

# Intelligenza artificiale costituzionalmente orientata come veicolo positivo di eguaglianza: algoritmi “riparativi” e programmazione sanitaria\*

Francesca Morganti\*\*

**SOMMARIO:** 1. Introduzione. Le applicazioni medico-sanitarie dell'intelligenza artificiale, tra rischi di discriminazione e potenzialità riparative. – 2. Modelli predittivi e programmazione sanitaria: il problema della “formulazione del problema”. – 3. Dalle logiche dell'anti-classificazione a quelle dell'anti-subordinazione, dalla mera *fairness* a veri e propri algoritmi “riparativi”. – 4. Algoritmi “riparativi” e azioni positive: considerazioni conclusive.

## ABSTRACT:

Nel contributo, a partire da alcuni esempi pratici, si ipotizza che le applicazioni medico-sanitarie dell'intelligenza artificiale, sempreché “costituzionalmente orientate”, possano contribuire a mitigare *bias* – algoritmici e non – preesistenti; si propone, poi, il perseguimento di finalità re-distributive mediante il ricorso a veri e propri algoritmi “riparativi”, i quali, addestrati sulle specificità e le storture del contesto d'utilizzo, saranno in grado di tenerne conto e agevolarne il superamento.

*This paper explores, through selected practical examples, the hypothesis that medically-oriented applications of artificial intelligence – provided they are guided by constitutional principles – may help mitigate both algorithmic and non-algorithmic pre-existing biases. It further advocates for the pursuit of redistributive objectives through the deployment of genuinely “reparative” algorithms. These systems, when trained on the specific characteristics and distortions of their operational con-*

\* Il contributo costituisce la rielaborazione dell'intervento svolto durante la quinta conferenza annuale di ICON-S Italia, dedicata a “Lo stato delle transizioni” (Università degli Studi di Trento, 18-19 ottobre 2024) e muove dalle ricerche svolte nell'ambito del PRIN 2020 “Il diritto costituzionale della salute e dell'organizzazione sanitaria dopo l'emergenza della pandemia” (p.i. prof. Renato Balduzzi).

\*\* Assegnista di ricerca in Diritto costituzionale e pubblico presso l'Università degli Studi di Roma “Tor Vergata”, francesca.morganti@outlook.com.

*text, would be capable of recognizing and addressing such disparities, thus contributing to their reduction.*

## 1. Introduzione. Le applicazioni medico-sanitarie dell'intelligenza artificiale, tra rischi di discriminazione e potenzialità riparative

È ormai ampiamente ipotizzato – e, almeno in parte, dimostrato<sup>1</sup> – che il ricorso a sistemi di intelligenza artificiale, specialmente in ambito medico-sanitario, comporti dei rischi per gli utenti (*rectius*, per le persone coinvolte) e possa incidere negativamente sui loro diritti, in particolare di libertà<sup>2</sup>, oltreché sulle “condizioni di eguaglianza” che andrebbero garantite loro nell’accesso ai servizi (sanitari, nel caso di specie).

Si potrebbe pensare, forse in maniera semplicistica, che l’implementazione su larga scala di sistemi di intelligenza artificiale sia destinata a risolvere definitivamente il problema del pregiudizio – e della discriminazione, per via indiretta – in ambito medico-sanitario e non solo: le macchine, a differenza delle persone, non avrebbero simpatie o antipatie, non di-

<sup>1</sup> V. *infra*, *passim*.

<sup>2</sup> Nell’ambito medico-sanitario, in particolare, è l’*opacità* dei sistemi di apprendimento automatico – la loro impostazione “a scatola nera” – a rappresentare un problema per il diritto alla salute *sub specie* di libertà, e in particolare per l’impianto del consenso informato (costituzionalmente fondato sul «profilo di libertà» del diritto alla salute medesimo: v. almeno D. MORANA, *La salute come diritto costituzionale. Lezioni*, IV ed., Giappichelli, Torino, 2021, spec. pp. 164 ss.). Con l’espressione “consenso informato” si richiamano, infatti, «diritti di diversa natura e struttura», e in particolare: (i) da un lato, il «diritto alla salute come diritto di libertà, o più precisamente la libertà di cura come *libertà di acconsentire alle cure* (o rifiutare le stesse)»; (ii) dall’altro, il «diritto alla salute come diritto a ricevere informazioni sulle cure, cioè il *diritto ad essere informato sulle cure* (ma più in generale su tutti i trattamenti, anche non curativi in senso stretto), dunque di ricevere le informazioni necessarie alla formazione del consenso (o dissenso) all’interno dell’alleanza terapeutica col medico» (così ancora D. MORANA, *A proposito del fondamento costituzionale per il «consenso informato» ai trattamenti sanitari: considerazioni a margine della sent. n. 438 del 2008 della Corte costituzionale*, in *Giur. cost.*, 6/2008, p. 4976). Quando il medico si avvalga di strumenti digitali *opachi*, che forniscono proposte di diagnosi o terapia delle quali non vi è modo di verificare la correttezza – perché l’*iter* di “pensiero” che ha condotto la macchina dal punto A (i dati di *input*) al punto B (la classificazione o predizione in *output*) non è conoscibile, e il professionista sanitario, anche non fosse d’accordo, può solo decidere se *accogliere* o *respingere* il parere del sistema – è materialmente impossibile, per il medico stesso, fornire informazioni complete e circostanziate al paziente. La macchina, dunque, inserisce un elemento di discontinuità all’interno dell’alleanza terapeutica, impedendo al medico di avere un quadro d’insieme e al paziente di ricevere gli elementi necessari per una scelta *informata*. Se, in altri termini, il diritto a ricevere informazioni sulle cure è strumentale al godimento, in forma “attiva” o “passiva”, del diritto alla salute *sub specie* di libertà – se, in particolare, esso si pone come presupposto logico-giuridico della *consensualità* del trattamento, a sua volta «corollario necessario» della *volontarietà* di quest’ultimo (ivi, p. 4973) – il ricorso all’intelligenza artificiale in ambito clinico, quando consista nell’impiego di sistemi (obiettivamente o intenzionalmente) opachi, come di regola sono quelli di *machine learning*, inevitabilmente incide sulle condizioni pratiche di godimento della *libertà di salute*. Sul tema v. anche C. DE MENECH, *Intelligenza artificiale e autodeterminazione in materia sanitaria*, in *BioLaw Journal – Rivista di BioDiritto*, 1/2022, pp. 181 ss., la quale sottolinea, tra l’altro, come «l’opacità di alcuni sistemi di intelligenza artificiale «risult[is] evidentemente malacconcia rispetto all’obiettivo di prospettare al paziente uno scenario diagnostico-terapeutico chiaro ed affidabile, con riguardo al quale egli possa riuscire ad autodeterminarsi razionalmente» (pp. 185 s.).

stinguerebbero irragionevolmente sulla base del genere o della «razza»<sup>3</sup>, non opererebbero (più o meno consapevoli) associazioni stereotipiche.

Va evidenziata, tuttavia, in prima battuta, la natura *oggettiva* e *strutturale* di molte forme di discriminazione, che non dipendono (esclusivamente) dai comportamenti individuali, ma dall'organizzazione della società e dei suoi servizi. Una condotta, detto altrimenti, può rivelarsi discriminatoria, *a prescindere* dall'intenzione e dai motivi di chi l'ha tenuta, ogniquale volta tenda a rafforzare dinamiche di potere preesistenti e inique, fondate sulla marginalizzazione di uno o più gruppi sociali – definiti in base a tratti identitari perlopiù “costruiti”<sup>4</sup> – in favore di altri<sup>5</sup>. L'apparente *neutralizzazione* dei processi decisionali che si realizzerebbe esternalizzandoli (*i.e.*, affidandoli a sistemi presentati come imparziali-per-definizione) potrebbe ridurre l'impatto delle simpatie e antipatie personali, delle discrimi-

<sup>3</sup> Nella piena consapevolezza del “peso” della nozione storico-biologica di razza, ampiamente confutata, si fa nondimeno ricorso all'espressione – comunque inserita, con scelta deliberata e implicitamente “di rottura”, nell'art. 3, comma 1, della nostra Costituzione – per richiamare brevemente i significati e le categorizzazioni (di matrice storica, etnica, socio-economica) sintetizzati nell'inglese “*race*”.

<sup>4</sup> Com'è noto, ad esempio, tanto la «razza» (v. nt. prec.) che il genere hanno natura “costruita” e matrice socio-culturale (in opposizione a storico-biologica) – il che, ovviamente, non li rende meno tangibili, meno rilevanti nell'esperienza quotidiana e ordinamentale.

<sup>5</sup> Secondo A. PACE, *Eguaglianza e libertà*, in *Pol. dir.*, 2/2001, p. 171, parafrasando Norberto Bobbio, «è l'odio: è il disprezzo per il diverso, per l'altro sesso, per un dato colore della pelle, per chi non condivide le proprie idee politiche o la propria fede religiosa, ciò che caratterizza la discriminazione» e la distingue dalla mera differenziazione irragionevole o arbitraria; al legame che sussisterebbe tra «illegittima personale discriminazione» e «odio» – del quale, stando a questa ricostruzione, sarebbe spia proprio la non-giustificatazza della differenziazione – fa riferimento, in maniera meno diretta, L. PALADIN, voce *Eguaglianza (dir. cost.)*, in *Enc. dir.*, XIV, Giuffrè, Milano, 1965, p. 525, secondo il quale «[l]a disegualianza (...) si traduce in illegittima personale discriminazione, allo stesso modo e per gli stessi motivi che storicamente producono il trapasso delle prerogative in privilegi: venendo meno o mancando in partenza quell'“utilità sociale”, cui debbono tenere le classificazioni normative, esse risultano ispirate per ciò solo (a prescindere dalle non rilevanti intenzioni del legislatore) dall'odio o dal favore nei confronti degli interessati» (enfasi aggiunta): l'odio, in questo caso, pare quasi reificarsi, o comunque “sperzonalizzarsi”. Questo *elemento soggettivo*, tuttavia, per quanto rilevante, non basta a connotare la discriminazione come tale: se così fosse, davvero si potrebbe parlare – e non solo per compiere una forzatura dialettica – di “discriminazione alla rovescia”, quando sia il componente di un gruppo-prevaricatore a essere deliberatamente prevaricato (o, comunque, sfavorito); in realtà, le pratiche discriminatorie sono strettamente intersecate ai meccanismi e alle dinamiche del *potere* – politico, socio-economico, culturale perfino – e non esistono al di fuori delle relative gerarchie. Cfr., in tema, B. PEZZINI, *Costruzione del genere e Costituzione*, in EAD. (a cura di), *La costruzione del genere. Norme e regole*, I, *Studi*, Bergamo University Press, Bergamo, 2012, p. 16 e p. 54, la quale assimila le condizioni di genere a un vero e proprio «assetto di potere», costruendo l'eguaglianza (di genere) come sovversione di quell'assetto, e più in generale dello *status quo*: al mero «principio anti-discriminatorio», il quale mira a un «processo ascensionale di inclusione, che apre al soggetto discriminato l'accesso ad uno spazio prima o altrimenti precluso», l'A. affianca un vero e proprio «principio anti-subordinazione», teso a generare «il rovesciamento ed il superamento [dell']assetto di potere che produceva l'esclusione» e ad avviare «un processo circolare di ri-definizione di tutti i soggetti coinvolti e dello stesso spazio in cui agiscono e su cui insistono». Più in generale, come si vedrà anche *infra*, § 4, l'idea sottesa a simili riflessioni è che la discriminazione non sia “neutra”, che non rappresenti una mera differenziazione, ancorché irragionevole: essa consisterebbe, piuttosto, nel trattamento peggiore di soggetti (o categorie di soggetti) *storicamente* e *sistematicamente* subordinati rispetto alle loro “controparti” (le donne rispetto agli uomini, i neri rispetto ai bianchi *etc.*); la discriminazione, in altri termini, è per definizione *in linea con* le dinamiche di subordinazione che caratterizzano il contesto sociale in esame, ed è per ciò che non sembra avere senso, sul piano logico prima che giuridico, parlare di “discriminazione alla rovescia”: perché il trattamento favorevole di individui o gruppi generalmente “sfavoriti” va *contro* la grana del tessuto sociale, va – in maniera occasionale e a-sistematica – *contro* quelle relazioni e quelle gerarchie che le discriminazioni tendono invece a rafforzare.

nazioni dirette e consapevoli, di ogni arbitrio squisitamente umano, ma non certo degli squilibri “ingranati” nell’organizzazione sociale.

I sistemi di intelligenza artificiale, in seconda battuta, sono prodotti *dall’uomo*, e *dell’uomo* incorporano, tra l’altro, le prevenzioni. Se ne distinguono, nella letteratura sul tema, di diversi tipi: (i) vi sono i *bias* preesistenti, propri dell’organizzazione che “disegna” il prodotto o sviluppa il programma, inevitabilmente ritenuti da quest’ultimo (in altri termini: se l’organizzazione è *biased*, il sistema sarà *biased*); (ii) vi sono, poi, i *bias* tecnici, frutto di scelte o vincoli altrettanto tecnici (attinenti, ad esempio, alla “scrittura” dell’algoritmo); (iii) vi sono, ancora, i *bias* emergenti, i quali, appunto, *emergono* man mano che il sistema di apprendimento automatico opera e si evolve; (iv) vi sono, da ultimo, i *bias* di misurazione, legati a criticità nella raccolta dei dati di addestramento<sup>6</sup>. Le distorsioni che attengono alla fase di progettazione e programmazione del sistema o a limiti dei *set* di dati possono considerarsi “derivate”, laddove quelle operate direttamente dalla macchina – quando, ad esempio, «il sistema di *machine learning* individui alcune caratteristiche che rinviano indirettamente a categorie protette (...) associando ai loro detentori un trattamento deteriore (...) in modo sostanzialmente autonomo» – sono, appunto, “autonome”, perpetrate dalla macchina *indipendentemente* da pregiudizi o mancanze o errori dell’uomo<sup>7</sup>.

Le problematiche relative ai dati e al *design* del sistema, per quanto potenzialmente impattanti, sono anche le più “controllabili” e contrastabili. Allo stato, ad esempio, il Regolamento europeo sull’intelligenza artificiale<sup>8</sup> richiede, per i sistemi “ad alto rischio”<sup>9</sup>, che i «*set* di dati di addestramento, convalida e prova» siano «pertinenti, sufficientemente rappresentativi e, nella misura del possibile, esenti da errori e completi nell’ottica della finalità

<sup>6</sup> Si tratta, ovviamente, di *una delle* classificazioni possibili, accolta – tra gli altri – da A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell’intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Riv. fil. dir.*, 1/2019, p. 96.

<sup>7</sup> Cfr. P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, in AA.VV., *Liber amicorum per Pasquale Costanzo – Diritto costituzionale in trasformazione*, I, *Costituzionalismo, reti e intelligenza artificiale*, Consulta Online, Genova, 2020, p. 461.

<sup>8</sup> Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, «che stabilisce regole armonizzate sull’intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828», d’ora in poi: *Regolamento sull’intelligenza artificiale* o *AI Act*.

<sup>9</sup> Stando all’art. 6, par. 1, dell’*AI Act*, «un sistema di IA è considerato ad alto rischio se sono soddisfatte entrambe le condizioni seguenti: a) il sistema di IA è destinato a essere utilizzato come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell’Unione elencata nell’allegato I; b) il prodotto, il cui i componente di sicurezza a norma della lettera a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto, è soggetto a una valutazione della conformità da parte di terzi ai fini dell’immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell’Unione elencata nell’allegato I». L’Allegato I menziona, tra gli altri, il Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, «relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio», del 5 aprile 2017, e il Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, «relativo ai dispositivi medico-diagnostici *in vitro* e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione», anch’esso del 5 aprile 2017: sul punto più diffusamente *infra*, nt. 11. Sono considerati “ad alto rischio”, inoltre, come precisa il par. 2 dello stesso articolo 6, «anche i sistemi di IA di cui all’allegato III»; fra questi, a mero titolo esemplificativo, specifiche categorie di sistemi riconducibili a settori quali la «Biometria», le «Infrastrutture critiche», l’«Istruzione e formazione professionale», l’«Accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali», l’«Attività di contrasto», l’«Amministrazione della giustizia» *etc.*

prevista», e che possiedano «le proprietà statistiche appropriate anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone relativamente ai quali il sistema (...) è destinato a essere usato»<sup>10</sup>. Poiché i sistemi di intelligenza artificiale impiegati in ambito medico-diagnostico e terapeutico sono per ciò stesso da considerarsi “ad alto rischio”<sup>11</sup>, una previsione di questa sorta vieta, tra l’altro, l’impiego di *software* diagnostici addestrati su *set* di dati *incompleti*, ovvero *non adeguatamente rappresentativi* della varietà – quanto al genere, alla “razza”, alle condizioni personali – della popolazione interessata.

Si pensi ai sistemi di *deep learning* impiegati come ausilio nella diagnostica (perché “abili”, tra l’altro, nel riconoscimento di immagini); si pensi, in particolare, a titolo di esempio, alla rete neurale convoluzionale che aveva “imparato” a distinguere e classificare diverse lesioni dermatologiche pigmentate<sup>12</sup>, e che avrebbe potuto essere ed era, dunque, d’aiuto ai medici dermatologi nella diagnosi del melanoma. Il sistema si era formato su una base di dati assai ricca, comprensiva di 129.450 immagini cliniche<sup>13</sup>; è stato rilevato, tuttavia, all’esito dell’addestramento e in sede di *deployment*, come la base in questione, nonostante l’innegabile “profondità”, mancasse di “ampiezza”, comprendendo in massima parte figurazioni riferibili

<sup>10</sup> Così l’art. 10, par. 3, del Regolamento sull’intelligenza artificiale. Si v. anche il Considerando (67), secondo il quale «[d]ati di alta qualità e l’accesso a dati di alta qualità svolgono un ruolo essenziale nel fornire una struttura e garantire le prestazioni di molti sistemi di IA, in particolare quando si utilizzano tecniche che prevedono l’addestramento di modelli, al fine di garantire che il sistema di IA ad alto rischio funzioni come previsto e in maniera sicura e che non diventi una fonte di discriminazione vietata dal diritto dell’Unione» (enfasi aggiunta).

<sup>11</sup> Come visto, infatti, l’art. 6, par. 1, lett. b), *AI Act*, qualifica come “ad alto rischio”, tra l’altro, «il prodotto, il cui componente di sicurezza a norma della lettera a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto» che sia «soggetto a una valutazione della conformità da parte di terzi ai fini dell’immissione sul mercato o della messa in servizio (...) ai sensi della normativa di armonizzazione dell’Unione elencata» nell’Allegato I. Ora, quest’ultimo Allegato menziona, fra gli altri, il Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, «relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio», del 5 aprile 2017, e il Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, «relativo ai dispositivi medico-diagnostici *in vitro* e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione», anch’esso del 5 aprile 2017. Il Regolamento (UE) 2017/745, in particolare, distribuisce i dispositivi medici (tra i quali, *ex art.* 2, pt. 1, sono inclusi anche i *software*) in classi di rischio, e stabilisce che i soli dispositivi della classe I non siano soggetti a valutazione di conformità da parte degli organismi notificati a ciò preposti; nella classe I andrebbero inclusi, a rigore, tutti i dispositivi «non invasivi» (cfr. Regola I, Allegato VIII), ma la Regola 11 dello stesso Allegato VIII attribuisce alle classi IIa o superiori qualunque «*software* destinato a fornire informazioni utilizzate per prendere decisioni a fini diagnostici o terapeutici»: i sistemi di intelligenza artificiale utilizzati in ambito medico-sanitario, dunque, *purché siano impiegati in contesto clinico* (e non, ad esempio, a fini di programmazione sanitaria), saranno *per ciò stesso* considerati “ad alto rischio”, vista la loro riconducibilità alle classi IIa o superiori del Regolamento sui dispositivi medici e la loro conseguente necessaria sottoposizione alle valutazioni di conformità in esso disciplinate (che, come detto, è condizione perché il sistema-prodotto si consideri “ad alto rischio” in base all’art. 6 dell’*AI Act*). I sistemi impiegati in ambito medico-sanitario ma *al di fuori del contesto clinico* (ad esempio a fini di programmazione sanitaria) potranno, ad ogni modo, essere verosimilmente ricondotti all’interno del settore n. 5 di cui all’Allegato III (v. *supra*, nt. 9), ossia l’«Accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi», in quanto, in particolare, «sistemi (...) destinati a essere utilizzati dalle autorità pubbliche o per conto di autorità pubbliche per valutare l’ammissibilità delle persone fisiche alle prestazioni e ai servizi di assistenza pubblica essenziali, compresi i servizi di assistenza sanitaria, nonché per concedere, ridurre, revocare o recuperare tali prestazioni e servizi».

<sup>12</sup> A. ESTEVA *et al.*, *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, in *Nature*, vol. 542, 2017, pp. 115 ss.

<sup>13</sup> *Ibidem*.



a persone «*fair-skinned*»<sup>14</sup>, e come la macchina, di conseguenza, tendesse a riconoscere e classificare con maggiore accuratezza – inevitabilmente – proprio le lesioni su pelle chiara. Più in generale, un sistema che viene addestrato su *data set* non sufficientemente rappresentativi – è questo il problema, in ottica giuridico-costituzionale – offrirà prestazioni di livello diverso con riferimento a sotto-popolazioni differenti, e un servizio deteriore, in particolare, ai componenti di gruppi minoritari e/o sotto-rappresentati nei dati di *training*. Quando questi gruppi, poi – come accade nella più parte dei casi – siano individuabili e circoscrivibili sulla base di alcune delle caratteristiche “protette” di cui all’art. 3 Cost., e soprattutto del «sesso» o della «razza», le criticità che si presentano e si presenteranno sono molteplici.

Il principio di eguaglianza è, prima di tutto, un principio *per il legislatore*, che impone a quest’ultimo di garantire – *in condizioni di eguaglianza*, appunto – il godimento di determinati diritti o, come si legge talvolta, «di un dato bene della vita»<sup>15</sup>. Gli articoli 3 e 32 della Costituzione, letti assieme, prescrivono – per utilizzare il lessico della legge 23 dicembre 1978, n. 833, di «Istituzione del servizio sanitario nazionale», inserita idealmente in quel solco – che si assicuri «l’eguaglianza dei cittadini nei confronti del servizio» e che quest’ultimo, una volta costituito, sia orientato «alla promozione, al mantenimento ed al recupero della salute fisica e psichica di tutta la popolazione *senza distinzione di condizioni individuali e sociali*»<sup>16</sup>.

Con riferimento al diritto alla salute, allora, il principio di eguaglianza rileva due volte: quando impone che i servizi sanitari siano erogati ai cittadini in condizioni di eguaglianza, senza distinguere *irragionevolmente* sulla base del «sesso», della «razza», della «lingua», della «religione», delle «opinioni politiche», delle «condizioni personali e sociali» della persona-paziente; e quando richiede che apposite misure vengano adottate – in nome dell’eguaglianza (anche) sostanziale<sup>17</sup> – per «consentire l’accesso alle cure a coloro che non siano in grado di sopportarne il costo»<sup>18</sup>.

<sup>14</sup> Cfr. A.S. ADAMSON, A. SMITH, *Machine Learning and Health Care Disparities in Dermatology*, in *JAMA Dermatology*, 11/2018, *passim*.

<sup>15</sup> «[I]n tanto un preteso diritto all’eguaglianza potrebbe essere invocato giudizialmente», rileva, ad esempio, A. PACE, *Eguaglianza e libertà*, cit., pp. 156 s., «in quanto venga in considerazione una discriminazione *rispetto al godimento di un dato bene della vita*» (enfasi aggiunta); in termini analoghi già C. MORTATI, *Istituzioni di diritto pubblico*, t. II, IX ed., CEDAM, Padova, 1976, p. 1023: «[c]i si chiede (...) se un atto che contraddica al principio di eguaglianza sia suscettibile di impugnativa da parte di chi appartenga alla categoria colpita dalla discriminazione, anche all’infuori o indipendentemente dalla personale ed effettiva *privazione del bene della vita* che senza quell’atto avrebbe potuto conseguire, o se invece solo una privazione di tal genere renda proponibile la doglianza. Più esatto è ritenere che il principio di eguaglianza non sia direttamente regolativo di situazioni giuridiche soggettive, ma piuttosto criterio obiettivo che opera sulla disciplina dei rapporti» (l’enfasi è aggiunta). Sulla natura dell’eguaglianza, più che come diritto dell’individuo, come «principio oggettivo» e come «*matrice* di una pluralità di diritti, che da essa traggono alimento», v. anche M. LUCIANI, da ultimo ne *I principi di eguaglianza e di non discriminazione, una prospettiva di diritto comparato*, Servizio ricerca Parlamento europeo, Bruxelles, 2020, spec. p. 4.

<sup>16</sup> Così, appunto, l’art. 1, co. 3, della l. 23 dicembre 1978, n. 833, di «Istituzione del servizio sanitario nazionale»; l’enfasi è aggiunta.

<sup>17</sup> Sul legame intercorrente fra misure latamente promozionali ed eguaglianza (anche) *formale* si tornerà *infra*, § 4.

<sup>18</sup> Sul punto – e sul carattere *necessariamente* relativo della nozione di «indigenza» incorporata nell’art. 32., co. 1, Cost., in base al quale la Repubblica «garantisce cure gratuite agli indigenti» – v., *ex multis*, D. MORANA, *La salute come diritto*

Ciascuno, dunque, ha diritto a godere del bene-salute senza incorrere in discriminazioni di sorta (o, se si vuole, è titolare di un «diritto al pari trattamento»<sup>19</sup> con riguardo al bene-salute); l'erogazione di servizi diseguali – migliori per gli uomini rispetto alle donne, per i bianchi rispetto ai neri – *che si basi o meno su strumenti digitali*, violerebbe, con tutta evidenza, una situazione soggettiva così sostanziata. In ambito medico, ovviamente, disparità e discriminazioni non nascono con la digitalizzazione<sup>20</sup>, ma rischiano di essere acuite dal ricorso massiccio alle macchine e ai trattamenti (del tutto o in parte) automatizzati: è compito *anche* del diritto fare in modo che ciò non accada.

---

costituzionale, cit., pp. 93 ss.; nonché, più risalenti, i contributi di L. CARLASSARE, *L'art. 32 della Costituzione e il suo significato*, in R. ALESSI (a cura di), *L'ordinamento sanitario*, I, *L'amministrazione sanitaria*, parte di AA.Vv., *Atti del Congresso celebrativo del centenario delle leggi amministrative di unificazione*, Neri Pozza, Vicenza, 1967, pp. 103 ss.; G. CORSO, *I diritti sociali nella Costituzione italiana*, in *Riv. trim. dir. pubbl.*, 3/1981, pp. 759 ss.; B. PEZZINI, *Il diritto alla salute: profili costituzionali*, in *Dir. soc.*, 1/1983, spec. pp. 26 ss., la quale sottolinea, in particolare, come «[l]a garanzia di cure gratuite agli indigenti non va[da] intesa in senso limitativo, ma come rafforzamento dell'impegno generale per la tutela della salute, nel quale è implicitamente ricompresa; non va intesa cioè come attività di beneficenza, distinta e separata dall'attività di assistenza sanitaria da prestarsi in favore della generalità: ogni concezione di "carità" rimane del tutto estranea ai principi dell'ordinamento costituzionale, fondato sulla *solidarietà sociale* e non modellato nelle forme dello stato assistenziale» (ivi, p. 27). D'altro canto, a chi in sede di Costituente proponeva di espungere ogni riferimento all'indigenza – «Noi respingiamo il concetto di indigenza consacrato nella Costituzione italiana. Oggi che tutti tendono alla solidarietà sociale, non possiamo pensare ad una categoria di cittadini che siano considerati indigenti e che, quindi, continuino ad avere bisogno della carità: non c'è più carità; c'è la solidarietà sociale», sosteneva l'on. Merighi nella seduta del 24 aprile 1947 – l'on. Arata, nel corso della stessa seduta, rispondeva così: «Mi duole di non essere d'accordo con il collega Merighi su questo punto. Noi accettiamo la sua opposizione a questa frase come un augurio che nella Repubblica italiana non vi saranno prossimamente più indigenti, ma siccome per il momento vi sono degli indigenti, noi riteniamo di votare il testo della Commissione per questo dovere dello Stato di dare loro cure gratuite».

<sup>19</sup> Così ancora A. PACE, *Eguaglianza e libertà*, cit., p. 157.

<sup>20</sup> Le disparità che pre-datano la digitalizzazione, tra l'altro, spesso complicano o «aggravano» quelle direttamente cagionate da quest'ultima: la «scarsità» di dati riferibili ai componenti di alcuni gruppi marginalizzati – che è alla radice di molte distorsioni algoritmiche – è dovuta, tra l'altro, alla loro sotto-rappresentazione nelle sperimentazioni, la quale a sua volta ha cause storiche, culturali, socio-economiche. Le donne, *e.g.*, tradizionalmente sono sotto-rappresentate nei *clinical trials*: «[n]ell'ambito della sperimentazione clinica», denunciava il Comitato Nazionale per la Bioetica già nel 2008, esse si pongono come «soggetti deboli», o comunque non oggetto di adeguata considerazione in ordine alla loro specificità sia in senso quantitativo (numero di donne arruolate rispetto al numero di uomini) sia in senso qualitativo (analisi dei dati rispetto alla differenza sessuale)» (COMITATO NAZIONALE PER LA BIOETICA, *La sperimentazione farmacologica sulle donne*, 28 novembre 2008, p. 8); questa sotto-rappresentazione ha principalmente ragioni «economiche» (l'arruolamento di donne e la stratificazione dei dati in base al sesso aumenta i tempi della ricerca e i relativi costi per le case farmaceutiche) e «biologiche» (le donne sono soggette a notevoli «fluttuazioni» ormonali nelle diverse fasi della vita), oltreché «sociali», «ambientali» *etc.* (ivi, pp. 11 ss.). Sul punto si tornerà *infra*, al § 3. Lo scetticismo di alcuni gruppi marginalizzati nei confronti della medicina «istituzionalizzata», inoltre, che ha spesso motivazioni storiche solide – si pensi a quanto emerso circa il *Tuskegee Study*, sul quale più diffusamente *infra*, spec. nt. 34, sovente annoverato tra le principali cause della sfiducia delle persone *African-American* nei confronti del *medical establishment* – inevitabilmente ne ha scoraggiato e ne scoraggia la partecipazione a studi e sperimentazioni. Sulla stretta correlazione tra «scarsa partecipazione delle minoranze etno-razziali ai *trial* clinici» e «impossibilità da parte di tali soggetti – già di per sé in una condizione di subalternità – di beneficiare delle innovazioni tecnologiche sviluppate anche con i dati provenienti da tali attività di ricerca», v., sempre con riferimento al contesto statunitense, V. LANDO, *Studi clinici, discriminazioni razziali e intelligenza artificiale: diversity and inclusion nel contesto statunitense*, in *BioLaw Journal – Rivista di BioDiritto*, spec./2024, pp. 155 ss. Ebbene, è la (relativa) «scarsità» di dati riferibili ai componenti dei gruppi minoritari o marginalizzati – dati che andrebbero raccolti, oltre che nella pratica clinica quotidiana, proprio nel corso di studi e sperimentazioni – che determina il prodursi, come anticipato, di diversi *bias* algoritmici.

È per questo che il Regolamento sull'intelligenza artificiale si occupa di qualità e completezza dei *set* di dati, esigendo, nell'ambito della loro *governance*, una valutazione preliminare ed *ex post* delle «distorsioni suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione»<sup>21</sup>, e imponendo che le basi di dati non solo siano rappresentative “in astratto”, ma tengano anche conto «delle caratteristiche o degli elementi particolari dello specifico ambito geografico, contestuale, comportamentale o funzionale all'interno del quale il sistema (...) è destinato a essere usato»<sup>22</sup>. Ed è per questo, guardando all'ordinamento interno, che la legge di delega «in materia di intelligenza artificiale» recentemente entrata in vigore<sup>23</sup> prescrive – nel suo art. 7, dedicato all'«Uso dell'intelligenza artificiale in ambito sanitario e di disabilità» – che i «sistemi di intelligenza artificiale utilizzati in ambito sanitario e i relativi dati impiegati» siano «affidabili, periodicamente verificati e aggiornati», sia al fine di «minimizzare il rischio di errori e migliorare la sicurezza dei pazienti» (comma 6), sia perché la (pur parziale) digitalizzazione dei servizi sanitari e socio-sanitari non può svolgersi secondo «criteri discriminatori» (comma 2).

Governare i dati, però, non è sufficiente: come si vedrà<sup>24</sup>, la stessa formulazione del problema sottoposto alla macchina può incidere sulle prestazioni di quest'ultima, conducendola a discriminare *indirettamente* nei confronti dei componenti di gruppi minoritari<sup>25</sup>.

Ciò che si tenterà di fare nel presente lavoro, tuttavia, è – dopo aver esplorato alcuni dei rischi connessi alle applicazioni medico-sanitarie dell'intelligenza artificiale, sia con riferimento al contesto clinico che alle attività di programmazione sanitaria – indagare le potenzialità *riparative* di quegli stessi sistemi; ci si chiederà, in altri termini, se gli algoritmi siano “destinati” a riprodurre e ad aggravare i *pattern* discriminatori ovvero possano, purché *costituzionalmente orientati*, trasformarsi in veicoli positivi di eguaglianza.

<sup>21</sup> Così l'art. 10, par. 2, lett. f), *AI Act*.

<sup>22</sup> Cfr. l'art. 10, par. 4, *AI Act*, che mira a evitare i cc.dd. *bias* contestuali, legati all'estrapolazione del sistema dal contesto in cui è stato disegnato e addestrato – magari più che adeguatamente – e al suo impiego in “ambienti” che presentano caratteristiche anche molto diverse (e.g., un programma sviluppato in «*high-resource settings: academic medical centers or state-of-the-art hospitals*» e utilizzato, a valle, in «*low-resource settings like community hospitals, community health centers, practitioners' offices*» rischia di offrire, nell'ambito d'uso, prestazioni notevolmente deteriori e, soprattutto, “inadeguate” alle circostanze; sul punto v., *ex multis*, W. NICHOLSON PRICE II, *Medical AI and Contextual Bias*, in *Harvard Journal of Law & Technology*, 1/2019, pp. 66 ss.).

<sup>23</sup> Il riferimento è alla l. 23 settembre 2025, n. 132, contenente «Disposizioni e deleghe al Governo in materia di intelligenza artificiale».

<sup>24</sup> *Infra*, spec. § 2.

<sup>25</sup> Dove “minoritario” è sinonimo di “subalterno”, e la marginalità è definita *qualitativamente* più che *quantitativamente*.



## 2. Modelli predittivi e programmazione sanitaria: il problema della “formulazione del problema”

Il problema della “formulazione del problema” è di particolare importanza nel *design* dei modelli predittivi impiegati a fini di stratificazione, ossia per distribuire una popolazione di riferimento all'interno di classi di rischio/bisogno (socio-sanitario, ad esempio). Sistemi di questa sorta, detto semplicemente, *sulla base di dati noti* (sesso/genere, età, “storico” clinico) effettuano una *stima* del rischio sanitario del paziente, e dunque una *predizione* del suo futuro bisogno di salute (*rectius*, di servizi sanitari), procedendo poi a collocarlo all'interno di una fascia di rischio.

Un modello simile, soprattutto se implementato a livello “istituzionale” (*i.e.*, di servizio sanitario, anziché lasciato alle singole strutture e reti), può rivelarsi di grande utilità, consentendo di identificare soggetti a forte rischio di eventi acuti o, più in generale, di un deterioramento delle condizioni di salute, e di coinvolgerli in appositi programmi di prevenzione: se il servizio si fa carico *tempestivamente* – prima di peggioramenti o eventi catastrofici – della persona-paziente, a beneficiarne saranno tanto l'individuo che il servizio stesso, le cui (limitate) risorse umane e finanziarie verranno impiegate più razionalmente. Il ricorso ad algoritmi di stratificazione, in altri termini, può creare le condizioni, almeno in teoria, per un godimento più pieno del diritto alla salute come diritto a prestazioni<sup>26</sup> – dal lato della persona – e consentire un “bilanciamento” ottimale tra quest'ultimo diritto e (*rilevanti*, ancorché non *fondamentali*) esigenze di contenimento della spesa, sanitaria e non<sup>27</sup> – dal lato del sistema.

Come è stato osservato, «gli obiettivi perseguiti nello sviluppare tali strumenti» rispondono, a un tempo, «ad (almeno) tre principî costituzionali: il principio di uguaglianza sostanziale,

<sup>26</sup> Sul “duplice volto” del diritto alla salute – come diritto di libertà e come «pretesa dell'individuo ad ottenere le prestazioni sanitarie che, di volta in volta, si rendano necessarie» – v. per tutti, D. MORANA, *La salute come diritto costituzionale*, cit., già pp. 1 ss.

<sup>27</sup> Sulla “problematicità” di una caratterizzazione «tralatizia» del diritto alla salute come «*finanziariamente condizionato*» – operata, tra gli altri, dalla Corte costituzionale – v., in senso critico rispetto ad alcune posizioni di quest'ultima, B. PEZZINI, *Il diritto alla salute a quarant'anni dall'istituzione del servizio sanitario nazionale: le criticità strutturali di un diritto sociale*, in *BioLaw Journal – Rivista di BioDiritto*, 2/2019, pp. 126 s., dove si sottolinea come, «[a]pplicata al diritto costituzionale alla salute, la formula tralatizia del *diritto finanziariamente condizionato* [appaia] semplicistica e fuorviante, rivelando tutta la sua ambiguità al confronto con la natura *fondamentale* del diritto[.] esso è certamente *condizionato all'attuazione del legislatore*, il quale, a sua volta, nel suo agire, non può prescindere dai vincoli finanziari nell'esercizio della sua discrezionalità, ma per il quale l'*equilibrio finanziario* – per quanta rilevanza abbia acquisito anche in riferimento alla formula costituzionale dell'art. 81 Cost. – *non costituisce uno scopo autosufficiente e irrelato*», bensì, tutt'al più, «una sorta di *finalità di terzo grado*, da giustificare nella sua coerenza alla *finalità secondaria* dell'efficienza, che, a sua volta, è strumentale alla realizzazione della prestazione dovuta per la *tutela della salute* (la *finalità diretta e primaria*)»; della stessa A. v. anche, sul medesimo tema ma più risalenti, B. PEZZINI, *Diritto alla salute e dimensioni della discrezionalità nella giurisprudenza costituzionale*, in R. BALDUZZI (a cura di), *Cittadinanza, Corti e salute*, CEDAM, Padova, 2007, p. 217, ed EAD., *Diritto alla salute e diritto all'assistenza tra tensione all'uniformità e logiche della differenziazione*, in R. BALDUZZI, G. DI GASPARÈ (a cura di), *Sanità e assistenza dopo la riforma del Titolo V*, Giuffrè, Milano, 2002, pp. 87 ss.

che impone equità e uniformità di accesso e presa in carico; il principio di appropriatezza organizzativa, a sua volta ricollegabile al principio di buon andamento della p.a., dal momento che tali strumenti migliorano l'efficienza (a parità di efficacia) dell'erogazione dei servizi; e il principio di autonomia e differenziazione, poiché permettono, nell'esercizio dell'autonomia organizzativa degli enti *sub*-statali, modalità di esecuzione delle prestazioni pubbliche in maniera più rispondente alle peculiarità territoriali<sup>28</sup>.

All'interno del d. Min. Salute 23 maggio 2022, n. 77 («Regolamento recante la definizione di modelli e *standard* per lo sviluppo dell'assistenza territoriale nel Servizio sanitario nazionale») – di attuazione della Missione 6 («Salute») del Piano nazionale di ripresa e resilienza – si sottolinea, non a caso, come la «stratificazione della popolazione» e il «monitoraggio per fattori di rischio» debbano servire, in quest'ordine: (i) a una «valutazione olistica dei bisogni dell'individuo», a sua volta finalizzata alla determinazione di un – personale e personalizzato – «Progetto di Salute»<sup>29</sup>; (ii) alla «programmazione e verifica dei risultati raggiunti dai servizi sanitari e socio-sanitari nella comunità di riferimento»<sup>30</sup>. I bisogni dell'individuo, della collettività e del sistema sembrano, in questo particolare caso, allinearsi perfettamente; va evidenziato, tuttavia, come nell'impiego su larga scala di algoritmi predittivi – è l'adozione di un «modello di stratificazione comune su tutto il territorio nazionale», si legge nel decreto, che «permetterà lo sviluppo di un linguaggio uniforme» e garantirà «equità di accesso ed omogeneità di presa in carico»<sup>31</sup> – si annidino rischi considerevoli, *in primis* di discriminazione.

<sup>28</sup> Così T. BALDUZZI, *La sanità digitale negli Stati decentrati: stratificazione e modelli predittivi del bisogno di salute nel riparto di competenze tra livelli territoriali*, in D. MORANA (a cura di), *La salute tra i diritti e nei territori. Questioni costituzionali nel rapporto Stato-Regioni*, Giappichelli, Torino, 2025, p. 186. V. anche EAD., *Gli indicatori del bisogno di salute: tra appropriatezza, autonomia ed equità nel Servizio sanitario nazionale*, in *federalismi.it*, 3/2025, spec. pp. 138 ss.

<sup>29</sup> Il quale consiste – come si legge nell'Allegato 1 («Modelli e *standard* per lo sviluppo dell'Assistenza Territoriale nel Servizio Sanitario Nazionale») al d. Min. Salute 23 maggio 2022, n. 77 («Regolamento recante la definizione di modelli e *standard* per lo sviluppo dell'assistenza territoriale nel Servizio sanitario nazionale»), al par. 3 – in «uno strumento di programmazione, gestione e verifica» che «associa la stratificazione della popolazione alla classificazione del “bisogno di salute”»; esso «si attiva all'accesso della persona al SSN» e tiene conto, oltre che di (eventuali) Progetti di assistenza individuale integrata (PAI) e Piani riabilitativi individuali (PRI), di tutta la documentazione acquisita nell'ambito dell'assistenza territoriale, consentendo la «definizione di azioni appropriate rispetto alle condizioni cliniche del paziente» e orientando «la persona e i professionisti nelle fasi di transizione tra i diversi *setting* di cura». Il Progetto di Salute, stando sempre al decreto, dovrà costituire «il filo conduttore che rappresenta la storia della persona e dei suoi bisogni clinico-assistenziali, in raccordo con i servizi sociali, seguendola prima ancora che sviluppi una patologia, dal momento in cui viene identificato [*sic*] come portatore di fattori di rischio o a partire dal primo contatto con il SSN, alimentando in modo coerente e tempestivo il relativo FSE»; il Progetto dovrà rappresentare, nel contempo, «uno strumento di programmazione, verifica e controllo della coerenza clinica e socio-assistenziale della presa in carico». L'idea, volendo semplificare, è che tutti i dati sanitari e socio-sanitari della persona debbano essere raccolti, aggregati e *valorizzati*, per consentire, a un tempo, un'attività di programmazione e controllo *individuale* (degli interventi assistenziali appropriati nel caso di specie) e *di comunità* (degli interventi – in senso ampio – necessari su larga scala).

<sup>30</sup> Cfr. d. Min. Salute 23 maggio 2022, n. 77 («Regolamento recante la definizione di modelli e *standard* per lo sviluppo dell'assistenza territoriale nel Servizio sanitario nazionale»), Allegato 1 («Modelli e *standard* per lo sviluppo dell'Assistenza Territoriale nel Servizio Sanitario Nazionale»), par. 3 (enfasi aggiunta).

<sup>31</sup> Per un'analisi, tra l'altro, dell'impatto del d. Min. Salute 23 maggio 2022, n. 77, sull'autonomia regionale in materia di stratificazione dei bisogni di salute, cfr. T. BALDUZZI, *La sanità digitale negli Stati decentrati*, cit., pp. 173 ss.; v. anche EAD., *Population Stratification for Public Health and the Ministry's Predictive Model: Constitutional Principles and Regulatory*

Sarà necessario fare particolare attenzione, come anticipato, alla formulazione del problema sottoposto al sistema, avendo cura di non valorizzare esageratamente i profili di contenimento della spesa; la stessa selezione delle variabili che si chiede al modello di analizzare (e.g., età della persona, eventuali patologie croniche o disabilità...) potrà incidere notevolmente sull'*output*.

Con riferimento ad alcuni algoritmi predittivi commerciali impiegati da strutture ospedaliere statunitensi a fini di stratificazione, ad esempio, un gruppo di ricercatori ha recentemente rilevato<sup>32</sup> come tendessero a sottostimare il bisogno di salute dei pazienti «*who self-identified as Black*» perché valutavano, nella distribuzione in classi di rischio/bisogno, i costi storici generati da ciascun individuo: le persone «*Black*», in America, si rivolgevano e si rivolgono più di rado al sistema sanitario, sia perché «*poor patients face substantial barriers to accessing health care, even when enrolled in insurance plans*»<sup>33</sup> (race e condizioni socio-economiche sono spesso interconnesse, negli Stati Uniti e non solo), sia perché «*Black patients have reduced trust in the health care system, a fact that some studies trace to the revelations of the Tuskegee study and other adverse experiences*»<sup>34</sup>.

Quella tra bisogno di salute e costi sanitari – perlopiù a carico dell'individuo in America, del servizio sanitario in Italia – è, fuor d'esempio, una falsa equivalenza, che rischia, se trasferita nel modello predittivo (o se tratteggiata autonomamente da quest'ultimo), di penalizzare gli appartenenti a gruppi minoritari *proprio perché storicamente penalizzati*. Il rischio, in termini più generali, è che una “formulazione del problema” frettolosa e la selezione di variabili connotate da criticità – l'utilizzo, in particolare, dei costi generati dalla persona come *proxy* del suo stato di salute – possano trasformare i modelli predittivi, da strumenti di «Medicina di Popolazione» e «Sanità di Iniziativa»<sup>35</sup>, in moltiplicatori di iniquità, portati a discriminare *direttamente* sulla base dello *status* socio-economico e *indirettamente* in ragione del genere, della “razza”, delle «condizioni personali» del paziente.

---

*Developments*, in *Corti supreme e salute*, 1/2024, pp. 287 ss., dove ci si interroga sul possibile utilizzo degli algoritmi di stratificazione, e in particolare del modello predittivo sviluppato dal Ministero della Salute a partire dal 2014 (oggi parte del PNRR), nella procedure di riparto tra le Regioni del Fondo sanitario nazionale.

<sup>32</sup> V. Z. OBERMEYER *et al.*, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, in *Science*, vol. 366, 2019, pp. 447 ss.

<sup>33</sup> Ivi, p. 450.

<sup>34</sup> *Ibidem*. La documentata sfiducia degli afroamericani nei confronti del *medical establishment* è stata sovente collegata alla «*historical disclosure of an unethical and deadly experiment, the Tuskegee Study of Untreated Syphilis in the Negro Male (TSUS)*», condotto tra il 1932 e il 1972 dallo stesso Ufficio per la Salute Pubblica statunitense: «*For 40 years (...) the U.S. Public Health Service (PHS) followed hundreds of poor, black men in Tuskegee, Alabama, the majority of whom had syphilis, for the stated purpose of understanding the natural course of the disease. The men were denied highly effective treatment for their condition (most egregiously, penicillin, which became standard of care by the mid-1940s) and were actively discouraged from seeking medical advice from practitioners outside the study. [...] News of the Tuskegee Study became public in 1972 in an exposé by Jean Heller of the Associated Press, and detailed narratives of the deception and its relationship to the medical establishment were widespread. By that point, the majority of the study's victims were deceased, many from syphilis-related causes*» (M. ALSAN, M. WANAMAKER, *Tuskegee and the Health of Black Men*, in *The Quarterly Journal of Economics*, 1/2018, pp. 408 s.).

<sup>35</sup> Cfr. d. Min. Salute 23 maggio 2022, n. 77, Allegato 1, *passim*.

Il caso degli algoritmi di stratificazione impiegati a fini di programmazione sanitaria, in sintesi, è particolarmente interessante, anche oltre le sue specificità, perché dà riprova che: (i) in ogni situazione della vita i singoli tratti identitari di ciascuno si pongono in inter-relazione, producendo effetti combinati<sup>36</sup>; (ii) (quasi) tutte le caratteristiche protette dell'individuo incidono, più o meno direttamente, sulla sua posizione socio-economica, la quale assurge, dunque, nel contempo a ricapitolazione e a minimo comun denominatore delle condizioni di minorità; (iii) non dare rilievo “a monte” a un tratto identitario non sempre equivale a escluderne l'impatto sulla decisione “a valle”, anzi: il rischio di discriminazioni “by proxy”<sup>37</sup>, anche emergenti<sup>38</sup>, è costantemente presente; (iv) una delle possibili soluzioni risiede in un approccio “consenso” – *race-conscious, gender-conscious...* – che celi aspirazioni riparative<sup>39</sup>, che valorizzi il contesto storico-culturale e chieda alla macchi-

<sup>36</sup> Che coincidono, sul versante “patologico”, con forme di discriminazione *multipla* o, più di frequente, *intersezionale*. La compresenza di condizioni di minorità, infatti, tende a creare un effetto moltiplicatore: raramente i singoli fattori di discriminazione saranno valutabili in maniera indipendente (c.d. discriminazione multipla); più comunemente, «i motivi di discriminazione si intrecciano, creando una tipologia unica di discriminazione», c.d. intersezionale (così definita nel Considerando A della Risoluzione del Parlamento europeo sulla «Discriminazione intersezionale nell'UE: la situazione socioeconomica delle donne di origine africana, mediorientale, latinoamericana e asiatica», del 6 luglio 2022). Quando più tratti identitari marginalizzanti si combinano, in altri termini, la loro incidenza sarà superiore alla somma delle discriminazioni e dello stigma risultanti da ciascun fattore considerato singolarmente. Nelle parole di Kimberlé Crenshaw, fondatrice ideale dell'intersezionalismo, che pensava soprattutto all'intrecciarsi del sesso e della “razza”: «*the intersectional experience is greater than the sum of racism and sexism*» (K. CRENSHAW, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, in *University of Chicago Legal Forum*, 1/1989, p. 140). Sulla nozione di intersezionalità v., nella dottrina italiana, almeno B.G. BELLO, *Diritto e genere visti dal margine: spunti per un dibattito sull'approccio intersezionale al diritto antidiscriminatorio in Italia*, in *Diritto e questioni pubbliche*, 2/2015, pp. 141 ss.; EAD., *Intersezionalità. Teorie e pratiche tra diritto e società*, FrancoAngeli, Milano, 2020; nonché, per un'intervista della stessa B.G. BELLO e di L. MANCINI a Kimberlé Crenshaw, *Talking about Intersectionality. Interview with Kimberlé Crenshaw*, in *Sociologia del diritto*, 2/2016, pp. 11 ss. Sullo specifico rischio di discriminazione/i intersezionale/i connesso alle applicazioni medico-sanitarie dell'intelligenza artificiale, e in particolare alla «*AI-powered m-Health*», v., con maggiore riguardo al contesto statunitense, M.A. WÓJCIK-SUFFIA, *Algorithmic Discrimination in M-Health: Rethinking the US Non-Discrimination Legal Framework Through the Lens of Intersectionality*, in *BioLaw Journal – Rivista di BioDiritto*, 1/2024, pp. 367 ss.

<sup>37</sup> Ossia di discriminazioni “indirette”, che non si basano sulla caratteristica distintiva protetta («sesso», «razza» *etc.*) bensì su un elemento altro – il *proxy*, appunto – formalmente “neutro” ma sovente associato, nel concreto, alla caratteristica protetta esclusa.

<sup>38</sup> Se è vero, infatti, come sottolinea P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., p. 462, che «le *proxy discriminations* rappresentano una pratica risalente, cui si è fatto ricorso anche prima della nascita dei sistemi di *machine learning* per occultare pratiche discriminatorie», nondimeno «[l]a novità rappresentata dall'avvento dell'intelligenza artificiale è legata alla circostanza che in questo caso il carattere *proxy* potrebbe essere individuato autonomamente dal sistema di apprendimento automatico, senza che l'utilizzatore ne sia consapevole: così, ad esempio, se un sistema informatico generasse autonomamente un modello che impone premi più elevati per l'assicurazione sanitaria a richiedenti che risultino iscritti ad un gruppo *Facebook* costituito per sollecitare un aumento della disponibilità del *test* genetico per le mutazioni dei geni BRCA correlate ad un alto rischio di sviluppare alcuni tumori, opererebbe una *proxy discrimination* sulla base dei dati genetici» (l'esempio è tratto da A.E.R. PRINCE, D. SCHWARCZ, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review*, 3/2019-2020, p. 1261). Come evidenzia ancora P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., p. 463, oltretutto, «le *proxies* di origine umana sono più facilmente individuabili e sono assoggettabili ad un controllo preventivo; quelle “inintenzionali”, che originano dagli algoritmi, risultano invece di norma più difficilmente individuabili e tendono di conseguenza a sfuggire alle successive verifiche».

<sup>39</sup> Nelle parole di J.L. DAVIS, A. WILLIAMS, M.W. YANG, *Algorithmic reparation*, in *Big Data & Society*, 2/2021, p. 1, che propongono la “fusione” di intersezionalismo e prassi riparative e la loro applicazione all'algoritmica: «*By default, technologies reflect and reinforce existing social orders, expressing and materializing hierarchical relations. However, technologies*

na di tenerne conto *non solo* per evitare nuove discriminazioni, *ma anche*, soprattutto, per correggere o depotenziare distorsioni già esistenti o emergenti.

### 3. Dalle logiche dell'anti-classificazione a quelle dell'anti-subordinazione, dalla mera *fairness* a veri e propri algoritmi "riparativi"

È stato sostenuto che le politiche "riparative" (*i.e.*, volte a porre rimedio alle discriminazioni *sistemiche* e, più in generale, alle "ingiustizie" storicamente perpetrate nei confronti di formazioni e gruppi sociali) debbano radicarsi «nella giustizia distributiva e nel controfattuale»<sup>40</sup>: occorre chiedersi come sarebbe strutturata la società – come sarebbe distribuita la ricchezza, come sarebbero allocate le risorse – se non vi fossero state quelle discriminazioni, quelle "ingiustizie", e tentare di avvicinare la realtà materiale a quella controfattuale, ipotetica e *alternativa*.

Se è vero, come è stato detto, che i sistemi di intelligenza artificiale spesso tendono ad «accelerare, esacerbare e amplificare l'impatto di "forze di oppressione" preesistenti»<sup>41</sup> – perché, tra le altre cose, ricevono dati non sufficientemente rappresentativi delle differenze e disparità esistenti, o dati *biased* presentati e recepiti come "neutri"<sup>42</sup> – è anche vero che quegli stessi sistemi possono essere utilizzati, in chiave riparativa, per individuare alcune storture e tentare di correggerle.

In un recente studio<sup>43</sup>, ad esempio, si è fatto ricorso a sistemi di intelligenza artificiale, e in particolare a modelli predittivi, per tentare proprio di "predire" l'efficacia e l'impatto sociale di plurimi, ipotetici *Special Purpose Credit Programs (SPCPs)*, programmi pubblici ideati per favorire l'accesso al credito di gruppi storicamente marginalizzati (ed esclusi, dunque, da diverse forme di ricchezza). Alcuni *SPCPs* seguivano un approccio *race-neutral*, altri adottavano un linguaggio *race-conscious*. Soltanto questi ultimi, stando ai modelli, si sarebbero dimostrati effettivamente in grado di *favorire* la comunità *African-American*, contribuendo così – ancorché in maniera indiretta – a ridurre il «*racial housing wealth gap*», che era l'obiettivo designato.

---

*can also be tools of liberation. They can expose, undo, and reshape status quos. This latter project necessitates concerted and targeted efforts, underpinned by socially informed perspectives. In service of such efforts, we present algorithmic reparation as a concept and a scaffold for Intersectional approaches to machine learning (ML) systems, displacing fairness in favor of redress. Beyond improving code, a reparative approach uses computational tools for social intervention, while critically assessing when and where computation does not belong».*

<sup>40</sup> W. SO, C. D'IGNAZIO, *Race-Neutral vs Race-Conscious: Using Algorithmic Methods to Evaluate the Reparative Potential of Housing Programs*, in *Big Data & Society*, 2/2023, p. 3, tr. it. mia.

<sup>41</sup> Ivi, p. 5, tr. it. mia.

<sup>42</sup> V. *supra*, § 1.

<sup>43</sup> Il riferimento è sempre a W. SO, C. D'IGNAZIO, *Race-Neutral vs Race-Conscious*, cit.



Oltre i confini dello studio, sembrerebbe di poter trarre due distinte conclusioni: che una politica concretamente (re-)distributiva non può prescindere dal dare rilievo alle disparità che mira a risanare; che, in seconda battuta, ogni apparente favoritismo – ogni “discriminazione alla rovescia”<sup>44</sup> – può essere inquadrato o inquadrata, in ottica (re-)distributiva, come un passaggio necessario nell’approssimazione della realtà materiale a quella (controfattuale) più equa. Se la popolazione *African-American* è stata sistematicamente sfavorita nell’accesso al credito e, anche di conseguenza, alla proprietà immobiliare – prima in maniera diretta, attraverso politiche segregazioniste, poi con modalità più subdole, dal c.d. *redlining* alla individuazione delle comunità *post-segregate* come *target* d’elezione per i mutui *subprime* – l’adozione di misure correttive e latamente promozionali non avrà i tratti di un *favor*, ma di una riparazione. Può essere utile immaginare una realtà (controfattuale) nella quale non vi siano o non vi siano stati schiavismo, segregazione, discriminazioni dirette e indirette, e utilizzare quello scenario ipotetico come parametro e come obiettivo: questo aiuterà, tra l’altro, a inquadrare le operazioni di re-distribuzione delle risorse, più opportunamente, come attività di “correzione del privilegio” e non di indebita sottrazione<sup>45</sup>.

Sembrerebbe di poter affermare, poi – nel solco dello studio sui *SPCPs* e muovendo oltre quest’ultimo – che l’intelligenza artificiale sia in grado, a determinate condizioni, di agevolare la suddetta operazione di approssimazione della realtà materiale a quella controfattuale, sia fornendo letture del tempo presente, sia aiutando a “predire” scenari futuri e, dunque, a indirizzare la programmazione e le attività pubbliche. *Context is key*: perché i sistemi di intelligenza artificiale possano effettivamente aiutarci a immaginare e a realizzare un futuro più equo, non basterà “depurare” i dati, e dunque rimuoverne i tratti stigmatizzanti, assicurarsi che siano rappresentativi; sarà necessario, in aggiunta a questo, *arricchire i dati di contesto* – storico, socio-economico – perché il sistema possa, non solo divenire meno dannoso, ma rivelarsi uno strumento e un veicolo positivo di eguaglianza. Nello specifico caso degli algoritmi di stratificazione, ad esempio, una soluzione possibile è addestrare la macchina a “sovrastimare” il rischio rilevato con riferimento agli appartenenti a determinate categorie protette, dal momento che quella “sovrastima” sarebbe in realtà correttiva di una “sottostima” fisiologica (*i.e.*, dovuta a peculiarità intrinseche del/nel contesto socio-economico-culturale).

Ai sistemi di intelligenza artificiale “deleghiamo” operazioni che per gli esseri umani sarebbero assai complesse o, talvolta, impossibili: chiediamo loro di analizzare sterminate

<sup>44</sup> V. *supra*, spec. nt. 5.

<sup>45</sup> Secondo W. SO, C. D’IGNAZIO, *Race-Neutral vs Race-Conscious*, cit., p. 4, «there is a portion of white wealth that cannot be rightfully claimed by white people and should not be protected by policy because it was stolen and expropriated through exclusionary practices. By centering distributive justice as an objective of reparative policies, we can reorient the reparative policy debate from the legal protection of “innocent” individual white people to redressing the intergenerational structural harms enacted by chattel slavery; then we can ask and measure the counterfactual: in the absence of racial oppression, what would be the present-day distribution of wealth between Black people and white people?».

quantità di dati, rinvenire dei *pattern*<sup>46</sup> – delle relazioni frequenti, che “tornano”, tra variabili inter-connesse ed esiti – ed effettuare *predizioni* sulla scorta di quanto ravvisato. La completezza e la rappresentatività dei dati sono strumentali *in primis* al buon funzionamento del sistema e all’accuratezza *across-the-board* dei risultati offerti in *output*, ma servono anche, come visto<sup>47</sup>, a garantire la *fairness* del sistema stesso (*i.e.*, la correttezza e l’imparzialità *formali*).

Sul piano *sostanziale*, tuttavia – quello della rimozione degli «ostacoli di ordine economico e sociale», per attingere a un lessico giuridico-costituzionale – c’è bisogno di qualcosa in più: i dati, per quanto completi e rappresentativi possano essere, costituiscono una “fotografia” dello *status quo*, con tutte le sue imperfezioni e storture (e dis-eguaglianze). Se i dati medico-sanitari, ad esempio, vengono raccolti nell’ambito della pratica clinica quotidiana e delle sperimentazioni, non si può non prendere atto, quando li si dà “in pasto” alla macchina, della grave sotto-rappresentazione delle donne all’interno delle sperimentazioni medesime, nonché, più in generale, di come le strutture patriarcali informino tanto la scienza medica che l’erogazione dei servizi sanitari e socio-sanitari<sup>48</sup>.

<sup>46</sup> Come sottolineato da A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell’intelligenza artificiale*, cit., p. 98, i sistemi di apprendimento automatico, e le reti neurali in particolare, «basano le loro predizioni» – ma lo stesso può dirsi di decisioni, stime, misure – «su *pattern* nascosti all’occhio del programmatore inferiti direttamente dai dati» (motivo per cui, tra l’altro, «non sappiamo di fatto quali correlazioni stiano sfruttando per costruire l’*output* fornito»).

<sup>47</sup> *Supra*, spec. § 1.

<sup>48</sup> Storicamente, infatti, la condizione maschile era considerata la norma, lo *standard*, e quella femminile un accessorio, una deviazione. Le sperimentazioni erano (e sono, spesso) condotte soprattutto su uomini; le specificità dell’essere-donna, nel corpo come nell’esperienza della realtà, venivano (e vengono) sistematicamente trascurate. «Nell’ambito della sperimentazione clinica», denunciava il Comitato Nazionale per la Bioetica già nel 2008, «le donne risultano essere “soggetti deboli”, o comunque non oggetto di adeguata considerazione in ordine alla loro specificità sia in senso quantitativo (numero di donne arruolate rispetto al numero di uomini) sia in senso qualitativo (analisi dei dati rispetto alla differenza sessuale)» (COMITATO NAZIONALE PER LA BIOETICA, *La sperimentazione farmacologica sulle donne*, cit., p. 8); considerazioni analoghe si rinvenivano in MINISTERO DELLA SALUTE, *Il genere come determinante di salute. Lo sviluppo della medicina di genere per garantire equità e appropriatezza della cura*, in *Quaderni del Ministero della Salute*, 2016, pp. 74 s., dove si sottolinea, tra l’altro, come i farmaci vengano studiati, generalmente, «su un tipo di maschio ideale per lo più giovane, bianco e sui 70 kg di peso», e finiscano per «non garanti[re] alle donne gli stessi risultati di beneficio documentati negli uomini». Alcuni dei passi compiuti da allora, anche sul piano simbolico – ad esempio, la creazione di un Centro di Riferimento per la Medicina di Genere presso l’Istituto Superiore di Sanità – sono ripercorsi in MINISTERO DELLA SALUTE, *Piano per l’applicazione e la diffusione della Medicina di Genere*, 6 maggio 2019, pp. 13 ss., redatto in attuazione dell’art. 3, co. 1, l. 11 gennaio 2018, n. 3 («Delega al Governo in materia di sperimentazione clinica di medicinali nonché disposizioni per il riordino delle professioni sanitarie e per la dirigenza sanitaria del Ministero della salute»), che per la prima volta collocava la medicina di genere all’interno del Servizio sanitario nazionale (cfr. M. TOMASI, *Sperimentazioni cliniche e medicina di genere. La ricerca dell’eguaglianza attraverso la valorizzazione delle differenze*, in B. PEZZINI, A. LORENZETTI [a cura di], *70 anni dopo tra uguaglianza e differenza. Una riflessione sull’impatto del genere nella Costituzione e nel costituzionalismo*, Giappichelli, Torino, 2019, p. 229). Si v., sul punto, anche il Reg. (UE) n. 536/2014 del Parlamento europeo e del Consiglio, del 16 aprile 2014, «sulla sperimentazione clinica di medicinali per uso umano e che abroga la direttiva 2001/20/CE», entrato in vigore nel gennaio 2022, che già al Considerando (14) precisa come, «[s]alvo diversa disposizione motivata nel protocollo, i soggetti che partecipano a una sperimentazione clinica dovrebbero essere rappresentativi dei gruppi della popolazione, ad esempio i gruppi per sesso ed età, che siano tra i possibili utilizzatori del medicinale oggetto della sperimentazione»; stando al pt. 17, lett. y), dell’Allegato I, il c.d. protocollo, che va incluso nel fascicolo di domanda (di autorizzazione alla sperimentazione) iniziale, dovrà contenere, fra le altre cose, «una giustificazione della scelta del genere e dell’età dei soggetti e, nel caso in cui un determinato gruppo di genere o di età sia escluso dalle sperimentazioni cliniche o vi risulti sottorappresentato, una spiegazione dei motivi e una giustificazione dei criteri di esclusione». Come rilevato ancora in MINISTERO DELLA SALUTE, *Il*

Sarà necessario, allora – perché l'intelligenza artificiale possa (contribuire a) rimediare ad alcune delle storture e dis-eguaglianze che caratterizzano la realtà presente – fornire questo contesto alla macchina, “istruendola” a tenerne conto in sede di elaborazione degli *output*. La sotto-rappresentazione delle donne all'interno delle sperimentazioni cliniche e farmacologiche fa sì che, tra l'altro, le stesse siano più esposte al rischio di reazioni avverse ai farmaci (*adverse drug reactions*, o *ADRs*)<sup>49</sup>. Sul piano del dover-essere, ovviamente, sarà importante adottare misure che contrastino il fenomeno (multi-causa<sup>50</sup>) della sotto-rappresentazione<sup>51</sup>; su quello scientifico e medico, nel frattempo, si potrà, tra le altre cose, “delegare” ad algoritmi di farmacovigilanza – come *AwareDX*, recentemente sviluppato<sup>52</sup> – il calcolo del «rischio differenziale» che un farmaco *x* cagioni reazioni avverse negli uomini o nelle donne, e dunque l'individuazione di eventi avversi ai farmaci «specifici per sesso» nonché, nell'ottica del singolo, un'accurata “predizione” circa l'appropriatezza di ciascuna terapia farmacologica per pazienti specifici individuati sulla base del sesso e/o del genere<sup>53</sup>. Com'è evidente, un sistema di questo tipo non si limita a non-aggravare i «*bias* basati sul sesso» che caratterizzano la sperimentazione e la somministrazione di far-

---

*genere come determinante di salute*, cit., pp. 74 s., «[l']approccio cautelare legato alla sperimentazione clinica nelle donne» tende a essere ricondotto «a ragioni sociali, ambientali, economiche e soprattutto biologiche (variazioni ormonali come ciclo mestruale, gravidanza, allattamento, menopausa o assunzione di anticoncezionali)»; tutte queste “condizioni”, tuttavia, «sono caratteristiche costanti della popolazione femminile che utilizzerà il farmaco una volta in commercio».

<sup>49</sup> Cfr., *ex multis*, A.C. BRABETE *et al.*, *A Sex- and Gender-Based Analysis of Adverse Drug Reactions: A Scoping Review of Pharmacovigilance Databases*, in *Pharmaceuticals*, 3/2022, *passim*. V. anche, con maggiore attenzione al contesto interno e per ulteriori riferimenti bibliografici, MINISTERO DELLA SALUTE, *Il genere come determinante di salute*, cit., spec. pp. 73 ss., dove si sottolinea come le donne rispondano ai farmaci in maniera diversa rispetto agli uomini in virtù di differenze «fisiologiche, anatomiche e ormonali»: le donne stesse, infatti, «mostrano un profilo farmacocinetico diverso rispetto all'uomo sia per assorbimento, sia per distribuzione, metabolismo ed eliminazione del farmaco», e «[h]anno un peso corporeo medio inferiore, una percentuale di massa grassa più alta, un minore volume plasmatico e un profilo di legame tra farmaco e proteine plasmatiche diverso»; non può sorprendere, allora – se, come anticipato, «il farmaco viene studiato su un tipo di maschio ideale per lo più giovane, bianco e sui 70 kg di peso» – che molti farmaci non garantiscano alle donne «gli stessi risultati di beneficio documentati negli uomini» e che le donne medesime risultino «maggiormente esposte a possibili reazioni avverse», poiché ricevono terapie (in media) sovradosate e sono immunologicamente più “forti”.

<sup>50</sup> V. *supra*, nt. 20.

<sup>51</sup> Come in parte è già accaduto: si pensi alle disposizioni sul punto contenute nel Reg. (UE) n. 536/2014 del Parlamento europeo e del Consiglio, del 16 aprile 2014, «sulla sperimentazione clinica di medicinali per uso umano e che abroga la direttiva 2001/20/CE», riguardo al quale più diffusamente *supra*, nt. 48. Secondo M. FASAN, C.M. REALE, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, un'occasione mancata?*, in *BioLaw Journal – Rivista di BioDiritto*, 4/2022, pp. 251 ss., tuttavia, il Regolamento in questione – che pure ha «innovato, in termini di maggiore tutela e garanzia, la disciplina precedente» – ha nondimeno rappresentato «un'occasione mancata per la tutela e la promozione dell'eguaglianza di genere nella sperimentazione clinica», sia perché «adotta una nozione di genere come sinonimo di donne, non soffermandosi sulla necessità di considerare le differenze derivanti dalle variazioni sessuali né tantomeno sui meccanismi socio-culturali connessi al sistema di genere binario», sia, soprattutto, perché «l'approccio normativo seguito dal legislatore europeo risulta disallineato» e “arretrato” «rispetto alle esperienze normative nazionali (...) e al concetto di medicina di genere (...) adottato e riconosciuto» già in quegli anni.

<sup>52</sup> V. P. CHANDAK, N.P. TATONETTI, *Using Machine Learning to Identify Adverse Drug Effects Posing Increased Risk to Women*, in *Patterns*, 7/2020.

<sup>53</sup> Ivi, *passim*, tr. it. mia.

maci, ma contribuisce a mitigarli<sup>54</sup>, contestualmente (e incidentalmente) approssimando la realtà presente della pratica clinica, informata da secoli di distorsioni patriarcali, a una realtà (controfattuale) più equa.

Si tratta, detto altrimenti, di superare le logiche della mera *fairness* e dell'anti-classificazione in favore di un approccio anti-subordinazione. (Soprattutto) nella dottrina nord-americana, l'idea che un processo decisionale – umano o artificiale che sia – debba ritenersi equo quando *non* prende in considerazione alcun tratto identitario protetto della persona interessata (la “razza”, il genere, la religione *etc.*) è ricondotta al c.d. principio anti-classificazione; quest'ultimo, in opposizione al paradigma dell'anti-subordinazione, prescriverebbe un identico trattamento di coloro che – a prescindere, di nuovo, dalla “razza”, dal genere, dalla religione – si trovino in situazioni assimilabili<sup>55</sup>. Un sistema di intelligenza artificiale, in quest'ottica, sarà giudicato equo se non riconosce rilevanza ai tratti protetti di cui sopra (o a *proxy* di quei tratti) e se offre un livello di *performance* uniforme quando impiegato su/per/con riferimento a persone appartenenti a gruppi sociali diversi, marginalizzati e non<sup>56</sup>.

Non sempre, tuttavia, la “cecità” – rispetto al colore, al sesso – porta a decisioni *giuste*. In particolare quando il processo decisionale è automatizzato, anzi, appare molto discutibile che ignorare la particolare conformazione della società *nella* quale e *sulla* quale il sistema di intelligenza artificiale è destinato a dispiegare i suoi effetti possa contribuire alla equità

<sup>54</sup> Ivi, spec. pp. 2 s., tr. it. mia. Occorre riconoscere, ad ogni modo – sottolinea L. PIVA, *Vulnerability in the Age of Artificial Intelligence: Addressing Gender Bias in Healthcare*, in *BioLaw Journal – Rivista di BioDiritto*, spec./2024, p. 175 – che le «technical solutions» non sono pienamente in grado di «answer the underlying societal problems that have led to the creation of gender stereotypes or to disregard the importance of conducting medical research adopting a gender perspective» (enfasi aggiunta). Le soluzioni tecniche e tecnologiche, detto altrimenti, possono aiutare ad arginare le conseguenze, ma non correggono le cause (storiche, socio-economiche...) delle differenze e delle dis-parità.

<sup>55</sup> Si parla di principio “anti-classificazione” perché, «roughly speaking» – come osservano, *ex multis*, J.M. BALKIN, R.B. SIEGEL, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, in *University of Miami Law Review*, 9/2003, p. 10 – «this principle holds that the government may not classify people either overtly or surreptitiously on the basis of a forbidden category: for example, their race»; fuor di contesto, le medesime considerazioni possono applicarsi ai processi decisionali affidati a soggetti non-istituzionali, umani o artificiali che siano. Sulla contrapposizione tra anti-classificazione e anti-subordinazione – non necessariamente in questi esatti termini: cfr. J.M. BALKIN, R.B. SIEGEL, *ivi*, p. 9 – v. anche, almeno, O.M. FISS, *Groups and the Equal Protection Clause*, in *Philosophy & Public Affairs*, 2/1976, pp. 107 ss., primo teorico dell'anti-subordinazione e critico dell'anti-classificazione (che lui chiamava, tuttavia, forse poco opportunamente, anti-discriminazione; contrari a questa nomenclatura, tra gli altri, J.M. BALKIN, R.B. SIEGEL, *ivi*, p. 10); D. BELL, *And We Are Not Saved: The Elusive Quest for Racial Justice*, Basic Books, New York, 1989, *passim*; K.L. KARST, *Belonging to America: Equal Citizenship and the Constitution*, Yale University Press, New Haven, 1989, *passim*; C.A. MACKINNON, *Feminism Unmodified: Discourses of Life and Law*, Harvard University Press, Cambridge (MA), 1987, soprattutto pp. 32 ss.; in ambito interno v., con particolare attenzione alle dinamiche di sesso/genere, B. PEZZINI, *Costruzione del genere e Costituzione*, cit., soprattutto pp. 49 ss., dedicate proprio a d'uguaglianza uomo-donna come principio anti-subordinazione.

<sup>56</sup> Cfr., anche per ulteriori riferimenti, V. KESWANI, E. CELIS, *An Anti-Subordination Approach to Fair Classification*, 2022, disponibile su: [papers.ssrn.com/abstract=4116835](https://papers.ssrn.com/abstract=4116835). Come rilevano le Autrici, «anti-classification has been successful in mitigating biases in many contexts (...). Crucially, in settings that are marked by historic discrimination against minority groups, guaranteeing the fact that protected attributes of individuals are not considered by a decision-maker is an important step towards equitable treatment. Nevertheless» – come si vedrà – «anti-classification cannot address all kinds of discrimination» (*ivi*, p. 6).

di quest'ultimo. Come dimostrato in diversi ambiti – si pensi al *predictive policing* e al c.d. caso COMPAS<sup>57</sup> – la mancata integrazione nell'*iter* decisionale automatizzato di riferimenti “consapevoli” al contesto sociale sullo sfondo finisce per determinare una *riproduzione e moltiplicazione* delle storture – dei *bias* – propri della decisione umana e che il ricorso ad algoritmi dovrebbe/potrebbe contribuire a correggere o a rimuovere.

#### 4. Algoritmi “riparativi” e azioni positive: considerazioni conclusive

Vi è una differenza essenziale tra algoritmi come *AwareDX* e i modelli predittivi impiegati nello studio sui *SPCPs*. I primi hanno come obiettivo precipuo l'accuratezza dei risultati in *output*, la “vicinanza alla realtà” delle classificazioni o previsioni effettuate dalla macchina; i secondi sono più propriamente strumenti di giustizia sociale, pensati per modificare la realtà stessa in senso più equo. Ai primi, detto altrimenti, si chiede di considerare la dif-

<sup>57</sup> Alcuni anni fa, un'inchiesta di *ProPublica* (J. ANGIN et al., *Machine Bias. There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, in *ProPublica*, 23 maggio 2016, disponibile su: [www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing](http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)) ha portato alla luce alcune problematiche connesse all'utilizzo di un sistema, noto come COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), impiegato negli Stati Uniti per coadiuvare la stima del rischio di recidiva, anche al fine di valutare l'opportunità che sia applicata al condannato la misura della (*rectius*, una misura equivalente alla) libertà vigilata; COMPAS, in particolare, è stato osservato, tende, pur avendo un tasso di accuratezza complessivamente comparabile tra le “razze”, a *sovrastimare* il rischio di recidiva negli *African-American* e a *sottostimare* il medesimo rischio nei bianchi, producendo un numero di falsi positivi doppio per i primi rispetto ai secondi (accade l'esatto opposto, invece, con riguardo ai falsi negativi). Ebbene: il questionario sottoposto ai «*prospective parolees*» e usato nella valutazione, che *non* fa riferimento alla “razza”, contiene domande relative ai precedenti penali e alla storia criminale della persona interessata, ai suoi titoli di studio, alla sua occupazione, alla situazione finanziaria, alle “caratteristiche” del suo quartiere (*neighborhood*) d'origine o residenza; come è stato messo in evidenza, «*all these variables are structured by racial domination – from job market discrimination to ghettoization – the survey measures the extent to which an individual's life chances have been impacted by racism without ever asking an individual's race*» (R. BENJAMIN, *Race After Technology*, Polity, Cambridge, 2019, p. 82). Molti dei fattori presi in considerazione per valutare la pericolosità sociale dell'individuo, in altri termini, sono strettamente connessi alla “razza” (o meglio, alle conseguenze socio-economiche dell'appartenenza a una “razza”); la scelta di utilizzare quei parametri e non altri è *politica*, come *politica* sarebbe la decisione di escluderli o riconfigurarli. Non solo: che gli *African-American* versino (tendenzialmente) in condizioni socioeconomiche deteriori e che i quartieri a prevalenza *African-American* presentino indici di criminalità più elevati *non* è un caso, ma è diretta conseguenza (delle molte sfaccettature del) razzismo sistemico. La non-rilevanza della “razza”, detto altrimenti, è un'illusione: i sistemi che non prendono in considerazione la “razza”, come COMPAS, finiscono per dimostrarsi, spesso, quanto e più iniqui di quelli *race-conscious*. Sul “caso COMPAS” v. anche, tra i contributi in lingua italiana, C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pubbl. comp. eur.*, spec./2019, pp. 114 s.; A. CELOTTO, *Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto*, in *Analisi giuridica dell'economia*, 1/2019, pp. 47 ss.; G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Pol. dir.*, 2/2019, pp. 215 s.; A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale*, cit., pp. 93 ss. – questi ultimi, in particolare, rilevano, a partire da e oltre il “caso COMPAS”, quanto sia estesa la «forza di argomentazione pratica dell'algoritmo», che pare sottovalutata dai sostenitori del (pur centrale) principio di non-esclusività della decisione algoritmica (art. 22 GDPR): il principio in questione, si legge, «sembra non riconoscere l'evidente forza “pratica” di qualsiasi automatismo valutativo che, da un lato, solleva il decisore dal *burden of motivation*, dal peso dell'esame e della motivazione, dall'altro, gli consente di “etichettare” la propria decisione con quel crisma di “scientificità”, ovvero di “neutralità”, che circonda la valutazione algoritmica» (ivi, p. 100), con tutte le conseguenze del caso.



ferenza *nella misura in cui* abbia un'incidenza pratica; ai secondi di soppesare quell'incidenza in ottica riparativa.

Alla base del ricorso agli uni e agli altri, tuttavia, vi sono considerazioni e intenti simili; vi è, in particolare, la piena consapevolezza di quanto sia importante, per giungere a esiti equi, che si analizzi l'impatto delle differenze e si riconosca loro rilievo. Gli algoritmi *stricto sensu* "riparativi", come i modelli predittivi impiegati nello studio sui *SPCPs*, muovono oltre queste premesse comuni e servono, in senso più proprio, a correggere le storture prodottesi nello storico stratificarsi di "parzialità" e iniquità; essi non si limitano, detto altrimenti, a non-aggravare o finanche a mitigare i *bias*, ma forniscono al decisore (in senso a-tecnico) gli strumenti per operare una vera e propria, ancorché parziale, *redistribuzione* di risorse.

Questa medesima impronta redistributiva connota anche, guardando al diritto, le *azioni positive*<sup>58</sup>.

Esse sono spesso inquadrare, nella prospettiva costituzionale, come deroghe al principio di eguaglianza formale, talvolta come vere e proprie "discriminazioni alla rovescia"<sup>59</sup>. Se, tuttavia, è improprio parlare di "discriminazioni" per riferirsi a condotte e contegni, pure illeciti o illegittimi o comunque ingiusti, che muovano *contro* la grana delle gerarchie socio-culturali<sup>60</sup>, è altrettanto discutibile che una differenziazione in funzione promozionale o inclusiva debba necessariamente rappresentare una *deroga* all'eguaglianza, ancorché formale. Appare quantomeno ipotizzabile una lettura delle azioni positive che non le

<sup>58</sup> Non vi è spazio, in questa sede, per una riflessione compiuta attorno alla nozione di "azioni positive": basti ricordare, con la migliore dottrina, che si tratta di un «modello 'plurale', al cui interno vanno ricomprese tanto le azioni positive in senso stretto, miranti a realizzare *immediatamente e direttamente* obiettivi di inclusione – l'esempio più immediato, ovviamente, è quello delle "quote" – quanto le misure latamente promozionali, orientate, piuttosto, a creare le condizioni perché l'inclusione possa, quasi *naturalmente*, realizzarsi; il tratto comune a interventi di questa sorta risiederebbe, con ogni evidenza, «nella ambizione a fungere non tanto da limite negativo (nei confronti di pratiche discriminatorie) quanto da fattore di cambiamento, per questo 'positivo' (o 'affirmative' nella dizione americana), delle strutture che hanno (anche solo) oggettivamente favorito i processi di discriminazione e di penalizzazione a carico di alcuni gruppi o categorie» (così A. D'ALOIA, *Eguaglianza sostanziale e diritto diseguale. Contributo allo studio delle azioni positive nella prospettiva costituzionale*, CEDAM, Padova, 2002, p. 247). Caratteristiche immancabili delle azioni positive, stando a questa (condivisibile) ricostruzione, parrebbero essere: (i) sul piano della struttura, l'impostazione "in positivo", che le distingue con immediata evidenza dai meri divieti di discriminazione; (ii) sul piano eziologico, l'esigenza di contrastare forme *sistemiche* di esclusione, connesse al "funzionamento" della società più che a sommatorie di comportamenti individuali; (iii) sul piano dei fini, la tendenza a *promuovere* piuttosto che a *proteggere*. Sull'impostazione – non compensatoria-risarcitoria, ma – promozionale e, soprattutto, redistributiva (di opportunità e risorse) delle azioni positive, v. ancora A. D'ALOIA, *ivi*, pp. 127 ss., spec. pp. 216 ss.

<sup>59</sup> Cfr., *ex multis*, A. PACE, *Eguaglianza e libertà*, cit., pp. 171 s., il quale si interroga sulla «ammissibilità costituzionale di *affirmative actions* per imporre una "discriminazione alla rovescia" in favore del sesso femminile». V. anche M. AINIS, *Cinque regole per le azioni positive*, in *Quad. cost.*, 2/1999, p. 364, a parere del quale «non si può (...) sostenere, come hanno ripetutamente fatto i *leaders* democratici statunitensi (da Clinton a Ted Kennedy[]), che bast[i] eliminare il sistema delle quote perché l'*affirmative action* non sia più discriminante, perché non comporti più una *reverse discrimination*»; l'unica "colpa" di cui le quote sarebbero responsabili, prosegue l'A., «è (...) a ben vedere un merito: quello di mettere in chiaro l'effetto dirompente delle azioni positive, delle misure di "diritto diseguale", sul principio di astratta parità degli individui».

<sup>60</sup> Sulla non-neutralità della nozione di "discriminazione" v. *supra*, § 1, spec. nt. 5.

ponga in contrapposizione con le logiche dell'eguaglianza, dipingendole come “discriminazioni alla rovescia”, né le intenda esclusivamente quali correttivi della sua declinazione giuridico-formale in chiave valutativa<sup>61</sup> o di eguaglianza sostanziale<sup>62</sup>. Il nodo è, forse, nella valorizzazione del significato “polemico”<sup>63</sup> e della portata trasformatrice del principio (storico e storicizzato) di eguaglianza: se si intende l'eguaglianza, anche formale, come forza trasformativa, le azioni positive non si porranno come deroghe, ma si collocheranno *nel solco* di quell'eguaglianza. La differenza tra eguaglianza formale ed eguaglianza sostanziale, in quest'ottica, si collocherebbe esclusivamente sul piano degli strumenti – “negativi” da un lato, “positivi” dall'altro – e non, invece, sul piano dei significati, dei contenuti<sup>64</sup>. Non vi è spazio, in questa sede, per una riflessione completa o approfondita sul rapporto tra eguaglianza formale e sostanziale, né per una proposta di “rilettura” delle azioni positive. Il riferimento a queste ultime, tuttavia, può essere d'aiuto per giustificare e fondare – *in primis* sul piano giuridico-costituzionale – il ricorso ad algoritmi “riparativi”, o comunque a sistemi di intelligenza artificiale “consapevoli” (*race-conscious, gender-conscious...*).

<sup>61</sup> Secondo parte della dottrina, infatti, l'adozione di misure latamente promozionali potrebbe “giustificarsi” alla luce di una lettura propriamente *valutativa* – e non *paritaria* – del principio di eguaglianza formale. Se è vero, come è vero, che “eguaglianza” non vuol dire “*unicuique idem*”, ma “*unicuique suum*” – che la regola dell'eguaglianza, lungi dall'imporre di trattare tutto e tutti allo stesso modo, prescrive, piuttosto, un trattamento adeguatamente differenziato di situazioni dissimili – già su queste basi potrebbe fondarsi l'adozione di azioni positive. Non sarebbe necessario, in altri termini, scomodare l'eguaglianza sostanziale: le «esigenze (...) dell'eguaglianza ‘effettiva’» sarebbero riconducibili, e agevolmente, «nell'ambito delle differenziazioni ragionevoli» (così A. D'ALOIA, *Eguaglianza sostanziale e diritto diseguale*, cit., p. 242, che pure si discosta, almeno in parte, da questa ricostruzione).

<sup>62</sup> Solitamente, la “giustificazione” costituzionale delle azioni positive è rinvenuta nell'art. 3, co. 2, Cost., e dunque nel principio di eguaglianza sostanziale, il quale si porrebbe in tensione con le logiche dell'eguaglianza formale-procedurale e opererebbe come loro quasi-correttivo. «È compito della Repubblica», si legge nel secondo comma dell'art. 3 Cost., «*rimuovere gli ostacoli di ordine economico e sociale*, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese» (enfasi aggiunta). Si tratterebbe, è stato autorevolmente osservato, di un esempio di «*programma* in ragione del quale si deroga al *precetto*», accettando di porre (occasionalmente) tra parentesi l'«eguaglianza giuridico-formale», che pure rappresenta la regola, per raggiungere «una più vera parità di fatto» (così L. PALADIN, *Il principio costituzionale d'eguaglianza*, Giuffrè, Milano, 1965, p. 332). Ebbene, il “programma” di eguaglianza sostanziale di cui all'art. 3, co. 2, Cost., nell'imporre la rimozione degli ostacoli di ordine socio-economico frapposti al «pieno sviluppo della persona umana» e alla «partecipazione di tutti i lavoratori» alla vita del Paese, a un tempo legittimerebbe e prescriverebbe, tra l'altro, l'adozione di misure *promozionali* rivolte ai componenti di gruppi sociali minoritari o marginalizzati, tanto sul piano economico che socio-culturale, in vista del raggiungimento – in (apparente) deroga all'eguaglianza formale – di una «pari dignità sociale» effettiva (quest'ultima, dunque, fungerebbe da “cerniera” tra primo e secondo comma dell'art. 3 Cost.).

<sup>63</sup> L'espressione è mutuata da L. PALADIN, voce *Eguaglianza* (*dir. cost.*), cit., p. 540, il quale si riferiva, a rigore, al «valore polemico» – piuttosto che «innovativo sul piano giuridico» – della «menzione della razza» da parte dell'art. 3, co. 1, Cost.; in senso analogo v., tra gli altri, M. MASSA, *Diseguaglianza e condizioni personali. Una polemica sull'eguaglianza*, in M. DELLA MORTE (a cura di), *La dis-eguaglianza nello Stato costituzionale*, Editoriale Scientifica, Napoli, 2016, p. 166, per il quale «l'eguaglianza non è un principio neutrale: al contrario, è un principio polemico, frutto e matrice di scelte».

<sup>64</sup> «[L]e azioni positive», d'altronde, come evidenzia A. D'ALOIA, *Eguaglianza. Paradigmi e adattamenti di un principio 'sconfinato'*, in *Rivista AIC*, 4/2021, p. 47, «ci mostrano che l'idea di eguaglianza corrisponde ad un *continuum* fatto di stadi e significati differenti ma non slegati gli uni dagli altri», e che «eguaglianza formale» ed «eguaglianza sostanziale» non sono affatto «due termini in opposizione inconciliabile tra di loro».

Un algoritmo di stratificazione che, come proposto prima, sia addestrato a “sovrastimare” il bisogno di salute dei componenti di sotto-gruppi storicamente marginalizzati (e, dunque, più o meno direttamente disincentivati a rivolgersi al Servizio sanitario), non dev’essere inteso come uno strumento eccezionale, che – in deroga al principio di eguaglianza nella presa in carico – indebitamente sottragga a qualcuno per dare ad altri. Andrebbe valorizzata, piuttosto, la sua funzione (re-)distributiva; esso non discriminerebbe “alla rovescia”, ma distorcerebbe l’attuale assetto di poteri e interessi *al fine di* renderlo più equo. Esso, soprattutto, non si porrebbe nettamente in contrasto con le logiche e le dinamiche dell’eguaglianza, anche formale, ma ne esalterebbe lo storico significato “di rottura”. Questa, ovviamente, non è l’unica lettura possibile dello sviluppo e dell’impiego di algoritmi “riparativi”. Che li si intenda – similmente a quanto accade con le azioni positive – come trasposizioni sul piano pratico del valore “polemico” dell’eguaglianza tutta, anche formale, o come strumenti di eguaglianza sostanziale, ovvero, perfino, come “discriminazioni alla rovescia” o comunque misure di “diritto diseguale”, rimangono indiscutibili, deve ritenersi, il loro fine-ultimo equitativo e la loro compatibilità a Costituzione.