

# CULTURE VISUALI TRA UMANO E MACCHINA: I CAPTCHA E IL RUOLO DELL'IMMAGINE NELLE INTERAZIONI TRA UTENTI E SISTEMI ALGORITMICI

*Simone Natale*

## *Introduzione*

I CAPTCHA sono una misura di sicurezza utilizzata online per stabilire se la richiesta di accedere a un servizio o a una pagina arriva da un utente umano o da un programma informatico. Nonostante il termine non sia universalmente conosciuto, ogni utente con un'esperienza di navigazione significativa sul web avrà incontrato dei CAPTCHA al momento di creare o semplicemente di accedere al proprio account su una piattaforma di social media o su un qualsiasi sito. Nella sua forma attualmente più comune, l'utente è invitato a identificare in quali quadranti di un'immagine si scorge un determinato oggetto – per esempio, una bicicletta, un ponte o un semaforo. Solo dopo aver superato il test sarà in grado di continuare l'operazione, per esempio registrare un account su una piattaforma digitale (fig. 1).

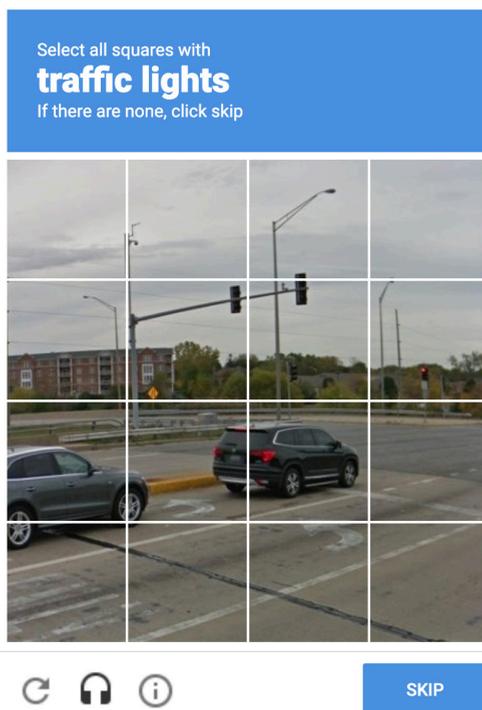


Fig. 1

La reazione della maggior parte di noi, quando incontriamo i CAPTCHA, è di indifferenza o fastidio: essi sono letteralmente degli ostacoli tra noi e l'operazione che desideriamo effettuare. Per questo a prima vista possono sembrare poco importanti, persino irrilevanti. In fondo, quando proviamo a capire le implicazioni della tecnologia nel mondo contemporaneo, ci concentriamo più spesso su oggetti apparentemente eccezionali, che stimolano la nostra immaginazione, come i robot o la realtà virtuale. Eppure, come hanno mostrato autori come Michael Billig, spesso le cose più banali e apparentemente irrilevanti nascondono le chiavi di lettura più fertili per capire la realtà in cui viviamo. Sono spesso le cose che non notiamo, che ci sfuggono proprio perché immerse e nascoste nelle pieghe delle nostre esperienze e abitudini quotidiane, a dirci qualcosa di più sul mondo contemporaneo<sup>1</sup>.

Nonostante, o forse proprio a causa della loro onnipresenza nella rete, i CAPTCHA rappresentano infatti un eccezionale punto di ingresso per comprendere alcune delle implicazioni più profonde delle relazioni tra umani e sistemi algoritmici nella società contemporanea. La loro origine va ricondotta alla fase di espansione del Web, i primi anni Duemila, in cui l'invadenza di programmi informatici in grado di effettuare operazioni online creò l'esigenza di strumenti appositi per riconoscere ed escludere tali programmi, privilegiando invece l'accesso degli utenti umani<sup>2</sup>. Se in un primo periodo molti sistemi di CAPTCHA richiedevano la trascrizione di un testo stampato, presto si aggiunsero altri tipi di quesiti, tra i quali l'immagine ha di recente assunto un ruolo centrale<sup>3</sup>.

Questo saggio si propone di mettere in questione l'apparente insignificanza dei CAPTCHA, per mostrare come essi ci permettano di precisare e capire aspetti spesso lasciati ai margini della discussione sulla cultura visuale nell'era del digitale<sup>4</sup>. Mi concentro, in particolare, su due aspetti cruciali. In primo luogo, come un Test di Turing alla rovescia (l'acronimo CAPTCHA significa infatti *Completely Automated Public Turing test to tell Computers and Humans Apart*), i CAPTCHA richiamano le crescenti difficoltà di distinguere tra umani e macchine. Tale difficoltà è destinata a diventare sempre più cruciale a mano a mano che le tecnologie di Intelligenza Artificiale (IA) acquisiscono la capacità di simulare forme di comunicazione tradizionalmente attribuite a esseri umani. In questo contesto le culture e competenze visuali sono destinate ad assumere un ruolo sempre più decisivo. In secondo luogo, i CAPTCHA sono oggi non solo una misura di sicurezza ma anche un sistema di addestramento utile a migliorare i software di

1 M. Billig, *Nazionalismo banale*, Rubbettino, Soveria Mannelli 2018 [ed. or. *Banal Nationalism*, Sage, London 1995]. Si veda anche W.H.K. Chun, *Updating to Remain the Same: Habitual New Media*, MIT Press, Cambridge (Mass.) 2016.

2 L.V. Ahn, M. Blum, N.J. Hopper, J. Langford, *CAPTCHA: Using hard AI problems for security*, in E. Bihani (a cura di), *Advances in Cryptology — EUROCRYPT 2003. EUROCRYPT 2003. Lecture Notes in Computer Science*, vol. 2656, Berlin, Heidelberg, Springer 2003, pp. 294-311.

3 V. P. Singh, P. Pal, P. *Survey of different types of CAPTCHA*, "International Journal of Computer Science and Information Technologies", 5.2, 2014, pp. 2242-2245. Tra le diverse tipologie di CAPTCHA, vi sono anche CAPTCHA basati su video; questa tipologia potrebbe diventare sempre più frequente mano a mano che gli algoritmi di riconoscimento visuale diventano più efficienti, rendendo impotenti e dunque obsoleti i CAPTCHA basati sulle immagini.

4 Per un approccio che, in sintonia con questo saggio, coniuga gli strumenti degli studi culturali sull'immagine a un'enfasi sugli aspetti tecnici e materiali dei dispositivi digitali, si veda R. Eugeni, *Capitale algoritmico. Cinque dispositivi postmediali (più uno)*, Morcelliana, Milano 2021.

riconoscimento visivo basati sulle reti neurali. In questo senso essi ci stimolano a considerare come le competenze visuali di singoli utenti vengano mobilitate da sistemi algoritmici al fine di fornire i dati necessari a riprodurre meccanicamente queste stesse competenze.

### 1. Rovesciando il test di Turing

Nel 1950, Alan Turing pubblicò un articolo sulla rivista *Mind* destinato a diventare una delle pubblicazioni più lette e citate sull'intelligenza artificiale. L'articolo cominciava con la seguente domanda: "Le macchine possono pensare?" Considerato il dibattito che sarebbe emerso intorno a tale questione nei decenni successivi e che prosegue fino ai giorni nostri, tale domanda non era certo insignificante. Ma Turing, che amava i paradossi, dedicò le righe successive dell'incipit a smontare proprio la domanda con cui aveva aperto l'articolo. Dal momento che non siamo in grado, ragionò il matematico inglese, di trovare un accordo su cosa significa il termine "pensare", porre un tale quesito si rivela completamente inutile. Turing propose dunque di mettere da parte quella domanda sostituendola con un esperimento pratico, il gioco dell'imitazione, oggi noto come test di Turing<sup>5</sup>.

Nel test di Turing, un'interrogatrice o interrogatore umano (rappresentato dalla lettera C nel diagramma di fig. 2) scambia messaggi scritti con un interlocutore sconosciuto. L'interrogatrice deve stabilire, a partire dal contenuto della conversazione sostenuta, se si tratta di un essere umano (B) o di un programma informatico (A). Un computer avrà passato il test se capace di passare per essere umano con una frequenza statisticamente rilevante.

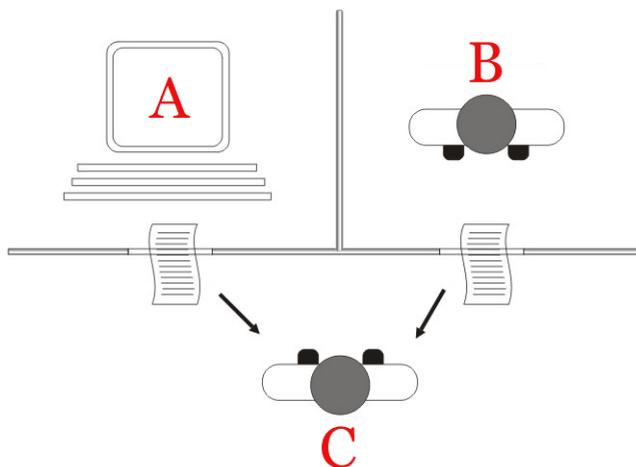


Fig. 2

<sup>5</sup> A. Turing, *Computing Machinery and Intelligence*, "Mind", 59.236, 1950, pp. 433-460 (trad. it., *Macchine calcolatrici e intelligenza*, in id., *Intelligenza meccanica*, trad. di Gabriele Lolli e Nino Dazzi, Bollati Boringhieri, Torino 1994, pp. 121-157).

Il test di Turing viene spesso discusso come se fosse in grado di decidere se l'intelligenza artificiale sia raggiunta: passare il test equivarrebbe dunque a stabilire che è possibile creare un computer "intelligente". Ma questa interpretazione non corrisponde alla proposta originale di Turing, che si era premurato di mettere da parte la domanda sulle macchine pensanti *sostituendola* con il test – e non di usare il test per trovare una risposta alla domanda. Considerare il test di Turing da un tale punto di vista non significa soltanto interpretarlo in maniera fuorviante, ma anche lasciarsi sfuggire le sue implicazioni più importanti. Assegnando a un interrogatore umano la responsabilità di valutare il comportamento della macchina, Alan Turing rifiutò infatti di definire l'IA in termini assoluti per concentrarsi sul modo in cui gli esseri umani percepiscono e comprendono l'IA: sono gli attori umani a decidere se la macchina passa il test, attribuendole o meno l'apparenza di umanità. Il vero "messaggio" del test di Turing, dunque, non ha a che fare con l'intelligenza delle macchine, ma piuttosto con le modalità con cui noi umani ci relazioniamo a loro. La questione non è se le macchine siano in grado o meno di pensare. È se noi umani siamo disposti a credere che lo facciano – in altre parole, se siamo disposti ad accettare il comportamento delle macchine come intelligente<sup>6</sup>.

Nella loro qualità di Test di Turing alla rovescia, volto a identificare gli esseri umani piuttosto che a misurare la capacità di mimetizzazione delle macchine, anche i CAPTCHA ci stimolano a considerare la relazione tra umani e macchine. Il motivo per cui i CAPTCHA sono diventati onnipresenti nella rete, infatti, è che essa è popolata da una quantità di programmi informatici che, come fantasmi, non solo contribuiscono a far funzionare l'infrastruttura della rete, ma in certi casi provano anche a sostituirsi, a spacciarsi, per utenti umani<sup>7</sup>. La nostra capacità di identificare questi programmi come tali rappresenta perciò un requisito fondamentale per assicurare la funzionalità di sistemi che utilizziamo nel nostro quotidiano, dalle piattaforme di social media alle e-mail, dalle banche dati della sanità ai servizi bancari online.

In questo senso, la più importante differenza tra il test di Turing e il suo doppio-opposto, il CAPTCHA, risiede nella posta in gioco. Mentre il test di Turing era un esercizio in fondo innocuo, appunto un "gioco" dell'imitazione secondo l'appellativo originariamente scelto da Turing, o al limite una risorsa per misurare la funzionalità dei sistemi di intelligenza artificiale applicati alla produzione linguistica, il CAPTCHA diviene necessario nella misura in cui ci si deve proteggere da veri e propri attacchi informatici portati avanti da sistemi programmati per effettuare un login, creare nuovi account o utilizzare determinate risorse di rete. La loro presenza richiama, in questo senso, un elemento di carattere bellico insito nei rapporti tra umani e macchine che la società delle reti sta contribuendo a plasmare.

<sup>6</sup> Per una discussione più estesa del test di Turing e delle sue implicazioni, si veda S. Natale, *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, Einaudi, Torino 2022 [ed. or. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, Oxford University Press, New York 2021].

<sup>7</sup> F. McKelvey, *Internet Daemons: Digital Communications Possessed*. University of Minnesota Press, Minneapolis 2018.

Parlare di guerra o conflitto tra umani e macchine può sembrare a prima vista sproporzionato o del tutto fuori luogo, e richiede una precisazione importante. Non si tratta qui di uno scenario alla Terminator, in cui un'entità informatica muove guerra agli esseri umani per sterminarli o dominarli. Nonostante questa eventualità sia considerata da alcuni come possibile o persino plausibile<sup>8</sup>, ciò che abbiamo di fronte nel presente e perlomeno nel prossimo futuro è un elemento di conflitto con le macchine più sottile ma non per questo meno denso di conseguenze. Dietro i programmi informatici che cercano di sottrarsi alla sorveglianza dei CAPTCHA, infatti, ci sono sempre dei programmatori o degli utilizzatori: per quanto il programma sia in grado di operare in maniera autonoma, è sufficiente ricostruire la sua catena di azioni abbastanza in profondità per giungere a un certo punto all'attore o attrice umana che lo ha programmato o messo in azione. Non si tratta, quindi, di un conflitto che mette a confronto umani e macchine come due fazioni distinte, secondo una dinamica che la fantascienza ci ha abituati a considerare, quanto di una concatenazione di atti potenzialmente ostili in cui macchine e umani al tempo stesso collaborano e collidono<sup>9</sup>.

È significativo, in questo senso, come l'immagine sia divenuta il mezzo fondamentale con cui i CAPTCHA garantiscono il servizio di *polizia*, relativo alla sicurezza, e al tempo stesso di *pulizia* dallo spam e da tutto ciò che non è utile alla circolazione di informazione sulla rete internet. Mentre il test di Turing metteva il linguaggio e la competenza verbale al suo centro, la gran parte dei sistemi di CAPTCHA che oggi troviamo online chiedono agli utenti di mettere a frutto la propria competenza visuale per dimostrare di essere creature in carne e ossa<sup>10</sup>. Paradossalmente, d'altronde, è la nostra competenza visiva a doversi adattare agli strumenti e alle tecnologie del mezzo informatico, piuttosto che viceversa. Siamo noi umani a doverci avvicinare al linguaggio delle macchine. Non ci viene chiesto, per fare un esempio, di descrivere un'immagine, lasciando spazio alla variazione e alla complessità che caratterizza la nostra relazione e la nostra esperienza con il visuale. Ciò che ci viene chiesto è di etichettarla in termini discreti, persino binari: dobbiamo decidere se in una determinata fotografia compare una bicicletta o non compare, se c'è un ponte o non c'è. Non c'è spazio per chiedersi che cosa sia una bicicletta o cosa sia un ponte – se un monociclo può rientrare nell'insieme delle biciclette, se sotto un ponte deve esserci per forza un corso d'acqua o se anche un cavalcavia può essere considerato tale. O meglio, lo spazio c'è, ma la risposta deve adattarsi alla natura binaria del quesito: sì o no, bicicletta o non bicicletta, ponte o non ponte.

8 Si veda per esempio il problematico, ma interessante, N. Bostrom, *Superintelligenza. Tendenze, pericoli, strategie*, Bollati Boringhieri, Torino 2018 [ed. or. *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford 2014].

9 Ci troviamo cioè davanti a un contesto comunicativo in cui umani e macchine non sono in opposizione tra loro in quanto appartenenti alla specie umana o al mondo delle macchine, ma semplicemente perché le macchine sono diventati attori sociali e il conflitto è parte della società umana. Su questi temi si veda anche M. Airoidi, *Machine Habitus: Toward a Sociology of Algorithms*, Polity, New York 2021.

10 M. Foley, "Prove You're Human": *Fetishizing Material Embodiment and Immaterial Labor in Information Networks*, "Critical Studies in Media Communication", 31. 2014, pp. 365-379.

Uno dei principi fondamentali dell'interazione umano-computer, del resto, stabilisce che la comunicazione tra computer ed esseri umani sia possibile nella misura in cui il computer è in grado di “parlare” il linguaggio dell'utente<sup>11</sup>. Questo non significa, ovviamente, che per creare un'interazione con l'utente i computer debbano *parlare* in linguaggio naturale, come gli assistenti vocali Siri o Alexa<sup>12</sup>. È sufficiente che utente e computer condividano un ambiente comunicativo, costruito attraverso interfacce hardware o software, presso il quale sono in grado di avviare una certa modalità di interazione. Per esempio, un programma di scacchi permette a un utente umano di giocare contro un programma informatico perché essi condividono le condizioni di interazione date dalla scacchiera, dagli spostamenti potenziali di ogni pezzo, dalle regole del gioco e così via. Di conseguenza, l'interazione umano-computer si è storicamente sviluppata secondo delle modalità che riducevano la ricchezza percettiva e comunicativa degli esseri umani per limitarsi a una sua versione più limitata, ma che fosse gestibile dal computer – come gli scacchi, ma anche l'uso di comandi testuali, o l'interazione con tastiera e mouse, per fare solo degli esempi. Lo stesso test di Turing rappresenta qualcosa di simile: per mettere in comunicazione computer e utenti umani Turing scelse di utilizzare il linguaggio scritto, mettendo quindi da parte la voce e ripulendo la comunicazione da ogni traccia di corporalità: ed è proprio perché il computer non ha bisogno di un corpo per “giocare” al test di Turing che è relativamente semplice creare un programma per intrattenere una conversazione di questo tipo con un utente umano (anche se è molto più difficile crearne uno in grado di intrattenere una conversazione credibile)<sup>13</sup>.

I CAPTCHA visuali si collocano saldamente all'interno di questa tradizione: creano uno spazio di interazione in cui non è il computer a riprodurre o imitare la complessità della cultura visiva umana, ma piuttosto è l'essere umano ad adattarsi ai limiti dell'interazione umano-computer – per la quale, almeno in questo contesto, ciò che conta è riconoscere in termini assoluti e discreti (sì/no, 0/1) che cosa appaia nell'immagine.

In questo senso, i CAPTCHA ci stimolano a riconoscere la natura dialettica della cultura visiva digitale. Da una parte, tecnologie come la CGI, la realtà virtuale, ma anche solo i processi algoritmici che ci fanno apparire le fotografie scattate dai nostri smartphone così ricche a livello visivo, spingono le tecnologie digitali verso un'emulazione sempre più precisa della ricchezza del nostro mondo percettivo<sup>14</sup>. Dall'altra, i CAPTCHA ci rivelano un'altra tendenza, apparentemente opposta eppure altrettanto significativa, per la quale noi umani siamo chiamati a utilizzare le nostre facoltà visive per effettuare operazioni più vicine alla

11 J. C. R. Licklider, and W. T. Robert. *The Computer as a Communication Device*, “Science and Technology”, 2.3, 1968, pp. 2-5.

12 Sulla voce come strumento di comunicazione e interazione tra computer e utenti umani, si veda D. Napolitano, *La voce artificiale. Un'indagine media-archeologica sul computer parlante*, Editoriale Scientifica, Napoli 2022.

13 Il primo programma di questo tipo, il chatbot ELIZA, fu messo a punto già negli anni Sessanta del Novecento. S. Natale, *If Software Is Narrative: Joseph Weizenbaum, Artificial Intelligence and the Biographies of ELIZA*, “New Media & Society” 21.2019, pp. 712-728.

14 A. Pinotti, *Alla soglia dell'immagine. Da Narciso alla realtà virtuale*, Einaudi, Torino 2021. Su algoritmi e fotografia digitale, si veda anche S. Taffel, *Google's Lens: Computational Photography and Platform Capitalism*, “Media, Culture & Society” 43.2, 2021, pp. 237-255.

logica binaria e computazionale<sup>15</sup>. In questa direzione, infatti, non vanno solo i CAPTCHA ma anche una quantità di sistemi algoritmici deputati all'analisi di immagine: si pensi ai programmi di intelligenza artificiale addestrati a riconoscere un oggetto specifico all'interno di un corpus di immagini, a riconoscere i segni di una potenziale patologia in un'immagine diagnostica, o a descrivere con una breve didascalia il contenuto di una fotografia.

Ecco che i CAPTCHA, a prima vista un oggetto insignificante, qualcosa da sbrigare il più in fretta possibile per arrivare alla schermata successiva, ci aiutano a intravedere una caratteristica spesso ignorata della cultura visuale nell'età del digitale. Lunghi dall'essere orientata esclusivamente nella direzione di una mimesi sempre più accurata, l'incontro tra umani e macchine nella nostra epoca (e le culture visive emergenti che ne risultano) si sta configurando come un meccanismo più complesso, la cui direzione non è univoca ma multidirezionale: se le macchine sono progettate per imitare in maniera sempre più precisa la percezione visiva umana, allo stesso tempo gli umani sono incoraggiati ad adattarsi alla logica di funzionamento del computer digitale, a "vedere" le cose come le vedono i computer. Se questo processo può sembrare divergente, si tratta in realtà di una convergenza, dal momento che entrambe le traiettorie portano alla fine a un punto di incontro tra umano e macchina: ovvero a uno spazio in cui computer e utenti condividono un approccio comune all'immagine e sono così nella posizione per interagire in uno stesso spazio comunicativo.

## 2. *Vide et labora: addestrare le macchine con i CAPTCHA*

Il tema delle relazioni tra umano e macchina rappresenta solo uno degli aspetti per cui i CAPTCHA si rivelano un oggetto decisivo a comprendere la cultura visiva della contemporaneità digitale. Un secondo ordine di questioni ha a che fare con la maniera in cui essi si configurano non solo come dispositivi di sicurezza, ma anche come una sorta di grande palestra distribuita nella quale il micro-lavoro degli utenti viene mobilitato per addestrare gli algoritmi di intelligenza artificiale a riconoscere determinate immagini.

Per spiegare come questo sia possibile è necessario soffermarsi brevemente sul funzionamento degli algoritmi di *deep learning*, o apprendimento profondo, che stanno alla base del successo di una larga parte dei sistemi di IA oggi in uso. A differenza dei tradizionali sistemi di IA, che puntavano a descrivere con mezzi simbolici operazioni logiche che potevano essere elaborate dai computer, le reti neurali o deep learning vengono "addestrati" attraverso l'esposizione a banche dati con le quali "imparano" autonomamente ad assolvere un compito o funzione<sup>16</sup>. Perché l'addestramento funzioni, però, è necessario un corpus di

<sup>15</sup> Come notato da Foley, i CAPTCHA infatti elevano i computer al rango di giudici, al contrario dei test di Turing in cui questo ruolo era deputato a utenti umani. M. Foley, "Prove You're Human": *Fetishizing Material Embodiment and Immaterial Labor in Information Networks*, cit.

<sup>16</sup> I termini usati sono quelli del linguaggio tecnico sviluppato nel campo dell'IA. L'utilizzo di metafore legate a facoltà e azioni associate all'umano è sempre stato preponderante in questo contesto, e ha un ruolo significativo nella persistenza di esagerazioni e mistificazioni sull'"intelligenza" artificiale. Si veda a riguardo H.R. Ekbia, *Artificial Dreams: The Quest for Non-Biological Intelligence*, Cambridge University Press, Cambridge 2008.

dati estremamente ampio. Se si vuole sviluppare un software per riconoscere immagini in cui compare un ponte, per esempio, si partirà da un database di migliaia se non milioni di immagini, nelle quali sono state etichettate quelle in cui compare un ponte. L'algoritmo di deep learning opererà autonomamente complessi calcoli statistici per stabilire cosa accomuna le immagini contrassegnate nel database come ponti; a seconda dell'ampiezza e della qualità del corpus di dati utilizzato, il sistema diventerà in grado di stabilire con minore o maggiore affidabilità se una nuova immagine contenga un ponte o meno<sup>17</sup>.

La potenza di questi sistemi ha portato a sviluppare, nel giro di pochi anni, sistemi in grado di riconoscere potenziali patologie in immagini diagnostiche o di redigere didascalie a partire da singole immagini, ma anche, al di là del contesto specifico del visuale, sistemi di guida autonoma, strumenti di traduzione automatica e le voci sintetiche di assistenti vocali come Alexa e Siri. Il prerequisito per creare algoritmi di deep learning efficaci è però, come detto, la presenza di database sufficientemente ampi per portare a compimento il processo di addestramento<sup>18</sup>. Il problema è che non basta avere a disposizione una banca dati – per esempio un corpus costituito da un milione di immagini. Per renderlo utilizzabile ai fini dell'addestramento, questo corpus deve essere appropriatamente organizzato: per esempio, nel caso dell'algoritmo programmato per riconoscere immagini di ponti, le immagini di questo tipo presenti nel corpus devono essere etichettate e distinte dalle immagini che invece ne sono prive. Solo così l'algoritmo sarà in grado di “imparare” cosa caratterizza le immagini di ponti, e quindi diventerà in grado di riconoscerne di nuove<sup>19</sup>.

È per questo che lo sviluppo di sistemi sempre più efficienti di deep learning, specialmente ma non esclusivamente nel campo del riconoscimento delle immagini, è spesso legato a un sistema di “sfruttamento consensuale” del microlavoro degli utenti, di cui i CAPTCHA sono un esempio estremamente rappresentativo<sup>20</sup>. La maggior parte dei CAPTCHA che incontriamo sono sviluppati da reCAPTCHA, uno strumento di proprietà di Google che è implementato da innumerevoli piattaforme e siti sulla rete. A partire dalla sua acquisizione nel 2009, reCAPTCHA ha significato per Google la possibilità di sviluppare immense banche dati con le quali l'azienda di Cupertino ha addestrato i propri sistemi di deep learning. In una prima fase, reCAPTCHA proponeva quesiti basati sul riconoscimento di caratteri: il microlavoro degli utenti che risolvevano i problemi posti dai CAPTCHA di Google era utilizzato per migliorare i sistemi di trascrizione dei testi scannerizzati nella banca dati di

17 J.D. Kelleher, *Deep Learning*, MIT Press, Cambridge (Mass.) 2019.

18 Per questa ragione grandi colossi come Google, Meta, Amazon e Apple hanno un vantaggio strategico sulla concorrenza: i dati prodotti dalla massa enorme di utenti che utilizzano i loro servizi possono essere utilizzati per sviluppare sistemi sempre più precisi basati sul deep learning.

19 Recentemente, sono state sviluppate tecnologie di deep learning che prescindono da questa operazione di etichettamento; per esempio, *language models* come ChatGPT sono addestrati attraverso l'esposizione diretta a enormi masse di dati testuali. Si tratta di un'innovazione decisiva per il campo dell'intelligenza artificiale, che ha portato aziende come OpenAI e Google a sviluppare sistemi estremamente performanti anche in campi come gli agenti conversazionali o chatbot, che finora erano stati toccati in maniera solo marginale dal deep learning proprio a causa della difficoltà di organizzare ed etichettare i dati con cui addestrarli.

20 R. Mühlhoff, *Human-Aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Toward a Media Sociology of Machine Learning*, “New Media & Society”, 22.10, 2020, pp. 1868-1884. Si veda anche M. Airoidi, *Machine Habitus: Toward a Sociology of Algorithms*, Polity, Cambridge 2021.

Google Books, all'epoca in forte espansione. Più di recente, i quesiti di reCAPTCHA si sono concentrati sul riconoscimento di oggetti come biciclette, automobili, semafori, cartelli stradali: non una scelta casuale, perché i dati resi disponibili dalle selezioni degli utenti sono utili allo sviluppo di sistemi di guida automatica, che hanno necessità di identificare in maniera il più attendibile possibile tali oggetti<sup>21</sup>.

Con i CAPTCHA, dunque, le competenze visuali di singoli utenti vengono mobilitate al fine di fornire i dati necessari a riprodurre meccanicamente queste stesse competenze. È stato stimato che la collettività degli utenti di internet impieghi a risolvere CAPTCHA un tempo complessivo equivalente a circa cinquecento anni ogni giorno<sup>22</sup>. Il motivo per cui sistemi alternativi, che evitino una tale perdita di tempo agli utenti, difficilmente verranno sviluppati nel prossimo futuro, è che questo tempo è “perso” soltanto dal punto di vista degli utenti: per aziende come Google si tratta invece di un tempo utile, ampiamente monetizzato attraverso il miglioramento di algoritmi proprietari deputati al riconoscimento visivo.

Vale la pena di riflettere sulle implicazioni di questo processo<sup>23</sup>. Una prima conseguenza significativa risiede nella dinamica di apparente filiazione, per cui gli algoritmi di IA, basandosi sul microlavoro di utenti umani, imparano a riprodurre le competenze visive. Anche in questo caso, però, si tratta di un tipo specifico di visione, che non è davvero rappresentativo dell'esperienza visiva di un essere umano. Gli algoritmi di reCAPTCHA, infatti, non imparano a “osservare” un paesaggio: più strettamente, essi imparano a riconoscere determinati oggetti che sono significativi per assolvere il compito a cui questi algoritmi sono deputati. Siamo di fronte, come già notato, a una scelta discreta, spesso persino binaria, per la quale l'umano si adatta ai meccanismi profondi del computer digitale: questo è un ponte oppure non è un ponte. L'ironia aggiuntiva è che gli utenti sono costretti a operare, per passare il CAPTCHA e procedere con la loro navigazione, un tipo di operazioni ripetitive che non solo si adatta al *modus operandi* del computer digitale, ma ne nutre lo sviluppo contribuendo ad addestrare gli algoritmi di deep learning.

Questa situazione ci ricorda che, nonostante il termine *robot* venga dal termine ceco che significa schiavo, c'è un'ambivalenza o forse una reciprocità in questa schiavitù: le macchine sono al nostro servizio ma occasionalmente anche noi ci mettiamo al servizio delle macchine. I CAPTCHA in questo senso potrebbero essere visti come una continuazione dell'esperienza disumanizzante della fabbrica fordista, trasformata in icona dal Charlie Chaplin di *Tempi Moderni* (*Modern Times*, C. Chaplin, 1936), dove i gesti continui e ri-

21 Alphabet, la compagnia che fa capo a Google, è infatti nel processo di sviluppare sistemi di guida automatizzata attraverso un'azienda sussidiaria chiamata Waymo. N. Rivero, Why AI developers love those annoying CAPTCHAs, Quartz, 17 Maggio 2021, <https://qz.com/2009829/cloudflares-captcha-proposal-would-end-ais-source-of-free-labor/>

22 T. Meunier, *Humanity wastes about 500 years per day on CAPTCHAs. It's time to end this madness*, “The Cloudflare Blog”, 13 Maggio 2021, <https://blog.cloudflare.com/introducing-cryptographic-attestation-of-personhood/>

23 Di cui, come detto, i CAPCHA non rappresentano che un caso esemplare: la monetizzazione del microlavoro svolto dagli utenti nella rete è infatti uno dei meccanismi fondamentale che regolano il funzionamento di tutte le maggiori piattaforme, inclusi i social media assieme a innumerevoli altri sistemi, siti e servizi online. Si veda a riguardo D. W. Hill, *The Injuries of Platform Logistics*, “Media, Culture & Society”, 42.4, 2020, 521-536. <https://doi.org/10.1177/0163443719861840C>. Fuchs, *Social Media: A Critical Introduction*, Sage, Londra 2014.

petitivi sono parcellizzati e distribuiti presso una vasta collettività di utenti ignari del loro status di micro-lavoratori nella fabbrica dell’algoritmo.

Una seconda implicazione di queste dinamiche ha a che fare con il funzionamento tecnico che si nasconde dietro le competenze visive sviluppate dagli algoritmi di deep learning. Nonostante ci sia una derivazione tra le competenze visuali degli utenti e quelle degli algoritmi, dal momento che è attraverso il contributo degli utenti che il processo di addestramento delle reti neurali diviene possibile, permane una differenza profonda tra come noi vediamo un’immagine e come la “vede” una macchina. Le virgolette sono d’obbligo, poiché la macchina non vede niente: quello che fa è operare un’analisi condotta attraverso modelli statistici così complessi da rimanere opachi persino agli stessi programmatori che sviluppano questi sistemi<sup>24</sup>.

È nella tradizione della statistica, in questo senso, prima che in quella della cultura visiva e artistica, che si collocano gli algoritmi di deep learning addestrati dai nostri CAPTCHA, così come molte altre tecnologie digitali contemporanee della visione. Si pensi ai complessi algoritmi che permettono ai nostri smartphone, sempre attraverso calcoli statistici, di scattare fotografie di alta qualità pur senza l’ausilio del dispositivo che tradizionalmente definiva la *performance* di una macchina fotografica, ovvero la lente ottica<sup>25</sup>. Si considerino, inoltre, i recenti tentativi di sviluppare programmi informatici in grado di produrre opere d’arte, come AI-CAN, che hanno spesso stimolato affermazioni enfatiche (e affrettate) sulla fine dell’arte, senza considerare che la procedura di creazione di tali immagini è basata sul deep learning e derivata essenzialmente da modelli statistici piuttosto che artistici o concettuali<sup>26</sup>.

Studiosi di cinema e cultura visiva come André Gaudreault e Philippe Marion hanno usato il concetto di “serie culturale” per distinguere specifiche tradizioni culturali che regolano usi, interpretazioni e dinamiche profonde di un medium. Questo strumento concettuale consente di ricostruire «un tema, una pratica culturale, un tipo di spettacolo o di rappresentazione associati a un dispositivo o apparato nel tentativo di tracciare e comprendere la formazione dell’identità di un mezzo»<sup>27</sup>. I meccanismi di riproduzione e analisi delle immagini sotto l’egida del deep learning, in questo contesto, si collocano solidamente nella serie culturale della statistica e della computazione. I CAPTCHA ci ricordano dunque come le culture visive che stanno emergendo intorno a nuove tecnologie algoritmiche e digitali siano da

24 A meno che questi programmatori non usino appositi software per analizzare le operazioni condotte dagli algoritmi. Sulla questione dell’opacità, si veda J. Burrell, *How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms*, “Big Data & Society” 3, 2016, <https://doi.org/10.1177/2053951715622512>.

25 S. Taffel, *Google’s lens: Computational photography and platform capitalism*, “Media, Culture & Society”, 43, 2021, pp. 237-255.

26 In un saggio recente scritto con Leah Henrickson, abbiamo proposto di rovesciare il dibattito sull’arte generata da sistemi di IA, considerando non tanto gli oggetti in sé quanto lo sguardo di chi li vede e tutti i meccanismi che ci fanno interpretare immagini prodotte da sistemi informatici come oggetti d’arte: per esempio, la loro collocazione in gallerie e istituzioni artistiche, l’uso di cornici in cui le immagini sono collocate, l’identificazione del nome del programma come “firma” autoriale. Si veda S. Natale, L. Henrickson, *The Lovelace effect: Perceptions of creativity in machines*, “New Media and Society” 26, 2024, pp. 1909-1926.

27 A. Gaudreault, P. Marion, *The Double Birth Model Tested against Photography*, in N. Leonardi, S. Natale (a cura di), “Photography and Other Media in the Nineteenth Century”, Penn State University Press, University Park 2018, p. 201.

contestualizzare non solo nella traiettoria storica della cultura analogica fotografica e cinematografica, ma anche in quella dei numeri e della loro manipolazione tecnica.

### *Conclusione*

L'onnipresenza dei CAPTCHA, come suggerito nell'introduzione, è spesso scambiata paradossalmente per una prova della loro insignificanza. Raramente ci soffermiamo a pensare alle cose più ordinarie che incontriamo nel nostro quotidiano, e la stessa cosa accade nella nostra esperienza degli spazi della rete. Eppure oggetti come i CAPTCHA, come ho cercato di mostrare in questo saggio, possono aiutarci a capire molto di più sull'universo dell'immagine e dei media rispetto a tecnologie più appariscenti e apparentemente futuribili. La centralità delle competenze visuali in questo contesto ha creato l'opportunità di usare i CAPTCHA come una chiave per riesaminare alcune degli sviluppi della cultura visiva nell'epoca di internet, del digitale e delle tecnologie di IA.

Per tornare a una delle ultime questioni discusse in questo saggio, l'identificazione della statistica come serie culturale di una parte crescente della "cultura visuale" del digitale ci incita a riconfermare la distinzione tra la nostra relazione con le immagini e la relazione che le macchine hanno con le immagini. La metafora della vista, utilizzata in espressioni oggi di uso comune in informatica come "*computer vision*", tende a oscurare questa distinzione profonda e a farci dimenticare che non esistono al momento macchine in grado di replicare o persino di avvicinarsi all'esperienza umana di visione. Il ruolo dei CAPTCHA nelle piccole interazioni quotidiane sulla rete è dimostrazione di quanto sia diventato sempre più decisivo, d'altronde, mantenere la capacità tecnica e culturale per distinguere l'operato delle macchine da quello degli esseri umani. Altrove, ho mostrato come le tecnologie di IA attivino una serie di strumenti per assomigliare agli esseri umani, mobilitando per esempio le sensazioni che sollecita l'ascolto di una voce apparentemente umana per apparirci più credibili e "intelligenti"<sup>28</sup>. I CAPTCHA rappresentano una direzione complementare a queste dinamiche: test di Turing alla rovescia, sono utilizzati per garantire che le macchine siano riconosciute come tali ed escluse da servizi e sistemi deputati a utenti umani.

Proprio perché le tecnologie di IA sono in continua evoluzione, i CAPTCHA sono condannati a un continuo inseguimento: più le capacità di riconoscimento visivo delle macchine diventano sofisticate, più diventa necessario creare quesiti più complessi che solo utenti umani sarebbero in grado di risolvere<sup>29</sup>. Ci troviamo quindi di fronte a un corto circuito per certi versi paradossale: gli utenti risolvono quesiti per dimostrare che non sono macchine; risolvendo questi quesiti, contribuiscono all'addestramento delle IA deputate all'analisi e al riconoscimento delle immagini; migliorando le performance delle IA in questo campo,

<sup>28</sup> S. Natale, *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, cit.

<sup>29</sup> L. Fortunati, A.M. Manganelli, F. Cavallo, F. Honsell, *You Need to Show That You Are Not a Robot*, "New Media & Society" 21, 2019, pp. 1859-1876.

contribuiscono a rendere gli attuali CAPTCHA rapidamente obsoleti. Se proiettiamo questa dinamica verso il futuro, possiamo immaginare uno scenario in cui i CAPTCHA non funzionano più. Quando le macchine avranno raggiunto capacità di riconoscimento del contenuto delle immagini comparabili a quelle degli utenti umani – cosa sempre più plausibile, considerati i continui progressi delle tecnologie di deep learning e dell'IA generativa in questi e altri campi<sup>30</sup> –, i CAPTCHA non saranno infatti in grado di assolvere la propria funzione. Le implicazioni di un tale scenario possono apparire insignificanti e banali come gli stessi CAPTCHA, ma in realtà nascondono implicazioni importanti. Lo scenario in cui le tecnologie di riconoscimento delle immagini si sono sviluppate al punto che i CAPTCHA non funzionano più, infatti, corrisponde a un futuro in cui avremo perso parte della nostra capacità di distinguere tra macchine ed esseri umani.

<sup>30</sup> *Large language models* come ChatGPT e *text-to-image* come DALL-E rappresentano due esempi particolarmente evidenti dei progressi in questo campo.