

Algo(r)etica e immigrazione

Fabio Ciraci

Prima di entrare nel cuore del tema che mi propongo di trattare, occorre brevemente motivare la scelta del contributo e spiegare, in seguito, il titolo del mio intervento. La quinta edizione delle *Giornate di studio sul razzismo* si è svolta in buona parte grazie al contributo del ONLUS *G.U.S. Gruppo Umana Solidarietà* attraverso il progetto “G.E.A. - Global, Green, Generative and Equal Educational Activities”, finanziato dall’Agenzia Italiana per la Cooperazione allo Sviluppo, al quale partecipano in qualità di partner sia il Centro in *digital humanities* di Unisalento, che mi onoro di dirigere, sia l’Unità di ricerca IntegroLAB/Scuola Civica del Dipartimento di Scienze Giuridiche di Unisalento.

Il progetto prevede una serie di attività condivise, tra le quali l’attivazione di un Assegno di ricerca annuale su “Le discriminazioni prodotte dalla profilazione algoritmica” ovvero lo studio delle discriminazioni generate dai sistemi di classificazione automatica che mette capo a forme di discriminazione implicite, oggettivate o istituzionalizzate. Si tratta di strumenti informatici che sono solo apparentemente neutrali, che mettono in atto processi classificatori e identificativi esercitati sugli individui, espletati in maniera automatica da algoritmi. Un algoritmo traduce in termini matematici e poi in codice digitale una procedura, definendo i passaggi che devono essere eseguiti, in sequenza, secondo un certo diagramma di flusso. Facciamo un esempio banale: in base ai dati relativi a un determinato soggetto (età, residenza, genere, stato di salute) un algoritmo definisce il rischio relativo alla possibilità di sinistri automobilistici sulla base del quale calcolare la polizza assicurativa. Questo è il motivo per il quale il medesimo soggetto paga una polizza assicurativa diversa se cambia città di residenza oppure se cambia la sua età. Per esempio, pagherà una polizza assicurativa maggiore a Napoli e minore a Pavia, perché il rischio di sinistri è superiore nella città partenopea.

Nel loro complesso, gli algoritmi utilizzati nell’analisi predittiva stimano la probabilità o il rischio che un evento futuro, ad esempio la probabilità che un individuo paghi un mutuo regolarmente; così come la probabilità che un individuo possa commetta un illecito o compri un libro. Come è evidente, si tratta di un calcolo che avviene su base statistica e tuttavia il presupposto (errato) che l’algoritmo sia del tutto neutrale o oggettivo condiziona il risultato. Qualsiasi algoritmo è sempre contestuale, ovvero relativo a una cornice o *frame* all’interno del quale trova applicazione. Nessuna informazione è di per sé “neutrale”, se è vero che per informazione dobbiamo intendere il significato associato al dato – ma il significato,

ovvero l'interpretazione del dato, è già un atto ermeneutico, è già frutto di un soggetto che giudica.

Il medesimo principio relativo al dato si deve applicare agli algoritmi, poiché il loro schema deriva da un insieme di istruzioni che devono essere preliminarmente stabilite dal programmatore. Facciamo un esempio: "Accendi il boiler quando la temperatura è inferiore a 20 gradi". In questo caso, ci troviamo di fronte a un dato che è facilmente riducibile a quantità, la temperatura. Più complicato invece applicare una riduzione quantitativa agli esseri umani per ottenere dei dati. Per esempio, stabilirne il colore della pelle, la formazione culturale, l'origine, oppure quello che è considerato un "paese a rischio"¹.

Nell'era digitale, gli algoritmi computazionali e i sistemi decisionali automatizzati rappresentano un potentissimo strumento per migliorare l'efficienza delle politiche e delle organizzazioni pubbliche. Non occorre quindi adottare una visione distopica, invocando nuove declinazioni tecnocratiche del "capitalismo della sorveglianza"². Non vi è dubbio che dalla possibilità di un utilizzo democratico e solidale delle nuove tecnologie digitali dipende, per esempio, non solo la speranza di una medicina capace di diagnosticare e curare con largo anticipo alcune malattie, di operare a distanza o di prevenire il diffondersi di infezioni (si pensi alla Xylella fastidiosa) e pandemie (per es. il COVID-19), ma anche la possibilità di utilizzare con parsimonia e in maniera sostenibile le risorse del pianeta. La questione non è se le tecnologie digitali, nel particolare, siano o meno pericolose e possano condurre a un'era di dominio tecnologico delle macchine sull'uomo. La questione è come utilizzare queste tecnologie altamente evolute, performanti e *trasformative* in strumenti al servizio della società e del benessere degli individui. Qui risiede il principale problema. E poiché anche gli algoritmi possono essere determinati da schemi che incapsulano pregiudizi, allora occorre vigilare criticamente sul loro utilizzo³. Non vi è dubbio che Stati non democratici possano utilizzare le nuove tecnologie a vantaggio del dominio politico. Ma questa è una questione di struttura sociale oltreché di proprietà dei mezzi di produzione – economica, culturale e scientifico-tecnologica – per tradurlo in termini marxiani.

E tuttavia, un aspetto cruciale è la tendenza intrinseca degli algoritmi a codificare pregiudizi, compresi quelli legati alla "razza", termine che qui utilizzo come un *feticcio*

¹ Rimangono attuali le riflessioni dell'economista statunitense R. Nelson, il quale in *The moon and the ghetto* del 1977, in cui sosteneva che l'analisi quantitativa avanzata risolveva problemi altamente complessi, compreso lo sbarco sulla Luna, ma risultava inefficace nel risolvere i complessi problemi sociali del "ghetto", per usare il suo termine, ovvero sfide politiche e di gestione pubblica legate alla riduzione della povertà, al razzismo e alle disuguaglianze sociali. Per Nelson la causa era la debolezza o fluidità teorica dei problemi sociali e delle relazioni mezzi-fini poco chiare.

² S. Zuboff, *Il capitalismo della sorveglianza. Il futuro della umanità*, Luiss, 2019.

³ Come sostenuto dalla Noble, gli algoritmi possono trasformarsi in strumenti di oppressione e rafforzare le pratiche e gli assunti razzisti. Cfr. S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, New York 2018.

ideologico, e altri insiemi di caratteristiche e connotazioni correlati alla necessità di identificare soggetti umani⁴. Si tratta di fenomeno noto come razzismo sistemico, un pregiudizio istituzionalizzato (*bias*) che può influenzare in modo significativo le decisioni e le politiche pubbliche. Nel nostro caso, il razzismo algoritmico è una sua sottospecie, si oggettiva nei processi automatizzati classificatori a monte del processo di identificazione dei migranti⁵, che possono comprendere, per esempio, funzioni quali l'estrazione dei dati dai cellulari, il rilevamento di dichiarazioni mendaci (test della verità), il riconoscimento delle emozioni e il *matching* dei dati biometrici. Gli algoritmi sono determinati dai programmatori umani che quindi possono condizionare, coscientemente o inconsciamente, l'intervento di pregiudizi di

⁴ In teoria, gli algoritmi computazionali dovrebbero codificare le regole amministrative o procedurali nel software per produrre regole trasparenti, coerenti e applicate in modo neutrale. Nella migliore delle ipotesi, gli algoritmi dovrebbero replicare il *giudice imparziale* di humana memoria. Le tecnologie assolvono a numerose funzioni e sono utilizzate in maniera diversa a seconda dei paesi di accesso: previsione della mobilità futura; valutazione del rischio e profilazione; elaborazione di visti, autorizzazioni di viaggio e domande di cittadinanza; verifica dell'identità e rilevamento delle frodi, comprese le tecnologie utilizzate per la verifica dei documenti e il riconoscimento di comportamenti/emozioni; il riconoscimento vocale, estrazione di dati dai telefoni cellulari; categorizzazione delle domande in base ai rischi percepiti o al tipo/complessità; monitoraggio elettronico, distribuzione di prestazioni sociali; strumenti di *matching* utilizzati per la distribuzione, in generale, del luogo di residenza. Per esempio, Ungheria e Lettonia (in precedenza anche la Grecia) stanno testando strumenti per il monitoraggio della veridicità della dichiarazione del migrante (*lie detection*, test di verità), nell'ambito del progetto di ricerca iBorderCtrl (finanziato nell'ambito del programma *Horizon 2020* e del programma *Secure Societies*). Il progetto è stato testato tra il 2016 e il 2019 e ha combinato diverse tecnologie, tra cui la verifica biometrica, il rilevamento dell'inganno, l'autenticazione dei documenti e la valutazione del rischio, in un unico sistema. Un altro progetto di ricerca controverso, TRESSPASS, mira a implementare il progetto iBorderCtrl. Finanziato anch'esso da *Horizon 2020*, TRESSPASS si basa sui risultati del progetto H2020 iBorderCtrl (per il controllo delle frontiere terrestri) e dei progetti H2020 FLYSEC e FP7 XP-DITE (per il controllo delle frontiere aeroportuali), al fine di stabilire «una soluzione di sicurezza multimodello basata sui rischi per l'attraversamento delle frontiere». Nel progetto pilota, TRESSPASS ha esplorato la fattibilità e l'utilità dell'analisi del comportamento (comprese le emozioni) durante le interviste delle guardie di frontiera e dei funzionari doganali. Il team ha suggerito che «questo tipo di tecnologia potrebbe migliorare [...] la correttezza dei risultati di tali interviste». Il progetto ha sollevato polemiche relative, oltre alle questioni legate all'accuratezza tecnica, anche alla necessità di esaminare le implicazioni etiche di queste tecnologie. In Italia e in Turchia, si utilizza la tecnologia *speech-to-tech* per la trascrizione dei colloqui con i richiedenti asilo. L'Inghilterra è il paese che utilizza più tecnologia in assoluto: valutazione del rischio per il trattamento delle domande di visto per visitatori (interrotto); valutazione del rischio delle domande di matrimonio; categorizzazione delle domande per il programma di insediamento UE; identificazione e definizione delle priorità degli immigrati irregolari; monitoraggio elettronico; estrazione dei dati dai telefoni cellulari. Attendiamo ora di vedere come funzionerà il nuovo sistema di ingressi-uscite EES, che l'Unione Europea dovrà istituire entro la fine del 2024, tenendo conto delle norme disciplinate dall'*IA Act*. Il progetto si basa sui risultati precedenti della Francia sull'uso dei dati biometrici e sui sistemi IT di gestione delle frontiere su larga scala dell'UE.

⁵ Si veda J.E. Fountain, *The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms*, in «Government Information Quarterly», 2022, vol. 39, n. 2, pos. 101645.

valutazione alla base della realizzazione di un algoritmo oppure essere culturalmente condizionati.

Il razzismo algoritmico è un problema complesso, già noto, che si annida nei dati stessi utilizzati per addestrare gli algoritmi, così come nei modelli e nei risultati che producono⁶. Questo pregiudizio si manifesta in molteplici forme, dalle decisioni discriminatorie ai processi che possono svantaggiare le persone sulla base della “razza”, fino alle politiche che possono perpetuare la discriminazione razziale. Per comprendere appieno il ruolo degli algoritmi computazionali in questo contesto, è fondamentale analizzare attentamente come vengono progettati, sviluppati e utilizzati, in particolare in relazione alla Intelligenza Artificiale.

Le IA hanno capacità adattive e predittive, oltre a varie funzioni, come la visione artificiale, l'elaborazione del linguaggio naturale, la robotica. Le IA si basano sull'apprendimento automatico attraverso il *machine learning*, una tecnica che utilizza algoritmi che migliorano automaticamente con l'esperienza o imparano senza essere programmati esplicitamente. L'apprendimento automatico consente pertanto alle macchine di migliorare le correlazioni man mano che nel modello aumentano i dati, adattando i modelli nel tempo via via che i nuovi dati vengono correlati. E tuttavia, nemmeno le intelligenze artificiali sono neutrali. Quella che potremmo considerare la *forma mentis* di un'IA viene addestrata su *data-set* che costituiscono e determinano le sue “basi di conoscenza”. È un po' come avviene per l'apprendimento di un bambino, cui viene data una certa educazione: se verrà educato alla diversità, sarà inclusivo; se invece verrà educato alla paura e alla discriminazione, diventerà un adulto razzista. La base di conoscenza può inoltre essere personalizzata attraverso la tecnica del *fine-tune training*, cioè addestrando l'IA su specifiche basi di conoscenza, perimetrata da un certo dominio. La scelta del modello di base su cui performare l'IA per il *fine-tune*, con il fine di diminuire l'errore statistico, influenza sia il comportamento dell'algoritmo sia

⁶ Uno degli esempi più noti è la pratica del *redlining*: utilizzata soprattutto in ambito immobiliare e bancario, ha determinato e radicalizzato le disuguaglianze in base alla razza e al pregiudizio sociale. Nel 1935, la *Federal Home Loan Bank* chiese alla *Home Owners Loan Corporation* di indicare delle zone chiamate “*residential security maps*”, al fine di valutare il rischio legato all'acquisto di immobili in tutte le aree metropolitane statunitensi. Così il governo federale pubblicò mappe di 236 città, in cui ogni quartiere viene contrassegnato da un livello di rischio sugli investimenti immobiliari, utilizzando i colori verde, blu, giallo e rosso. Le zone in rosso erano quelle abitate da minoranze, afroamericana, ispanica o altre a seconda dello Stato. Negli anni Sessanta, John McKnight, sociologo e attivista della *Northwestern University*, conia il neologismo *redlining* in riferimento alla prassi di segnare in rosso i quartieri meno desiderabili. Il termine è stato poi impiegato, in senso esteso, anche alle persone, indipendentemente dal luogo in cui risiedono, ma in relazione al loro status sociale. Stando a questa classificazione, per cui, ad esempio, le persone di colore hanno maggiori probabilità di pagare tassi di interesse o premi più alti solo perché sono neri o latini, soprattutto se vivono in quartieri a basso reddito. Come ha dimostrato la Noble, nell'uso quotidiano di internet e nelle tecnologie di intelligenza artificiale che utilizziamo, la discriminazione è incorporata anche nel codice dei computer, in maniera invisibile e anonima.

i costi di gestione. Nella pratica, si riduce la base di conoscenza per avere risposte più precise, e quindi la si condiziona volontariamente.

Per riassumere, è possibile condizionare l'IA almeno in tre fasi: addestrandola su *data-set* che incorporano pregiudizi; performandola in maniera da circoscrivere la base di conoscenza; correggendo le risposte sulla base del principio di retroazione (*feedback*), in base al quale l'IA modifica e modella la propria base di conoscenza. Queste tre fasi sono "opache", non sono cioè aperte alla verifica di chi utilizza l'IA. Lasciando da parte la questione tecnica della *black box* relativa ai sistemi neurali complessi e multilivello che caratterizzano la struttura algoritmica dell'IA – un problema che rimane pur sempre di non poco momento – è spesso impossibile conoscere su quale *data-set* sono state addestrate le IA, in quale maniera sono state programmate e a chi e come è stato affidato l'operazione di correzione delle risposte per orientare la base di conoscenza.

L'IA è stata ampiamente utilizzata però nel processo decisionale in materia di migrazione. Come ricorda nel dossier del 2023 Derya Ozkul, ricercatrice senior del centro oxfordiano sui *refugee studies*:

L'uso delle nuove tecnologie sta gradualmente aumentando nei settori della migrazione e dell'asilo in tutta Europa. [...] L'uso delle nuove tecnologie, e in particolare dei sistemi decisionali automatizzati, può accelerare i processi decisionali a vantaggio delle agenzie governative e di alcuni richiedenti. Tuttavia, possono anche portare a nuove vulnerabilità. Se da un lato l'uso delle nuove tecnologie può potenzialmente facilitare alcuni processi decisionali, dall'altro i loro rischi intrinseci di parzialità, discriminazione e potenziali 'errori della macchina' rappresentano una minaccia significativa per i (potenziali) migranti e richiedenti asilo che sono già privi di diritti e si trovano ad affrontare difficoltà nel cercare rimedi. L'uso di nuove tecnologie può anche portare a nuove relazioni tra il settore pubblico e quello privato per sviluppare, sostenere e implementare queste tecnologie. [...] Poiché la proposta di legge [europea] sull'IA classifica come 'ad alto rischio' l'utilizzo dell'IA per l'immigrazione, l'asilo e il controllo delle frontiere, è necessario condurre un'indagine sistemica sulle pratiche attuali e sulla portata del loro utilizzo in Europa⁷.

⁷ D. Ozkul, *Automating immigration and asylum: the uses of new technologies in migration and asylum governance in Europe*, Refugee Studies Centre, University of Oxford, Oxford 2023, p. 5. Cfr. Id., p. 11: «In Europa, nel 2013, la Commissione europea ha proposto il pacchetto 'Frontiere intelligenti', seguendo i suggerimenti per l'istituzione di un sistema di ingresso/uscita e di un programma per viaggiatori registrati. Più di recente, la Commissione europea ha commissionato una relazione che esamina le opportunità e le sfide per l'uso delle tecnologie dell'intelligenza artificiale nel controllo delle frontiere, nella gestione della migrazione e della sicurezza in Europa, e promuove l'uso di nuove tecnologie, come le valutazioni automatizzate del rischio dei cittadini di paesi terzi (TCN)».

Le decisioni relative all'ammissibilità, all'accoglienza o all'espulsione di migranti e rifugiati possono letteralmente cambiare il destino e la vita delle persone coinvolte⁸. Di conseguenza, è essenziale verificare la neutralità degli algoritmi impiegati in questo contesto.

Qui non possiamo entrare nell'analisi dettagliata del *report*, ma essa mostra modelli algoritmici che determinano l'aumento della sorveglianza, la violazione della privacy, la partecipazione a pratiche di lavoro occulte. Gli studi più recenti mostrano come l'IA sia stata applicata al processo decisionale legato all'immigrazione, concentrandosi principalmente sulla questione dell'aderenza ai patti internazionali relativi alla protezione dei diritti umani, là dove esiste una frizione tra le legislazioni degli stati europei e occorre una tassonomia di base per chiarirne il funzionamento⁹.

Gli studi specialistici in ambito di *immigrazione e algoritmi* mettono in luce conseguenze inquietanti, che suggeriscono come le decisioni automatizzate in materia di immigrazione siano potenzialmente discriminatorie nei confronti dei migranti. Ancora più preoccupante è il fatto che questi pregiudizi sembrano essere stati intrinsecamente presenti sin dalla creazione degli algoritmi, a causa dei dati storici

⁸ Per esempio, come riportato nel report di Ozkul, nel Regno Unito è emerso che, tra il 2015 e il 2020, l'*Home Office* ha utilizzato un algoritmo per elaborare le domande di visto per i visitatori, che ha portato a risultati potenzialmente discriminatori. Questa pratica è stata successivamente interrotta grazie all'impegno costante delle organizzazioni della società civile - il *Joint Council for the Welfare of Immigrants e FoxGlove*. Nei Paesi Bassi, tra il 2014 e il 2021, il Servizio per l'immigrazione e la naturalizzazione (IND) ha utilizzato la valutazione del rischio per valutare l'affidabilità dei potenziali sponsor di migranti altamente qualificati. In questo processo, l'IND cercava e archiviava diverse caratteristiche delle aziende, tra cui la composizione etnica del consiglio di amministrazione. Questa pratica è stata interrotta grazie alle pressioni interne ed esterne. L'IND sta attualmente sviluppando un nuovo modello di rischio per valutare questi sponsor senza memorizzare informazioni sulla loro composizione etnica.

⁹ F. Palmiotto, *When is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis*, 10.07.2023. In corso di pubblicazione in «German Law Journal», disponibile su SSRN: <https://ssrn.com/abstract=4578761>, ultima consultazione: 14.02.2024. Il contributo affronta le urgenti questioni relative all'uso di sistemi automatizzati nel processo decisionale pubblico, con un'attenzione specifica al settore della migrazione, dell'asilo e della mobilità. Basandosi sulla ricerca empirica condotta per il progetto AFAR, il documento esamina il potenziale e i limiti del Regolamento generale sulla protezione dei dati e della proposta di legge sull'intelligenza artificiale nell'affrontare efficacemente le sfide poste dal processo decisionale automatizzato (ADM). Il documento sostiene che le attuali definizioni e categorizzazioni legali dell'ADM non riescono a cogliere la complessità e la diversità delle applicazioni reali, in cui i sistemi automatizzati assistono i decisori umani piuttosto che sostituirli completamente. Questa discrepanza tra il quadro giuridico e l'attuazione pratica evidenzia la necessità di un approccio ai diritti fondamentali per la protezione giuridica nell'era dell'automazione. Per colmare il divario tra l'ADM nella legge e nella pratica, il documento propone una tassonomia che fornisce chiarezza teorica e consente una comprensione completa dell'ADM nel processo decisionale pubblico. Questa tassonomia non solo migliora la nostra comprensione dell'ADM, ma identifica anche i diritti fondamentali in gioco per gli individui e la legislazione specifica del settore applicabile all'ADM. Il documento invita infine a raccogliere osservazioni empiriche e contributi da parte di esperti in altre aree del diritto pubblico per arricchire e perfezionare la tassonomia proposta, garantendo così quadri concettuali più chiari per tutelare gli individui nella nostra società sempre più algoritmica.

contaminati da pregiudizi e da problematiche strutturali nei sistemi stessi. L'*opacità tecnologica* dell'IA ha reso ancor più impellente la necessità di una *policy* di trasparenza e di 'spiegabilità' (*explainability*) del processo di scelta, oltreché sollevate questioni relative alla costruzione di quella che in questa sede abbiamo chiamato *forma mentis* della IA. La comprensione dell'impatto dell'utilizzo delle nuove tecnologie sui processi decisionali è direttamente correlata alla trasparenza di chi le utilizza e per le pubbliche amministrazioni corre vieppiù un obbligo legale e morale.

Tuttavia, non possiamo rinunciare a uno strumento così formidabile e potente¹⁰, che potrebbe rivelarsi esso stesso non solo un mezzo per analizzare i pregiudizi (si pensi a forme di IA antagoniste), ma anche per favorire l'inclusione. Inoltre, la trasparenza dei *data-set* sui quali sono addestrati le IA e che le alimentano modificandole, il processo di correzione e una vigilanza sul rispetto dei diritti fondamentali dell'essere umano sarebbero già una garanzia di un maggior equilibrio nell'utilizzo di una tale tecnologia.

Per concludere, ciò che qui si suggerisce è un passaggio dall'algoetica (l'etica fondata sugli algoritmi), per utilizzare un'efficace neologismo introdotto da Paolo Benanti, all'algoetica (dal greco *algo* = dolore) un'etica cioè che si fa carico del dolore altrui, fondata sulla compassione e sull'empatia, sollecitata ad interrogarsi sulle forme occulte di manipolazione e condizionamento algoritmico dei processi di discriminazione dei migranti, poiché è sulle loro spalle che si gioca la fondamentale partita della migrazione globale e non si può pensare di demandarla a sistemi automatici che non solo incapsulano strutturalmente pregiudizi, ma diventano strumenti invisibili di discriminazione razziale e di conservazione statistica del dato storico¹¹. Dobbiamo rendere quanto più trasparente questo processo, pena quello che Achille Mbembe ha recentemente chiamato *brutalismo*: un'epoca in cui tutte le sfere del vivere sono pervase dal capitale e tutta l'organizzazione umana è soggetta al calcolo digitale, da cui l'elemento umano, sovrabbondante e non categorizzabile, viene escluso a priori.

¹⁰ Gli strumenti di intelligenza artificiale possono essere impiegati con grande vantaggio per la collettività: per migliorare l'efficienza della città, per prevenire frodi e corruzione, per costruire reti sociali che rafforzano la fiducia pubblica, per razionalizzare e personalizzare l'erogazione dei servizi e modellare sistemi complessi e dinamici, come il flusso del traffico stradale o le emissioni inquinanti o la previsione di eventi climatici catastrofici.

¹¹ Per esempio: se storicamente le donne nere sono sottorappresentate a causa di discriminazioni passate, il modello può "imparare" (cioè dedurre sulla base di correlazioni) che le donne nere non sono idonee ad assumere determinate posizioni di funzione apicale nella società.

