto, viene considerata *critica*, la quale mira a decostruire la retorica opaca della *black box*. I casi presi in esame per la prima prospettiva sono quelli relativi alla rappresentazione dell'IA in contesto bellico; i casi analizzati per la seconda prospettiva riguardano due opere al confine tra arte e informatica, che cercano di render conto del funzionamento dell'IA.



4. Prospettiva mitica: IA bellica tra immaginario magico e religioso

Le tecnologie digitali sono da sempre oggetto di interesse per la ricerca e l'applicazione militare, soprattutto per quanto riguarda l'ambito della "visione aumentata" (Eugeni, 2021, pp. 83-96), che riguarda le implementazioni della capacità di vedere e agire efficacemente sul campo di battaglia; non a caso la prima versione della *black box*, il magnetron, era un dispositivo simile, un prototipo di radar. La storia delle tecnologie belliche ci mostra inoltre che, parallelamente a un aumento della qualità e dell'ampiezza dello sguardo protesico, aumenta anche la distanza da cui è possibile esercitare tale sguardo, e di conseguenza aumenta la distanza da cui si eseguono operazioni belliche. In questo senso i droni con il loro sguardo onnisciente, costante, quasi divino, "onniveggente" (Chamayou, 2013), condividono le caratteristiche delle più recenti tecnologie di IA; ovvero si basano sull'elaborazione di grandi quantità di dati.

Lo sviluppo di tecnologie IA ha avuto un impatto sull'universo discorsivo bellico, dove si possono notare processi di valorizzazione euforica dell'efficacia automatizzata delle tecnologie, o al contrario valorizzazioni disforiche e relative critiche. Ciò che preoccupa maggiormente rispetto all'introduzione di tecnologie di IA in ambito bellico riguarda l'etica decisionale che presiede al loro utilizzo e che entra in relazione con l'opacità a più livelli della *black box*.

Con lo scoppio della guerra russo-ucraina l'IA è entrata a pieno titolo tra le armi del conflitto, non senza polemiche, e a tal proposito Alex Karp, CEO dell'azienda "Palantir" leader in questo campo, ha sostenuto la necessità di "un'architettura che permetta la trasparenza sui dati-sorgente"²¹. L'attenzione alla trasparenza dei dati

²¹ Frase pronunciata da Karp, tradotta autonomamente, e riportata in Sterling, e van Den Berg (2023): <https://www.reuters.com/technology/ukraine-war-shows-urgency-military-ai-palantir-ceo-says-2023-02-15/>

emerge quindi come condizione necessaria per un uso etico delle tecnologie IA in ambito bellico (e non solo), ma come vedremo, nella realtà dei fatti questa trasparenza non è quasi mai realizzata.

Inoltre, nel discorso di Karp, pronunciato in un summit internazionale, il riferimento alla dimensione dell'immaginario è volontariamente escluso; ma, come visto in precedenza, le figure dell'immaginario si annidano anche nei contesti più reali, e in questo caso svolgono una funzione di mediazione tanto nella rappresentazione quanto nella percezione comune di tecnologie IA reali a tutti gli effetti.

Infatti, considerando il livello della rappresentazione identitaria dell'azienda di Karp, è possibile notare diversi riferimenti a dimensioni immaginarie, a partire dal nome stesso dell'azienda. La scelta del nome "Palantir" è infatti un esplicito riferimento al mondo de *Il signore degli anelli*: nello specifico, il Palantir è un oggetto magico che permette di vedere a distanza, ma il cui funzionamento non è del tutto chiaro (Tolkien, 1946). Il riferimento a un oggetto dai poteri magici sovrumani è funzionale a dare l'idea di una tecnologia potente ed efficace ben oltre le competenze e conoscenze umane, e in maniera non troppo velata evoca il principio di inconoscibilità caratteristico della *black box*. Prestando poi attenzione al breve video che viene riprodotto automaticamente appena si accede alla sezione "defense" del sito di Palantir²², si può notare un altro esempio di ibridazione di immaginari, in questo caso con il ricorso a un'estetica ricalcata su videogame e film fantascientifici: montaggio rapido, luoghi estremi, desertici e glaciali, dove compaiono veicoli militari, mentre gli attori umani osservano rigorosamente a distanza, il tutto accompagnato da una musica elettronica dai toni epici. L'immaginario tecnologico-bellico delle IA si carica di significati propri dell'immaginario fantasy, cinematografico e videoludico, affiancandosi a esso piuttosto che ritagliandosi un proprio spazio esclusivo, sfruttandolo a fini retorici, portando alla costruzione di una figura "mitica"²³ della tecnologia IA.

Come sostiene Barthes, il processo di mitizzazione di un oggetto – in questo caso la tecnologia bellica – dà vita a una "*parola depoliticizzata*" (1974: 223). Il mito



²² "Defense", Palantir, <https://www.palantir.com/offerings/defense/>.

²³ Riprendendo l'idea di mito di Barthes, possiamo sostenere che siamo di fronte a oggetti reali che acquistano qualità mitiche; si tratta di un processo di "deformazione del senso" di un oggetto, che si riveste di una significazione ulteriore, una "connotazione" (cfr. Barthes, 1974: 191-238) funzionale a uno scopo retorico, molto spesso di tipo ideologico. Per Barthes il mito "impoverisce" il senso, "lo allontana", ma allo stesso tempo "lo tiene a disposizione"; infatti, continua Barthes, "per quanto paradossale possa sembrare *il mito non nasconde nulla*: la sua funzione è di deformare, non di far sparire." (p. 203) Recentemente il concetto di mito è stato ripreso in semiotica per l'analisi di nuovi media (Stano e Ponzio, 2023); giustamente Stano sottolinea anche il carattere positivo, euforico, del mito, inteso come meccanismo di coesione sociale e comunitaria, aspetto che nella lettura barthesiana fosse rimasto in ombra. Convinti che il mito svolga anche e soprattutto tali funzioni, la nostra prospettiva si preoccupa maggiormente di mettere in luce i processi di connotazione e deformazione del senso che interessano l'oggetto del mito, nel tentativo di mostrare successivamente operazioni contrastanti, che definiamo critiche, che tentano in diverso modo di mettere in mostra i presupposti ideologici su cui si basa la costruzione, la rappresentazione e la percezione della tecnologia IA.

funziona dunque come “un metalinguaggio, basato su operazioni connotative, che naturalizza determinate visioni ideologiche” (Stano e Ponzio, 2023: 12).

Come ricordato in precedenza ogni utilizzo e ogni rappresentazione della tecnologia si carica anche di valenze politiche ed economiche; tuttavia, un simile processo di mitizzazione dell’oggetto tecnologico è funzionale a deformare quelle responsabilità politiche, intrinsecamente umane, che sono coinvolte nell’uso e nella rappresentazione della tecnologia, e che sono irriducibili a qualsiasi tipo di determinismo tecnologico (Eugeni, 2021).

a. *The Gospel: immaginario religioso e il mito dell’efficienza dell’IA*

La questione della responsabilità etica e politica trova spazio anche e soprattutto nel contesto dell’attuale guerra a Gaza, dove l’uso di un software di IA da parte dell’esercito israeliano ha scatenato diverse critiche.²⁴ In questo caso, è possibile osservare la costruzione di un legame tra immaginario religioso e immaginario bellico all’interno in un discorso teso a giustificare l’uso di armi IA e più in generale l’operato militare israeliano²⁵.

Il termine in ebraico che designa il software di IA sul sito dell’IDF è “Habsora”, tradotto in inglese con “Gospel”, contrazione di “God’s spell”, ovvero parola di Dio, che in italiano corrisponde a “Vangelo”, mutuato dal greco “*euangelion*”, ovvero “buone notizie” (Della Ratta, 2024). La funzione del software è generare obiettivi militari nemici da colpire sulla base di un lavoro di analisi computazionale su dati di Intelligence di cui dispone IDF. Come si può leggere nella pagina web di IDF, “Gospel è un sistema che consente l’uso di strumenti automatici per produrre obiettivi a un ritmo veloce e funziona migliorando materiale informativo accurato e di alta qualità secondo le esigenze”²⁶. Si evince dunque un forte legame tra l’automazione e la rapidità del processo di selezione degli obiettivi, misurata in quantità di edifici abbattuti. Anche qui ci si ritrova di fronte a un processo di mitizzazione che sfrutta il sostrato immaginario del discorso religioso per connotare l’efficacia e la rapidità del software, allontanandolo dal suo reale contesto di applicazione²⁷.

La retorica della precisione e dell’efficienza dell’IA si appoggia qui all’immaginario religioso per costruire una figura mitica della tecnologia e circolarmente anche dell’esercito che la usa. L’immaginario religioso costituisce un luogo di coesione



²⁴ Il primo articolo reperibile è stato pubblicato da *The Guardian* il 3/12/2023, ripreso e commentato da testate giornalistiche di diversi paesi; sono seguiti altri articoli sullo stesso tema pubblicati su *Local Call* e *+972 Call*.

²⁵ Già nel 2019 sul sito di IDF era possibile leggere la brochure di presentazione di un nuovo tipo di drone, “Harpy”, che “funziona in maniera totalmente automatica”. Non si trattava ancora di un vero e proprio modello di IA con deep learning ma la retorica atta a valorizzare l’efficacia di un funzionamento autonomo è simile. Anche qui il riferimento alla creatura mitologica è sintomatico di una tendenza all’ibridazione tra immaginario bellico e immaginario sovrumano.

²⁶ Citazione tratta dal sito in questione in traduzione italiana di Google, di cui il link alla pagina web di IDF dove viene presentato il software Gospel.

²⁷ Infatti, come scrive Donatella Della Ratta in un commento all’inchiesta di *The Guardian*: “despite its biblical namesake suggesting absolute truth and an infallible message, *Habsora*, like all machine learning systems, operates on correlation rather than causation and likelihood rather than certainty”.

sociale e politica per molte collettività, nello specifico quella israeliana – non a caso i soldati morti nella guerra sono designati da IDF come “martiri di guerra”²⁸ – e serve a interpretare la tecnologia in chiave teleologica, come qualcosa che naturalmente porta alla prosperità della comunità che la usa.

Sempre nella pagina web di presentazione di Gospel, compare anche un video che mostra proprio i risultati del lavoro del software, o meglio, la loro interpretazione pragmatica; si tratta di una serie di bombardamenti in inquadrature da pochi secondi l’una, con un effetto di valorizzazione estetica della velocità delle prestazioni dell’IA, all’interno di una strategia che tende a valorizzare maggiormente i risultati, piuttosto che i processi che presiedono a quest’ultimi, e da cui dipende la logica che li interpreta come efficaci. Paradossalmente nella narrazione dell’esercito israeliano l’accento viene posto anche sulla competenza umana di controllo delle informazioni, nel tentativo di umanizzare un processo che, supervisionato o meno da attori umani, è causa di morte e devastazione in quantità e modalità spaventose.

Questo processo di umanizzazione dell’utilizzo del software di IA da parte di Israele si appoggia da un lato a una retorica della disumanizzazione del nemico, e dall’altro è a sua volta funzionale a deformare e nascondere ogni tipo di responsabilità politica delle persone e delle istituzioni coinvolte nell’uso di tali armi. Se per la narrazione di IDF il controllo dell’umano è totale, le fonti intervistate raccontano invece di selezioni ed elaborazioni di dati grossolane, e di scelte strategiche che consapevolmente accettano la morte “collaterale” di numerosi civili²⁹. Si può parlare a questo proposito anche di discorso o immaginario ideologico, secondo la definizione di Eco³⁰; se per Barthes il mito deforma, per Eco l’ideologia nasconde effettivamente le ragioni della sua stessa enunciazione, “bloccando in questo modo la funzione metasemiotica del discorso” (Stano, 2022: 10). Un discorso come quello portato avanti da IDF, servendosi dell’immaginario religioso, diventa ideologico in quanto blocca la natura meta-semiotica del discorso, celando i criteri che sono alla base dell’emersione di un certo risultato.

È possibile allora ritornare all’idea di Karp sulla necessità di una trasparenza sui dati-sorgente; nel caso di Gospel i dati di intelligence vengono usati per etichettare gli abitanti di Gaza, i quali ricevono un punteggio indicativo della loro vicinanza al gruppo armato di Hamas. Tuttavia, questi dati non sono pubblici, e non si tratta di una *black box* tecnica, quanto piuttosto di una *black box* politica, di stampo ideologico, funzionale a celare i criteri di pertinenza per cui un soggetto è considerato più o meno vicino a Hamas, e la cui morte è quindi più o meno giustificata. Più profondamente, se manca una trasparenza relativa ai dati-sorgente



²⁸ Denominazione che compare sul sito di IDF, di cui il link.

²⁹ Si tratta di una specifica strategia di IDF, detta “dahiya doctrine”, utilizzata per la prima volta nella guerra in Libano nel 2006, e si basa su una distruzione pressoché totale delle infrastrutture civili, considerate come centri di potere del nemico.

³⁰ Secondo Eco, “l’ideologia è visione del mondo parziale e sconnessa [...] essa cela anche le ragioni pratiche per cui certi segni sono stati prodotti” (Eco, 1975: 369). L’ideologia, infatti, consiste proprio in un tipo di discorso che parte da un’interpretazione parziale e selettiva del mondo, e sceglie di considerare interpretativamente solo quei tratti di significato funzionali al suo progetto retorico. Sui rapporti tra mito e ideologia, concetti che a volte sembrano coestensivi, si veda Leone e Stano (ed.) 2023.

un algoritmo *può essere sfruttato per produrre qualsiasi risultato*, e non sempre è possibile ricostruire il processo che ha prodotto tale risultato.

Il panorama del discorso bellico sta cambiando attraverso l'uso e la rappresentazione delle novità tecnologiche, e si sta assistendo ad un processo di depoliticizzazione e de-responsabilizzazione degli attori umani all'interno del contesto operativo militare. Tale processo si avvale di una più o meno tacita valorizzazione euforica del progresso tecnologico, inteso come indispensabile, giusto, e quindi, di fatto, naturalizzato e assunto come mito della contemporaneità. Infine, il potere delle tecnologie chiamate "intelligenti" si basa su un processo di antropomorfizzazione delle stesse, e contemporaneamente permette agli attori umani implicati nelle operazioni militari di assumere un'identità quasi anonima, relegando la responsabilità dell'azione a un software di cui spesso non si conoscono tutti i dettagli relativi al funzionamento, e che si basa su una forma di estrazione massiva di dati.

A fianco di narrazioni che costruiscono le tecnologie attraverso rimandi agli immaginari fantastici e fantascientifici, si può intravedere un cambiamento reale, relativo alla percezione del rischio e delle conseguenze della guerra. Infatti la portata operativa delle tecnologie militari è arrivata a un'efficacia mai vista prima (misurata in termini di velocità di distruzione, come abbiamo visto con il software Gospel) tale da rendere la guerra stessa un fenomeno dalle caratteristiche fantascientifiche. Il coinvolgimento dell'umano è sempre più distante dal contesto fisico della guerra, poiché è distante lo sguardo protesico esercitato dai dispositivi tecnologici, i quali elaborano quantità di dati la cui scala è praticamente impossibile da maneggiare per l'umano. Questo cambiamento nel modo di condurre e pensare la guerra, già in atto prima delle recenti IA (cfr. Chamayou, 2013), si è consolidato anche grazie alla retorica dell'opacità, caratteristica di una black box miticizzata.

Per questa ragione vengono presentati ora quei casi di interpretazione critica della *black box* delle IA, che costruiscono un immaginario contrapposto a quello qui presentato, attraverso operazioni orientate alla visualizzazione dei processi di categorizzazione ed elaborazione di dati che spesso rimangono invisibili, e spesso vengono giustificati in base a una retorica dell'efficienza "naturale".

5. Prospettiva critica: arte del(l'in)visibile

All'interno dell'immaginario, l'articolazione semantica della metafora della *black box* imperniata sulle categorie dicotomiche di visibile e invisibile, interno ed esterno, delinea degli assi di azione su cui operare. Procedendo da un elemento della categoria all'altro o viceversa, tali assi costituiscono le logiche di operazioni che si strutturano e attualizzano in quanto tali in ragione delle possibilità offerte dalla delimitazione semantica insita nell'immaginario. All'interno del contesto dell'arte contemporanea, si è affermata una tendenza che agisce sull'asse della visualizzazione, ossia di rendere visibile l'invisibile, processo spesso indicato con la dicitura "opening the black box" (Wasielowski, 2023).



Tale tendenza³¹ ha costituito il nucleo progettuale di operazioni artistiche a partire già dal decennio scorso, in cui il flusso globale iconico e informativo, l'avvento dei Big Data e la presa di consapevolezza circa l'opacità degli algoritmi trainanti la società digitale han dato luogo alla necessità di rendere i processi algoritmici visibili, rendere conto delle responsabilità di quest'ultimi e sfidare l'apparente immaterialità del digitale.

In tal modo, l'arte si (ri)configura come una "scienza del visibile" che "funziona [...] in quanto contro-discorso orientato a profanare l'opacità strutturale delle tecniche di osservazione e dei processi di rappresentazione che manipolano la realtà (e la sua immagine) traducendola nel territorio della visibilità artistica" (Malavasi, 2017: 81). Ed è proprio questo processo traduttivo che mostra come l'immaginazione artistica si muova all'interno dell'immaginario sociotecnico della *black box* secondo un asse contrastivo e critico, che, nel caso specifico dell'IA, si oppone alla sua retorica.

La prospettiva critica del dispositivo artistico non può, dunque, non divenire politica nel momento in cui la *black box* dell'IA viene (relativamente) aperta, seguendo un approccio che Martin Zeilinger definisce tattico in senso decerteauiano nel momento in cui l'intento è di resistere a "strategic approaches that blackbox knowledge" e di opporsi al "blackboxing of technology" (2021: 51 e 154) perseguendo modalità di emancipazione dello spettatore o spettatrice (Rancière 2008) in accordo con una prospettiva di arte *critica* così come elaborata da Rancière³².

Pasquinelli e Joler dimostrano come l'IA consista in uno strumento estrattivo e organizzativo della conoscenza (2021) che viene diffratta nella realtà secondo le lenti del modello statistico: la conoscenza ingerita (i dati prelevati dal reale) viene sottomessa e poi restituita in una forma mediata da una specifica razionalità – scientifica, statistica, algoritmica – che, positivamente valorizzata e accettata come criterio decisionale, genera effetti sul reale. Rendere visibili i meccanismi della *black box* significa esporre la logica e la razionalità di questi "instrument of power" che operano secondo "an aesthetic and ideology of power" (Paglen, 2016).

Questa prospettiva critica è alla base di due opere che, intersecando arte, tecnologia e informatica, sono state esposte in spazi rilevanti per l'arte contemporanea. Come si procederà a dimostrare, queste due opere condividono il tentativo di visualizzazione di elementi che compongono la *black box* dell'IA, ossia delle logiche tassonomiche di organizzazione dei dataset che ne fungono da confini epistemici per il funzionamento e delle implicazioni visuali che la performatività algoritmica reca con sé.

a. Immersi nel dataset: speleologia dentro la black box

³¹ Si tratta di una tendenza fra le tante sfaccettature delle operazioni che caratterizzano il campo della cosiddetta *AI Art*. Per una panoramica generale, cfr. Barale, 2020; Audry, 2021; Pedrazzi, 2021; Grba, 2022 e 2024.

³² A questo proposito, sull'idea di arte critica di Rancière e il concetto di attivismo, si veda Trione, 2022, pp. 183-185.



From 'Apple' to 'Anomaly' (Pictures and Labels) è stata commissionata dal Barbican di Londra appositamente per la galleria The Curve all'interno del progetto *Life Rewired*. Inaugurata a fine settembre del 2019 e conclusasi nel febbraio successivo, pubblicitarimente descritta come "revealing the powerful, and often hidden, forces at play in artificial intelligence"³³, l'installazione contava e si componeva di approssimativamente 30'000 fotografie stampate e appese sulla lunga parete curva dello spazio immersivo della galleria (fig. 1).



263



Fig. 1 – Veduta dell'installazione "Trevor Paglen – From 'Apple' to 'Anomaly' (Pictures and Labels). Selections from the ImageNet dataset for object recognition", Barbican Centre, Londra. 26 settembre 2019 – 16 febbraio 2020. Foto di Max Colson. Courtesy dell'artista e del Barbican Centre –The Curve.

Le fotografie e le 200 categorie secondo le quali erano organizzate provengono da uno dei primi e più famosi dataset per l'allenamento dell'IA nel riconoscimento di oggetti, ossia Image Net, sul quale Paglen, in collaborazione con Kate Crawford, continuerà ad approfondire le proprie ricerche – segue infatti *Image Net Roulette*³⁴ – che sono poi culminate in un testo fondamentale intitolato *Excavating AI: The Politics*

³³Come si può leggere sulla pagina dedicata presente sul sito del Barbican, <https://sites.barbican.org.uk/trevorpaglen/>.

³⁴ Progetto del 2019, che consisteva in una rete neurale addestrata tramite i dati relativi alla categoria "persona" presenti in Image Net. La rete neurale addetta al riconoscimento facciale riproponeva, dunque, le classificazioni problematiche contenute in Image Net, <https://www.chiark.greenend.org.uk/~ijackson/2019/ImageNet-Roulette-cambridge-2017.html>.

of Images in Machine Learning Training Sets³⁵. Il titolo del saggio che corona questo percorso di ricerca è particolarmente interessante, perché permette di concepire retroattivamente la prima parte di questo scavo, ossia *From 'Apple' to 'Anomaly'*, come un tentativo di fornire un'esperienza speleologica all'interno della *black box* e, quindi, come esempio di una delle modalità di visualizzazione di quest'ultima.

ImageNet venne realizzato nel 2009 sotto la direzione di Fei Fei Li con l'intento di "mappare l'intero mondo degli oggetti"³⁶: composto di 14 milioni di immagini, consiste in un'organizzazione di quest'ultime secondo categorie e tassonomie che sono necessarie per il funzionamento degli attuali sistemi di IA *machine learning*. Il dataset di allenamento rappresenta di fatto il pacchetto di conoscenza mediata astratta dal mondo reale che ne diviene i confini epistemici di azione (Crawford, 2021). Il modo in cui questa conoscenza viene trattata, organizzata, gerarchizzata si poggia su associazioni visivo-verbali arbitrarie che pretendono di far convergere due significanti diversi di due semiotiche differenti per il medesimo referente, recando con sé problematiche a livello epistemologico e politico (Crawford e Paglen, 2021). Tali associazioni, una volta ingerite all'interno dello spazio latente dell'IA, lo strato più profondo della *black box*, divengono inoltre non più leggibili e comprensibili dall'occhio umano (Paglen, 2016).

L'installazione di Paglen espone, rendendo visualizzabile, la logica semiotica di cui la *black box* dell'IA ha bisogno per funzionare e che operativizza nei meccanismi algoritmici. Una logica che regola, nelle parole di Paglen, "the political economy of meaning that is produced through such apparatuses" (2020). La visualizzazione messa in mostra si realizza tramite la traduzione visiva e materiale delle immagini del dataset in fotografie stampate e appese, rendendo manifesto il bisogno quantitativo di informazioni di un sistema di IA assieme a quello delle categorie verbali, riproposte sotto forma di etichette, per l'organizzazione del materiale visivo.

Questa visualizzazione avviene inoltre in uno spazio che ricrea a sua volta un effetto *black box*, dove, nell'immersività della galleria in cui, a parte la parete curva illuminata, a dominare è il buio, la conformazione spaziale dell'installazione ha reso spettatori e spettatrici "walkers", richiedendo "an active traversal of the gallery space" (Zylinska, 2020: 97). In questo senso, *From 'Apple' to 'Anomaly'* rappresenta un tentativo di speleologia di un antro della *black box* – una sua parete, in questo caso. Annunciata già dal titolo, il quale ricorda come le molteplici immagini in mostra siano solamente quelle appartenenti a un breve segmento di lemmi raggruppati in ordine alfabetico secondo la lettera iniziale, questa limitatezza, che fa i conti con le *contraintes* materiali-fenomeniche, attesta l'impossibilità di visualizzare nella sua totalità anche solo un elemento della *black box* richiedente incommensurabili



³⁵ Il progetto di Paglen e Crawford raggiunse una popolarità tale che rese di dominio pubblico le problematiche razziali, sessiste – e non solo – insite nei dataset di addestramento. Ciò provocò la rimozione di quasi un milione di immagini nel tentativo di correggere i bias. A riguardo, si veda Ruiz 2019, <https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias>.

³⁶ Citazione tradotta in italiano che appare in Gershgorn (2017), <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>.

quantità di dati. Questo a riprova del fatto che "nevertheless, the full interpretability and explicability of machine learning statistical models remains a myth" (Pasquinelli e Joler, 2021: 1265).

b. Learning to See: sguardi stranianti sull'opacità algoritmica

Se l'opera di Paglen mette in mostra il dataset, cioè esibisce le logiche tassonomiche e semiotiche che, a monte, garantiscono l'operatività della *black box*, l'installazione di Memo Akten rende visibile il modo con cui i meccanismi operativi algoritmici interni applicano la conoscenza fornitagli su input visivi, definendo una propria *way of seeing* (Berger, 1972).

Presentata alla mostra *AI: More than Human* tenutasi sempre al Barbican tra i mesi di maggio e agosto 2019, l'installazione di Memo Akten *Learning to See* è compresa all'interno dell'omonima serie³⁷, a sua volta parte del suo progetto di ricerca iniziato nel 2017, culminato nella sua tesi di dottorato discussa nel 2021 alla Goldsmith University of London (Akten, 2021).

Parte della terza sezione, *Data World*³⁸, l'installazione interattiva (fig. 2) si compone di due schermi appesi alla parete che esibiscono in diretta ciò che la videocamera riprende sul tavolo posizionato di fronte – un insieme eterogeneo di oggetti quotidiani – e quello che "vede" la rete neurale. Tramite la manipolazione tattile e diretta degli oggetti da parte dei visitatori e delle visitatrici diviene possibile osservare i cambiamenti figurativi nei due schermi, sia quello dell'immagine ritraente la disposizione degli oggetti sul tavolo che quello della corrispondente immagine generata dalla rete neurale, addestrata specificamente con cinque diversi e ristretti dataset³⁹ che si alternano ogni trenta secondi.

La simultanea e comparata apparizione in tempo reale di due modi di vedere – a destra, quello della camera, che ripropone a un livello di alta iconicità gli oggetti presenti e, a sinistra, quello dell'IA, che mostra la trasmutazione degli stessi – fornisce allo spettatore o alla spettatrice la possibilità di uno sguardo straniante sulla rappresentazione in atto secondo una logica di *showing seeing* (Mitchell, 2002). Giocando sull'effetto di neutrale obiettività proposto dalle immagini della camera, che emerge in contrasto a quello di *irrealità* condizionata dell'IA, l'opera di Akten pone in discussione la percezione visiva di chi osserva e, contemporaneamente, di quella programmata della macchina intelligente.



³⁷ Attraverso la realizzazione, lo sviluppo e l'addestramento di una rete neurale profonda, il progetto di Akten intende indagare le modalità di comprensione visuale della realtà di tale "visual instrument" (Akten et al., 2019) e, a un livello più generale, i modelli di *deep learning* in quanto nuovi medium artistici. <https://www.memo.tv/works/learning-to-see/>

³⁸ Qui di seguito il comunicato stampa della mostra, <https://www.barbican.org.uk/sites/default/files/documents/2019-07/FINAL%20AI%20More%20than%20Human%20exhibition%20release%20280219.pdf>.

³⁹ Ciascuno dei cinque dataset si compone unicamente di immagini relative ai quattro elementi naturali (acqua, aria, terra e fuoco) e al cosmo.

Nel mettere in mostra la performatività dell'algoritmo, *Learning to See* permette di comprendere come la visione dell'IA coinvolga una dinamica interpretativa che dipende dalla conoscenza ingerita nella sua *black box*, rifratta sulla realtà secondo logiche statistiche; retroattivamente, mette in discussione anche una concezione di percezione pura attribuita alla vista umana (Audry, 2021; Bueno e Abarca, 2020: 1185). Nello stesso modo in cui si verifica nel caso analizzato da Voto (2022) seguendo le categorie di opacità e trasparenza di Marin (1994), la performance algoritmica in tempo reale, che emerge dal contrasto con un'immagine *più reale* di ciò che viene ritratto, *presenta* a chi osserva la rappresentazione dell'IA, smontandone la trasparenza e rendendo in tal modo opaco, quindi intelligibilmente "visibile", lo spazio latente che ne condiziona la forma, "redeeming it from the inaccessibility with which the uninitiated normally deal" (Voto, 2022: 50).

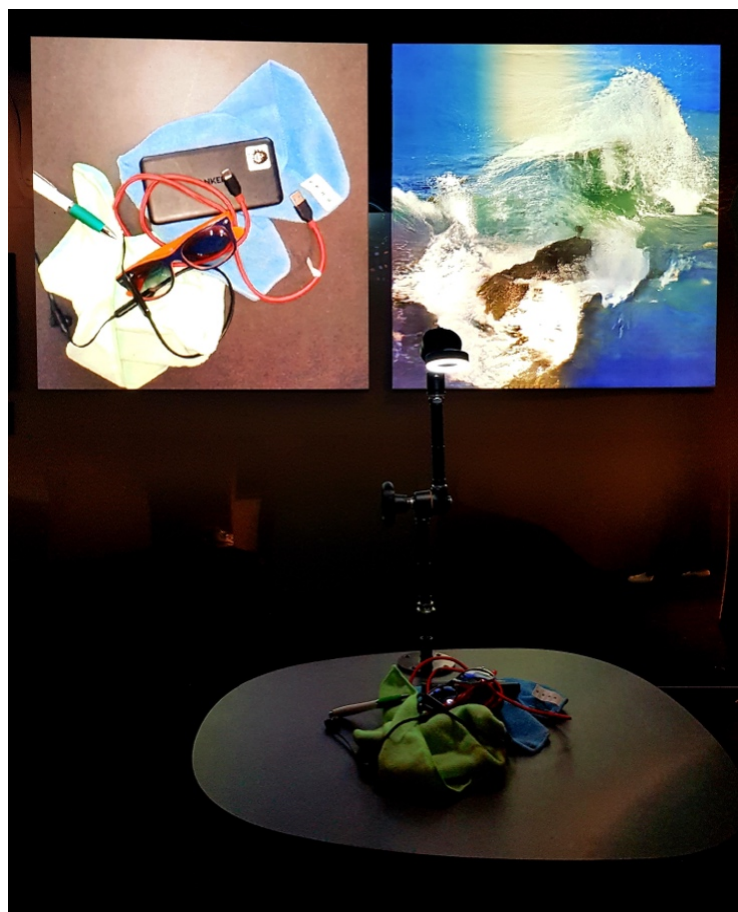


Fig. 2 – Veduta dell'installazione "Learning to See" inclusa in *AI: More than Human*, Barbican Centre, Londra, 2019. Courtesy dell'artista.

Tuttavia, la limitatezza di conoscibilità si riscontra anche qui: la *black box* rimane *inaccessibile* al suo interno e *invisibile* all'esterno. L'unica possibilità consiste nel coglierla a posteriori, negli output che realizza, con la consapevolezza della costruzione, parzialità e problematicità epistemologica del suo sguardo.

6. Conclusione

Semioticamente inteso, l'immaginario dell'IA si dimostra non solo come una rappresentazione finzionale della tecnologia, ma come una dimensione di senso che contribuisce attivamente alla sua creazione e interpretazione; allo stesso tempo e di conseguenza, l'immaginario plasma e viene plasmato dai relativi discorsi culturali e socio-tecnologici. All'interno di questa circolarità, la semantica della figura della *black box* apre possibilità di dinamiche e percorsi di senso che ne influenzano le attualizzazioni, ossia quelle pratiche e rappresentazioni che, coinvolgendo l'IA, ne rinforzano la retorica dell'automazione efficiente o la criticano tramite la decostruzione dell'opacità.

Dopo aver ricostruito la rete delle sfaccettature semantiche che compongono la figura della *black box* (profondità, opacità, automazione), tra oscurità e opacità a più livelli, talvolta discordanti, si è voluto mettere in luce come da tale metafora si articolino discorsi con valorizzazioni differenti a seconda dei contesti. La comparazione di porzioni locali dell'immaginario della *black box*, quello bellico e quello artistico, ha dunque permesso una ricognizione – che non pretende di essere esaustiva – del significato e dell'utilizzo di tale metafora nel caso dell'IA e, dunque, delle sue conseguenze politiche, sociologiche ed estetiche.

L'incursione nei territori dell'arte contemporanea potrebbe continuare approfondendo i diversi livelli dell'esperienza estetica con l'IA e le diverse strategie di enunciazione utilizzate nel tentativo di visualizzare i processi che presiedono all'emergenza di determinati risultati. Parallelamente si potrebbe proseguire con il confronto tra queste operazioni, definite critiche, e i discorsi ideologici e mitici discussi in ambito bellico con i relativi processi di deformazione che rendono opachi e dunque inconoscibili diversi aspetti dell'IA, in una costante negoziazione tra opacità tecnica e strategie di nascondimento e di inganno.

Ad ogni modo, l'analisi qui presentata ha permesso di far emergere e sottolineare la capacità dei discorsi critici di riflettere sulle implicazioni politiche che questa metafora porta con sé, soprattutto nelle sue attualizzazioni e nei relativi contesti operativi. L'approccio critico e straniante dell'ambito artistico nei confronti dell'invisibilità opaca della *black box* permette di rileggere a ritroso i discorsi che si avvalgono della potenza retorica di tale figura provenienti da altri ambiti in modo più critico, ponderandone gli effetti.

In conclusione, la metafora della *black box*, nella sua costitutiva parzialità e località, riflette e contribuisce alla complessità e alla diversità di narrazioni e pratiche relative all'immaginario dell'IA, rappresentando sia una sfida che un'opportunità



Lorenzo Aimo, Luca Brunet
Immaginario Black Box

nell'interpretazione e nella definizione del rapporto tra questa tecnologia pervasiva e la società.



Bibliography

- Akten M. (2021), *Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression*, Doctoral thesis, Goldsmiths, University of London [Thesis], <https://doi.org/10.25602/GOLD.00030191>.
- Akten M., Fiebrink R., Grierson M. (2019), "Learning to see: you are what you see", in ACM SIGGRAPH 2019 Art Gallery (SIGGRAPH '19). Association for Computing Machinery, New York, NY, USA, Article 13, 1-6. <https://doi.org/10.1145/3306211.3320143>.
- Alexander J.C. (2008), "Iconic consciousness: the material feeling of meaning", *Environment and Planning D: Society and Space*, 26, pp. 782-794.
- Arnold T., Tilton L. (2021), "Depth in Deep Learning: Knowledge, Layered, Impenetrable", in Redrobe K., Scheible J. (ed.) *Deep Mediations*, Minneapolis, University of Minnesota Press, pp. 309-329.
- Ashby R. W. (1957), *An Introduction to Cybernetics*, London, Chapman&Hall LTD.
- Audry S. (2021), *Art in the Age of Machine Learning*, London & Cambridge (MA), The MIT Press.
- Barale A. (2020), *Arte e Intelligenza Artificiale. Be my GAN*, Milano, Jaca Book.
- Barthes R. (1957), *Mythologies*, Seuil, Paris, (trad. it. *Miti d'oggi*, Torino, Einaudi, 1974).
- Berger J. (1972), *Ways of seeing*, London, Penguin Books Ltd.
- Bory S. e Bory P. (2015), "I nuovi immaginari dell'Intelligenza Artificiale", *Im@go. A Journal of the Social Imaginary*, n. 6, pp. 66-85, doi: <https://doi.org/10.7413/22818138047>
- Burrel J. (2016), "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, 3(1), pp. 1-12, <https://doi.org/10.1177/2053951715622512>.
- Castoriadis C. (1975), *L'institution imaginaire de la société II. L'imaginaire social et l'institution*, Paris, Editions du Seuil (trad. it. *L'istituzione immaginaria della società*, Torino, Bollati Boringhieri, 1995).
- Cave S., Dihal K (2020), "The Whiteness of AI", *Philosophy & Technology*, 33:685-703, <https://doi.org/10.1007/s13347-020-00415-6>.





- Cave S., Dihal K. (ed.) (2023), *Imaging AI. How the World Sees Intelligent Machines*, Oxford, Oxford University Press.
- Cave S., Dihal K., Dillon S. (ed.) (2020), *AI Narratives. A History of Imaginative Thinking about Intelligent Machines*, Oxford, Oxford University Press.
- Chabert J. (ed.) (1999), *A History of Algorithms: From the Pebble to the Microchip*. Berlin, Springer.
- Chamayou G. (2013), *Théorie du drone*, Paris, La Fabrique Éditions.
- Chun W. H. K. (2011), *Programmed Visions: Software and Memory*, Cambridge (MA), The MIT Press.
- Crawford K. (2021), *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*, New Haven & London, Yale University Press (trad. it. *Né intelligente, né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino).
- Crawford K., Annany M. (2018), "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability", *new media & society*, 20(3), pp. 973-989, <https://doi.org/10.1177/146144481667664>.
- Crawford K., Paglen, T. (2021), "Excavating AI: The Politics of Images in Machine Learning Training Sets", *AI & Society*, 36:1105-1116, <https://doi.org/10.1007/s00146-021-01162-8>.
- Daston L. (2022), *Rules. A Short Story of What We Live By*, Princeton & Oxford, Princeton University Press.
- De Seta G. (2023), "Digital Depth: A Volumetric Speculation", *magazén*, 4(2), pp. 245-270, doi:10.30687/mag/2724-3923/2023/02/004.
- Della Ratta D. (2024), "Killing intelligence: death by tech and other ordinary horrors in Gaza", *Untoldmag*, consultato il 25 aprile da: <https://untoldmag.org/killing-intelligence-death-by-tech-and-other-ordinary-horrors-in-gaza/>
- Eco U. (1964), *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Milano, Bompiani.
- Eco U. (1975), *Trattato di semiotica generale*, Firenze, Giunti editore.
- Eugeni C. (2021), *Capitale algoritmico. Cinque dispositivi postmediali (più uno)*, Brescia, Scholé.
- Ferraro G. (2011), "Dinamiche dell'immaginario: una prospettiva semiotica", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 91-102.



Ferraro G. (2023), "Kito e realtà dell'IA: uno sguardo semiotico", in Leone M. e Santangelo A. (ed.), *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 65-90.

Friedenberg J. & Silverman G. (2012), *Cognitive Science: An Introduction to the Study of Mind*. Thousand Oaks, SAGE Publications.

Gershorn D. (2017), "The data that transformed AI research – and possibly the world", *Quartz*. Consultato il giorno 5 maggio da: <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>.

Gigerenzer G. (2019). "How to Explain Behavior?", *Topics in Cognitive Science*, 12:4, pp. 1363-1381, <https://doi.org/10.1111/tops.12480>.

Gillespie T. (2014), "The Relevance of Algorithms" in Gillespie T., Boczkowski P., Foot K. (ed.) *Media Technologies: Essays on Communication, Materiality, and Society*, Cambridge (MA), MIT Press, pp. 167-193.

Galison P. (1994), "The Ontology of the Enemy: Nobert Wiener and the Cybernetic Vision", *Critical Inquiry*, 21, pp. 228-266.

Galison P., Daston L. (2007), *Objectivity*, New York, Zone Books.

Grba D. (2022), "Deep Else: A Critical Framework for AI Art", *Digital*, 2, pp. 1-32. <https://doi.org/10.3390/digital2010001>

Grba D. (2024), "The Mechanical Turkness: Tactical Media Art and the Critique of Corporate AI", *arXiv*, arXiv:2402.17490v1.

Hayles K. N. (2024), "Detoxifying Cybernetics: From Homeostasis to Autopoiesis and Beyond", in Yuk Hui (eds.), *Cybernetics for the 21st Century. Vol. 1: Epistemological Reconstruction*, Hong Kong, Hanart Press, pp. 85-100.

Hepp A. (2020), *Deep Mediatization*, London, Routledge.

Jasanoff S., Kim S. H. (2015), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, Chicago, University of Chicago Press.

Knuth D. (1997), *The Art of Programming. Vol. 1 Fundamental Algorithms*, 3rd Edition, Boston, Addison-Wesley.

Lakoff G., Johnson M. (1981), *Metaphors We Live By*, Chicago, The University of Chicago Press.

Latour B. (1999), *Pandora's Hope. Essays On the Reality of Science Studies*, Cambridge & London, The MIT Press.



Leone M. (2023), "I compiti principali di una semiotica dell'intelligenza artificiale", in Leone M. e Santangelo A (ed.). *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 29-44.

Leone M. (2024), "Semiotics of the black box: on the rhetoric of algorithmic images", *Visual Communication*, 23(3), pp. 426-45, <https://doi.org/10.1177/14703572241247120>.

Lorusso A. M. (2011), "A Semiotic Approach to the Category of Imaginary", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 113-124.

Malavasi L. (2017), "Nella terra degli algoritmi e dei big data: incursioni artistiche nell'invisibile", *piano B. Arti e culture visive*, vol. 2 n. 2, pp. 68-86, doi: 10.6092/issn.2531-9876/7962.

Marin L. (1994), *De la représentation*, Paris, Éditions Seuil.

McCulloch W. S., Pitts W. (1943), "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biology*, vol. 5, pp. 115-133, <https://doi.org/10.1007/BF02478259>.

Mitchell W. J. T. (2002), "Showing Seeing: A Critique of Visual Culture", *Journal of Visual Culture*, 1(2):165-181, doi:[10.1177/147041290200100202](https://doi.org/10.1177/147041290200100202).

Munn L. (2022), *Automation Is a Myth*, Stanford, Stanford University Press.

Natale S. (2022), *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*. Torino, Einaudi.

Norbert W. (1961), *Cybernetics; or, Control and Communication in the Animal and the Machine*, 2nd ed., New York, The MIT Press.

Paglen T. (2016), "Invisible Images (Your Pictures Are Looking at You)", *The New Inquiry*, 8 dicembre, 2016. Consultato il giorno 5 maggio da: <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/>.

Paglen T., Downey A. (2020), "Algorithmic anxieties: Trevor Paglen in conversation with Anthony Downey", *Digi War*, 1, pp. 18-28, <https://doi.org/10.1057/s42984-020-00001-2>.

Paolucci C. (2010), *Strutturalismo e Interpretazione*, Milano, Bompiani.

Paolucci C. (2020), *Persona. Soggettività nel linguaggio e semiotica dell'enunciazione*. Milano, Bompiani.





- Paolucci C. (2021), *Cognitive semiotics. Integrating Signs, Minds, Meaning and Cognition*, New York, Springer.
- Pasquale F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press.
- Pasquinelli M. (2023), *The Eye of the Master: A Social History of Artificial Intelligence*, London & New York, Verso Books.
- Pasquinelli M., Joler V. (2021), "The Nooscope manifested: AI as instrument of knowledge extractivism", *AI & Society*, 36(4), pp. 1263-1280, <https://doi.org/10.1007/s00146-020-01097-6>
- Pavan M. (2022), "La creazione immaginaria. L'elemento immaginario di Cornelius Castoriadis." Rivista online *Fatamorgana*. Consultato il 27 aprile da: <https://www.fatamorganaweb.it/elemento-immaginario-di-cornelius-castoriadis/>
- Pedrazzi R. (2021), *Futuri possibili. Scenari d'arte e intelligenza artificiale*, Milano, Jaca Book.
- Patrick E. R. (2020), "Building the Black Box: Cyberneticians and Complex Systems", *Science, Technology, & Human Values*, 45(4), pp. 575-595, <https://doi.org/10.1177/0162243919881212>.
- Ponzo J. e Stano S. (2023), "Introduzione", in Ponzo J. e Stano S. (ed.) *Nuovi media. Nuovi miti*. Lexia 49, Roma, Aracne editore, pp. 11-20.
- Rancière J. (2008), *Le spectateur émancipé*, Paris, La Fabrique éditions.
- Roberge J., Seyfert R. (ed.) (2016), *Algorithmic Cultures. Essays on Meaning, Performance and New Technologies*, Abingdon & New York, Routledge.
- Ruiz C (2019), "Leading online database to remove 600,000 images after art project reveals its racist bias", *The Art Newspaper*. Consultato il 5 maggio da: <https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias>.
- Santangelo A. (2023), "Essere umani al tempo dell'intelligenza artificiale. Narrazioni a confronto.", in Leone M. e Santangelo A. (ed.), *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 239-266.
- Stano S. (2023), "Ripensare l'ideologia: un'introduzione.", in Leone M. e Stano S. (ed.) *Ideologia*, Lexia 41-42, Roma, Aracne editore, pp. 9-18.
- Sterling T., and van den Berg S. (2023), "Ukraine war shows urgency of military AI, Palantir CEO says", *Reuters*, consultato il 25 aprile da:



<https://www.reuters.com/technology/ukraine-war-shows-urgency-military-ai-palantir-ceo-says-2023-02-15/>

Tilak S., Glassman M., Kuznetcova I., Pelfrey L. (2022), "Applications of cybernetics to psychological theory: Historical and conceptual explorations", *Theory & Psychology*, 32(2), pp. 298–325, <https://doi.org/10.1177/0959354321105380>.

Trione V. (2022), *Artivismo: arte, politica, impegno*, Torino, Einaudi.

Volli U. (2011), "L'immaginario delle origini", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 31-62.

Von Hilgers P. (2011), "The History of the Black Box: The Clash of a Thing and its Concept", *Cultural Politics*, 7(1), pp. 41-58.

Voto C. (2022), "From archive to dataset. Visualizing the latency of big data", *Punctum. International Journal of Semiotics*, 08(01), pp. 47–62.

Wasielewski A. (2023), *Computational Formalism. Art History and Machine Learning*, Cambridge (MA), The MIT Press.

Zeilinger M. (2021), *Tactical entanglements. AI Art, Creative Agency, and the Limits of Intellectual Property*, Lüneburg, meson press.

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

Sitography

Future of Life Institute (2023) "Pause Giant AI Experiments: An Open Letter", *FLI Open Letters*, 23 marzo. Consultato il 5 maggio da: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

Palantir, sito ufficiale, sezione "defense", consultato il 25 aprile da: <https://www.palantir.com/offerings/defense/>.

IDF, pagina web dedicata alla presentazione del software Gospel, consultato il 25 aprile da: <https://www.idf.il>.

Barbican, "Trevor Paglen: On 'From Apple to Anomaly'". Consultato il 5 maggio da: <https://sites.barbican.org.uk/trevorpaglen/>.

Paglen T., "Imagenet Roulette". Consultato il 5 maggio da:



Lorenzo Aimo, Luca Brunet
Immaginario Black Box

<https://www.chiark.greenend.org.uk/~ijackson/2019/ImageNet-Roulette-cambridge-2017.html>.

Akten M., "Learning to See (series)". Consultato il 5 maggio da:
<https://www.memo.tv/works/learning-to-see/>.

Barbican, "AI: More Than Human", Press Release. Consultato il 5 maggio da:
<https://www.barbican.org.uk/sites/default/files/documents/2019-07/FINAL%20AI%20More%20than%20Human%20exhibition%20release%20280219.pdf>



