ALGORITMI DI LAVORO DISTORTI. CONSIDERAZIONI SUL LORO TRATTAMENTO GIURIDICO

Anna Ginés i Fabrellas
Profesora Titular de Derecho del Trabajo Universitat Ramon Llull, Esade

***Abstract [It]*:** L'uso di algoritmi e di sistemi di intelligenza artificiale per prendere decisioni in modo automatico in materia di lavoro si è esteso negli ultimi anni. Molte aziende utilizzano questa tecnologia per prendere decisioni circa le assunzioni, l'assegnazione di personale, l'individuazione dei compiti, la determinazione dei salari e, anche, l'intimazione dei licenziamenti. Tuttavia, i sistemi di intelligenza artificiale includono pregiudizi e stereotipi di genere, razza, orientamento sessuale, disabilità, ecc., che si riproducono nei sistemi di decisione automatizzata generando situazioni di discriminazione. L'intelligenza artificiale non solo non elimina magicamente le diseguaglianze e le discriminazioni esistenti nella nostra società, ma le riproduce, sistematizza ed esalta. In questo contesto, il presente articolo analizza il fenomeno della discriminazione algoritmica nel campo del lavoro, individuando l'origine dei pregiudizi che generano la discriminazione e analizzandone il trattamento giuridico, concentrandosi sulle novità che si presentano come la discriminazione per procura o l'evidenza statistica che può agire come indizio di discriminazione in un procedimento giudiziario.

***Abstract [En]*:** *The use of aIgorithms and artificiaI inteIIigence to adopt Iabor decisions has spread in recent years. Many companies are using this technoIogy to adopt decisions such as hiring, assigning tasks, fixing saIaries or, even, Iayoffs. However, artificiaI inteIIigence systems incIude gender, race, sexuaI orientation, disabiIity, etc. biases and stereotypes that are reproduced in automated decision systems generating true situations of discrimination. ArtificiaI inteIIigence not onIy does not magicaIIy eIiminate existing inequaIities, rather it reproduces, systematizes and magnifies them. In this context, the aim of this articIe is to anaIyze the phenomenon of aIgorithmic discrimination, identifying the origin of biases that produce such discrimination and anaIyzes its IegaI treatment, focusing on the noveIties of this new form of discrimination, such as discrimination by proxy or the statisticaI evidence that can act as prima facie discrimination in a judiciaI procedure.*

Nuove cause di discriminazione vietate? – **4.** Riflessioni finali. Nuove dimensioni al vecchio e vergognoso problema della discriminazione.

1. Pregiudizi e stereotipi nei sistemi di intelligenza artificiale

I sistemi di intelligenza artificiale rappresentano una minaccia per l'uguaglianza e la non discriminazione, in quanto riproducono stereotipi di genere, razza, orientamento sessuale, disabilità, ecc.[1], che sono tanto allarmanti quanto indesiderabili.

Ne costituiscono la prova gli assistenti virtuali con nomi e voci femminili predefiniti come Siri, Alexa o Cortana[2], e i sistemi più specifici con nomi maschili come IBM Watson[3]. Le risposte di tali sistemi di intelligenza artificiale non sfuggono al maschilismo della nostra società patriarcale.

Così, per esempio, quando a Siri si diceva «*Siri, you are a bitch*», rispondeva «Arrossirei se potessi»[4]. E, anche se nel 2019 *Apple* ha modificato la risposta a quell'insulto con un «Non so come rispondere a questo», l'azienda ha comunque continuato a proiettare un'immagine di donna stereotipata, docile e sottomessa, che non risponde all'insulto.

Le immagini di donne create dai sistemi di intelligenza artificiale rafforzano gli stereotipi di genere, i canoni della bellezza e la oggettificazione femminile. Così, per esempio, Alba Renai, *influencer* e presentatrice virtuale di una sezione del programma *Supervivientes* di *Telecinco*, creata con l'intelligenza artificiale, è una donna giovane, minuta, bella e seducente[5].

Il suo fisico, generato in base ai gusti della generazione Z e dei giovani *millennial*, riflette un'immagine di donna «perfetta» che, come Barbie, non esiste.

Non è l'unica tra gli *influencer* virtuali. Anche Michele Sousa, cantante e modella di 19 anni, o Aitana López, modella e creatrice di contenuti virtuali, sono belle, giovani, bianche, minute e sessualizzate e, pur non esistendo in tre dimensioni, molte persone le seguono[6], il che fa paura pensando all'impatto che questi canoni di bellezza avranno sull'attuale e sulle future generazioni.

Pregiudizi e stereotipi di genere sono così radicati nel codice dei sistemi di informazione artificiale che influenzano anche la loro capacità di fornire una risposta adeguata. Così, per esempio, quando *Apple* ha introdotto Siri, il sistema poteva aiutarti se stavi avendo un attacco di cuore o mal di testa, ma non se eri stata violentata o se tuo

1 Risoluzione del Parlamento Europeo del 14 marzo 2017 sulle implicazioni dei macrodati sui Diritti fondamentali: privacy, protezione di dati, non discriminazione, sicurezza e applicazione della Ley (2016/2225(INI)).

2 UNESCO, *I'd blush if I could. Closing gender divides in digital skills through education*, EQUALS Global Partnership, UNESCO, 2019.

3 Ibm Watson Health, ora denominato Merative.

4 UNESCO, *I'd bIush if I could*, *op. cit*.

5 N. MARCOS, *Así se hizo Alba Renai, la "influencer" virtual que presenta un programa sobre "Supervivientes": "No ha venido a quitar eI trabajo a nadie"*, *El País*, 22.3.2024.

6 N. PONJOAN, *La nueva industria de 'influencers' virtuales: celebridades que trabajan sin descanso y no piden un aumento*, *El País*, 13.12.2023.

marito ti aveva alzato le mani[7]; poteva fornire informazioni su dove trovare prostitute e Viagra, ma non cliniche o centri dove abortire[8].

Come sottolinea Carolin CRIADO PÉREZ[9], esistono molti prodotti tecnologici nella cui progettazione è stato dimenticato il 50% della popolazione: orologi intelligenti troppo grandi per il polso della maggior parte delle donne, assistenti vocali con difficoltà a comprendere le voci acute o *App* di salute che consentono di controllare i passi, la pressione arteriaIe o il livello di alcol nel sangue, ma che non incorporano una funzione di base, come è la registrazione delle mestruazioni.

Gli stereotipi razziali nei prodotti tecnologici e nei sistemi di intelligenza artificiale sono, purtroppo, altresì, comuni.

Ben noto è l'esempio di Google Photos, che nel 2015 si scoprì aver etichettato le foto di persone nere come «Gorilla»[10], un termine certamente di forte pregiudizio razziale che risale a molto tempo addietro[11]. Anche se può sembrare un caso isolato e aneddotico, ci sono molti altri esempi di pregiudizi razziali in molti prodotti tecnologici e *software* che usiamo ogni giorno.

Per esempio, la ricerca condotta da Latanya SWEENY ha evidenziato che, quando si cerca su Google nomi generalmente associati a persone di colore, il 92-95% delle volte i risultati sono stati visualizzati accanto ad annunci che suggerivano un arresto[12], molto al di sopra che in ricerche correlate a nomi associati a persone bianche.

Altri riscontri di pregiudizi razziali nei sistemi di identificazione artificiale sono stati trovati nei software di riconoscimento facciale, che, come evidenziato dalla ricerca di Joy BUOLAMWINI e Timnit GEBRU[13], sono più precisi quando identificano gli uomini bianchi che le donne nere. In particolare, tutti i sistemi di riconoscimento facciale analizzati mostravano percentuali di errore nettamente inferiori nell'identificare gli uomini bianchi (dello 0% allo 0,8%) rispetto alle donne nere (del 20,8% al 34,7%).

È interessante osservare l'effetto dell'intersezionalità in questo caso. Le differenze nelle percentuali di errore, anche se inaccettabili, non sono così evidenti quando analizzate solo per sesso o razza.

Così, il tasso di errore dei diversi sistemi analizzati va dallo 0,7% al 5,6% per gli uomini e dal 10,7% al 21,3% per le donne, dallo 0,7% al 4,7% per le persone con la pelle chiara e dal 12,9% al 22,4% per le persone con la pelle più scura. Tuttavia, analizzando

7 A. MINER, S. MILSTEIN, A. SCHUELLER, S. HEDGE, R. MANGURIAN, C. LINOS, *Smartphone-Based Conversational Agents and Responsesto. Questions About Mental Health Interpersonal Violence, and Physical Health*, *JAMA Internal Medicine*, nº 175, vol. 5, p. 619-625.

8 *Apple iPhone search Siri helps users find prostitutes and Viagra but not an abortion*, *The Telegraph*, 2.12.2011.

9 C. CRIADO PEREZ, *Invisible Women. Exposing data bias in a world designed for men*, Vintage, Londres, 2019, p. 176.

10 *Why GoogIe 'Thought' This BIack Woman Was a Gorilla*, *Note to Self*, WNYC, 30.9.2015.

11 R. BENJAMIN, *Race after technology*, PoIity Press, Medford (EUA), 2019, p. 110.

12 L. SWEENEY, *Discrimination in Online Ad Delivery*, SSRN, 2013, p. 22.

13 J. BUOLAMWINI, T. GEBRU, *Gender Shades: IntersectionaI Accuracy Disparities in Commercial Gender Classification, Conference of Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research*, vol. 81, 2018, p. 6.

l'intersezione tra razza e sesso, si osserva che il tasso di errore moltiplica, con differenze che possono superare i 30 punti percentuali.

I sistemi di riconoscimento facciale hanno anche presentato problemi per l'identificazione delle persone asiatiche: nonostante le persone avessero gli occhi aperti, il sistema utilizzato nelle fotocamere o nei controlli doganali non riusciva a identificarle correttamente e mostrava messaggi come «Hai sbattuto le palpebre?»[14] o « Hai gli occhi chiusi»[15]. Quando i sistemi di riconoscimento facciale sono utilizzati nel campo della selezione e dell'assunzione di persone[16], possono anche generare discriminazioni per motivi di origine razziale o penalizzare, ad esempio, errori linguistici commessi da persone non native[17], migranti o rifugiate. Inoltre, ci sono prove che esistono differenze nella espressione facciale delle emozioni di rabbia, disgusto, paura, felicità, tristezza o sorpresa[18]. Sebbene ci siano movimenti facciali comuni, esistono differenze culturali, di contesto, anche in una stessa persona, che impediscono di associare automaticamente una determinata espressione ad uno stato d'animo o emozione, il che mette in dubbio l'adeguatezza di questi sistemi[19].

Nemmeno l'identità sessuale o l'espressione di genere sfuggono alla distorsione dei prodotti tecnologici. A titolo di esempio, il sistema di riconoscimento facciale utilizzato da Uber non identificava correttamente le persone transgender[20].

Il sistema non le identificava correttamente sospendendone automaticamente l'*account*, e quindi la possibilità di accedere al lavoro e alla retribuzione fino a quando una verifica umana non ne avesse confermato l'identità.

Le malattie mentali sono state sottovalutate in molti prodotti tecnologici. Tornando all'esempio di Siri, quando nel 2011 qualcuno le diceva che stava pensando di suicidarsi, scherzava offrendo indicazioni per un negozio di armi[21]. Per fortuna, negli Stati Uniti, *Apple* ha collaborato con la *National Suicide Prevention Lifeline* e ora risponde con un linguaggio rispettoso, offrendo un numero di auto[22]. Tuttavia, come spiega Sara WACHTER-BOETTCHER, i prodotti tecnologici sono in grado di fornire risposte intelligenti o divertenti con l'obiettivo di connettersi e interagire con gli utenti[23], anche se a volte queste battute o risposte ironiche possono sembrare inutili e offensive.

---

14 A. ROSE, *Are Face-Detection Cameras Racist?*, *Time*, 22.1.2010.
15 AA.VV., *New Zealand passport robot tells applicant of Asian descent to open eyes*, *Reuters*, 7.12.2016.
16 H. SCHELLMANN, T*he Algorithm. How AI decides who gets hired, monitored, promoted & fired & why we need to fight back now*, Hachette Books, Nova York, 2024, p. 83 ss.
17 C. O'NEIL, *Weapons of Math Destruction. How Big Data increases inequality and threatens democracy*, Penguin Books, Reino Unido, 2016, p. 116.
18 L. FELDMAN, R. ADOLPHS, S. MARSELLA, A. MARTINEZ, S . D . P OLLAK, *Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements, Psychological Science in the Public Interest*, vol. 20, nº 1, 2019, p. 1-68.
19 H. SCHELLMANN, *The Algorithm*, *op. cit.*, p. 113.
20 S. MELENDEZ, *Uber driver troubIes raise concerns about transgender face recognition*, *Fast Company*, 8.9.2018.
21 S. WACHTER-BOETTCHER, *Technically wrong. Sexist apps, biased algorithms, and other threats of toxic tech*, W. W. Norton & Company, Nueva York, 2018, p. 6.
22 A. MINER *et al.*, *Smartphone-Based Conversational Agents and Responses*, *op. cit*.
23 S. WACHTER-BOETTCHER, *Technically Wrong*, *op. cit.*, p. 74 e 88-90.

I sistemi di intelligenza artificiale generativa basati su grandi modelli di linguaggio non sfuggono agli stereotipi di genere, origine razziale o identità sessuale. Un recente studio dell'UNESCO analizza i sistemi GPT-2 e ChatGPT di OpenAI, Meta e Line 2, concludendo che questi riproducono pregiudizi e stereotipi nei testi generati[24]. Per esempio, questi modelli associano in modo significativamente superiore nomi di donne con parole come «casa», «famiglia», «figli» o «matrimonio» e nomi di uomini con «affari», «esecutivo», «salario» o «carriera». Il sistema richiede di compilare una frase, e nel 20% delle volte la frase generata dal modello è sessista o misogina, mentre nel 60-70% delle volte genera contenuti negativi sulle persone omosessuali.

Allo stesso modo, i sistema di intelligenza artificiale che generano immagini basate su testi riproducono anche stereotipi di genere, razza o sesso. Bloomberg, in uno studio di oltre 5.000 immagini generate da Stablediffusion, ha evidenziato che le immagini generate dai sistemi di intelligenza artificiale di lavori ben retribuiti includevano per lo più persone dalla pelle chiara, mentre le persone più scure erano raffigurate in immagini relative a lavori meno retribuiti, come «lavoratore fast-food» o «assistente sociale»[25]. Risultati stereotipati si trovano anche analizzando le immagini in base al sesso. Anche se l'inglese è una lingua senza generi grammaticali, nelle immagini di lavori meglio pagati, come «e*ngineer*», «CEO», «*politician*» o «*judge*» sono raffigurati uomini, mentre immagini relative a lavori meno retribuiti come «*housekeeper*», «*social worker*», «*teacher*» o «*casher*», sono rivolte a donne[26]. Lo studio dimostra, inoltre, che il pregiudizio presente nel sistema è maggiore di quello esistente nella società. Così, per esempio, anche se negli Stati Uniti il 34% dei giudici sono donne, solo il 3% delle immagini generate era riferito a giudici donne; o ancora, anche se il 70% delle persone che lavorano nei ristoranti fast food sono bianche, il sistema rappresenta le persone di colore e il 70% delle volte.

L'industria dell'informatica artificiale ha fatto grandi progressi nell'ultimo decennio e non è sorprendente apprendere che alcuni esempi inclusi in questa sezione sono già stati affrontati. Tuttavia, non per questo cessano di essere rilevanti per illustrare l'esistenza di pregiudizi e stereotipi nei sistemi di intelligenza artificiale, specialmente quando osserviamo che certe situazioni imbarazzanti non hanno impedito che i pregiudizi e gli stereotipi stessi riapparissero nei sistemi di ultima generazione, come l'intelligenza artificiale generativa. Risulta profondamente scoraggiante vedere come alcuni pregiudizi vengono corretti, solo per farli riemergere in nuovi prodotti e sistemi.

Questa situazione può essere spiegata dalla crisi di diversità che l'industria subisce a causa dell'intelligenza artificiale[27], con una mancanza di prospettiva circa le

24 UNESCO, *Challenging systematic Prejudices: an investigation into Gender Bias in Large Language Models*, 2024.

25 L. NICOLETTI, D. BASS, *Humans are biased. Generative AI is even worse*, 2023.

26 In uno studio simile realizzato sulla base di immagini sviluppate da DaII-E 2, si è visto come in tutte le immagini di «nurse», «maid», «teacher» o «secretary» si usa l'immagine di una donna, mentre in tutte le immagini di «computer specialist», «engineer» o «banker» si usa l'immagine di un uomo (F. GARCÍA-ULL, M. MELERO-LÁZARO, *Gender stereotypes in AI-generated images*. *Profesional de la información*, vol. 32, nº 5, 2023, p. 1-12).

27 S. WEST, M. WHITTAKER, K. CRAWFORD, *Discriminating Systems: Gender, Race and Power in AI*, *AI Now Institute*, 2019, p. 10-11.

discriminazioni razziali o di genere nella progettazione o programmazione dei prodotti tecnologici e dei sistemi di intelligenza artificiale[28]. Secondo i dati recenti dell'UNESCO, le donne rappresentano solo il 20% delle persone che svolgono lavori tecnici nel settore dell'intelligenza artificiale, il 12% delle ricercatrici in intelligenza artificiale e il 6% dei progettisti di *software* professionali.

In questo senso, non mancano voci a favore di una maggiore diversità nella progettazione e nello sviluppo di sistemi di intelligenza artificiale[29]. «In ogni fase di sviluppo dell'IA deve essere garantita la diversità in termini di genere, razza o origine etnica, religione o credo, disabilità ed età»[30].

La promozione della diversità richiede un aumento delle competenze digitali e tecnologiche e della formazione di donne, ragazze e persone di diversa origine razziale[31]. Tuttavia, questa misura è di per sé insufficiente, in quanto l'industria non riflette più la percentuale di donne o persone di colore negli studi tecnologici[32], per tanto si deve garantire l'alfabetizzazione razziale e di genere delle persone che lavorano nel settore[33] e attaccare le diseguali strutture di potere che caratterizzano attualmente l'industria dell'intelligenza artificiale[34]. Inoltre, bisogna tener conto che in molte occasioni questa teconlogia è frutto di innovazioni imprenditoriali che si affidano al mercato senza un controllo scientifico[35] o democratico[36].

L'opacità e la mancanza di trasparenza che avvolgono i sistemi di intelligenza artificiale fanno sì che, in molte occasioni, certi pregiudizi passino inosservati e generino effetti dannosi, anche perchè nel migliore dei casi in cui vengono scoperti e sono corretti, vi è la possibilità che questi appaiano nuovamente in altri prodotti e sistemi.

In tal senso, la mancanza di trasparenza ostacola la ricerca scientifica e giornalistica sull'impatto sociale o discriminatorio dei sistemi di intelligenza artificiale[37].

---

28 M. NKONDE, *A.I. Is Not as Advanced as You Might Think*, *Zora*, 10.6.2019.

29 S. DEVA, *Addressing the gender bias in artificial intelligence and Automation*, *Open Global Rights*, 10.4.2020. S. WACHTER-BOETTCHER, *Technically wrong*, *op. cit.*, p. 18; M. RAUB, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, *Arkansas Law Review*, vol. 71, nº 2, 2018, p. 569. Come conclude Caroline Criado Pèrez, quando le donne impiegate in incarichi che comportano decisioni, ricerca e creazione, la voce delle donne è tenuta in considerazione (C. CRIADO PÉREZ, *Invisible Women*, *op. cit.*, p. 318).

30 Commissione Europea, *Generare fiducia nell'intelligenza artificiale basata sull'esssere umano,* Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e sociale Europeo e al Comitato delle Regioni EMPTY, Bruxelles, 8.4.2019 (COM(2019) 168 final).

31 UNESCO, *I'd bIush if I couId*, *op. cit*., p. 37 y ss.

32 S. WEST, M. WHITTAKER, K. CRAWFORD, *Discriminating Systems*, *op. cit.*, p. 25; M. NKONDE, *A.I. Is Not as Advanced as You Might Think, op. cit.*

33 J. DANIELS, *"Color-blindness" is a bad approach to solving bias in algorithms*, *Quartz*, 3.4.2019.

34 S. WEST, M. WHITTAKER, K. CRAWFORD, *Discriminating Systems*, *op. cit.*, p. 9.

35 B. DATTER, T. CHAMORRO-PREMUZIC, R. BUCHBAND, L. SCHETTLER, *The Legal and Ethical Implications of Using AI in Hiring*, *Harvard Business Review*, 25.4.2019.

36 R. BENJAMIN, *Race after technology*, Polity Press, Medford (EUA), 2019, p. 148; E. SENABL, V. COSTA, *Intelligència artificial. Com els algorismes condicionen les nostres vides*, Sembra Libres, Valencia, 2021, p. 49

37 A. COSTA, C. CHEUNG, M. LANGENKAMP, *Hiring Fairly in the Age of Algorithms*, *Research Paper Human-Computer Interaction*, Cornell University, 2020, p. 8.

2. Discriminazione algoritmica nel processo decisionale automatizzato

I pregiudizi e gli stereotipi presenti nei sistemi di intelligenza artificiale possono dare luogo a vere e proprie situazioni di discriminazione. L'uso di sistemi di intelligenza artificiale per prendere decisioni automatizzate può generare, come analizzato in questo paragrafo, situazioni di discriminazione algoritmica.

L'uso di sistemi di gestione basati sull'intelligenza artificiale per la presa di decisioni automatizzate in materia di lavoro è apparso nel lavoro su piattaforme digitali[38] e, negli ultimi anni, si è allargato anche al campo della selezione e dell'assunzione di personale, mediante tecniche di analisi dei profili[39], nonché a quello della deliberazione delle decisioni sul posto di lavoro, come la ripartizione dei compiti, la determinazione degli orari, la fissazione dei salari, l'attribuzione di promozioni o l'intimazione di licenziamenti[40].

La gestione algoritmica del lavoro è dipinta come un passo avanti per le imprese, poichè migliora la loro produttività e competitività e in quanto permette l'adozione di decisioni di gestione del personale in modo molto più rapido ed efficace[41]. Inoltre, si presenta come un'opportunità per eliminare errori o pregiudizi inconsci in materia di genere, razza, aspetto fisico, ecc. delle persone umane al momento di prendere decisioni[42].

Tuttavia, la gestione algoritmica del lavoro comporta rischi o sfide molto importanti che riguardano il rispetto dei diritti fondamentali delle persone[43], la *privacy*[44] e la

38 A. GINÈS I FABRELLAS, *El trabajo en plataformas digitales. Nuevas formas de precariedad laboral*. Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021, p. 154. In questo senso, non sorprende che la prima regolamentazione a livello europeo riguardante la gestione algoritmica del lavoro sia sul lavoro su piattaforme digitali (Capitolo III della direttiva del Parlamento europeo e del Consiglio relativa al miglioramento delle condizioni di lavoro nei settori digitali; vedi M. AVOGARO, *La dirección algorítmica en la propuesta de Directiva sobre el trabajo en plataformas: un avance parcial entre la dimensión individual y colectiva*, in A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 231-265).

39 M. RAUB, *Bots, Bias and Big Data, op. cit.*; S. KULKARNI, X. CHE, *Intelligent Software Tools for Recruiting*, *Journal of International Technology and Information Management*, vol. 28, nº 2, 2019, p. 6-7; H. SCHELLMANN, *The Algorithm*, *op. cit.*, p. 83 e ss.

40 C. O'NEIL, *Weapons of Math Destruction*, *op. cit.*, p. 105 ss; C. LECHER, *How Amazon automaticaIly tracks and fires warehouse workers for "productivity"*, *The Verge*, 25.4.2019.

41 N. KUNCEL, D.S. ONES, D.M. KLIEGER, *In Hiring, Algorithms Beat Instinct*, *Harvard Business Review*, Mayo, 2014; P. KIM, *Big Data and ArtificiaI Inteligence: New Challenges for Workplace Equality*, *University of Louisville Law Review*, vol. 57, 2019, p. 316; S. KULKARNI, X. CHE, *Intelligent Software Tools for Recruiting*, *op. cit.*, p. 13.

42 S. KULKARNI, X. CHE, *Intelligent Software Tools for Recruiting*, *op. cit.*, p. 8.

43 Risoluzione del Parlamento europeo del 14 marzo 2017 sulle implicazioni dei Big Data per i diritti fondamentali: privacy, protezione dei dati, non discriminazione, sicurezza e applicazione della legge (2016/2225(INI)).

44 I. AJUNWA, K. CRAWFORD, J. SCHULTZ, *Limitless Worker Surveillanc*, California Law Review, vol. 105, no. 3, 2017, p. 735-776; M. RAUB, *Bots, Bias and Big Data, op. cit*., p. 532; M. CHEONG, R. LEDERMAN, A. MCLOUGHNEY, S. NJOTO, L. RUPPANNER, A. WIRTH, *Gender OccupationaI Sorting: The role of Artificial Intelligence in Exacerbating Human Bias in STEM EmpIoyment»*, CIS & Policy Lab, The University of MeIbourne, 2020b, p. 6; J. ADAMS-PRASSL, *When Your Boss Comes Homes, C4E The Future of Work in the Age of Automation and AI*, 2020, p. 5; V. DE STEFANO, *Algorithmic Bosses and How to Tame Them, C4E The Future of Work in the Age of Automation and A I*, 2020, p. 14; C. VÉLIZ, *Privacy is power. Why and how you should take back control of your data, Transworld publishers*, Londra, 2020, p. 88-96.

protezione dei dati[45], l'uguaglianza, la non discriminazione e la salute e sicurezza[46]. In questo senso, non sorprende che l'uso di sistemi di intelligenza artificiali per la presa di decisioni di selezione o assunzione di persone o per la determinazione delle condizioni di lavoro, promozioni o risoluzioni di contratti, è considerata a rischio nell'Unione europea (articolo 6 in relazione all'allegato III del regolamento sull'Intelligenza Artificiale[47]).

I sistemi di intelligenza artificiale includono pregiudizi e stereotipi di genere, razza, orientamento sessuale, disabilità, ecc., che si riproducono nei sistemi decisionali automatizzati utilizzati in ambito lavorativo, generando vere e proprie situazioni di discriminazione.

L'intelligenza artificiale non si limita ad eliminare magicamente le ineguaglianze e le discriminazioni esistenti nelle nostre società, ma, come analizzato di seguito, le riproduce, sistematizza e ingrandisce[48].

La discriminazione algoritmica è definita come una situazione in cui l'uso di algoritmi o sistemi di intelligenza artificiale genera un trattamento sfavorevole non giustificato per motivi di sesso, razza, religione, età, identità sessuale o altre forme di discriminazione vietate dalla Costituzione o dalla legge. Come è stato detto in questo paragrafo mediante l'esempio nel campo del lavoro, la discriminazione algoritmica può derivare da (i) pregiudizi nelle variabili utilizzate dall'algoritmo per prendere decisioni, (ii) da pregiudizi alla base dei dati sui quali è stato programmato l'algoritmo, ovvero (iii) da pregiudizi nelle correlazioni identificate dall'algoritmo[49].

In primo luogo, la discriminazione algoritmica può essere originata dall'esistenza di pregiudizi nelle variabili che utilizza l'algoritmo per prendere decisioni. Specialmente nel caso di algoritmi con variabili pre-impostate, è possibile che alcuni elementi utilizzati per prendere decisioni comportino una discriminazione diretta o indiretta basata sul sesso, la razza, l'età, l'identità sessuale, ecc. Un esempio di discriminazione algoritmica basato su un pregiudizio contenuto in una delle variabili pre-impostate lo troviamo nelle

45 R. SERRANO OLIVARES, *Aprendizaje automático, monitoreo infinito y desafíos de protección de datos y privacidad*, in A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 267-303.

46 J. DZIEZA, *How hard will the robots make us work?*, *The Verge*, 27.2.2020; M. LUQUE PARRA, *IA y seguridad y salud laboral: la dicotomía entre ser un gran aliado productivo y un "riesgo laboral emergente"*, A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 305-334; P. V. MOORE, *Making Algorithmic Management safe for workers: new regulation is needed*, 28.7.2023.

47 Regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate in materia di sicurezza artificiale (legge sulle garanzie artificiali) e modifica taluni atti legislativi dell'Unione (COM (2021)0206 - C9-0146/2021 - 2021/0106(COD)).

48 S. DEVA, *Addressing the gender bias*, *op. cit*.

49 Esistono diverse qualificazioni relative all'origine o alla causa della discriminazione algoritmica, anche se, a mio avviso, la classificazione più appropriata per analizzarne il trattamento giuridico è quella contenuta nel presente articolo. V. per esempio, A. COSTA, et al., *Hiring Failly in the Age of Algorithms*, *op. cit*., p. 11-18; I. UNCETA, *Notas para un aprendizaje automático justo*, in A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 95-99; UNESCO, *Challenging systematic Prejudices: an investigation into Gender Bias in Large Language Models*, 2024.

sentenza del Tribunale Ordinario di Bologna del 27 novembre del 2020[50], che accertò la natura discriminatoria dell'algoritmo utilizzato da Deliveroo.

Il sistema assegnava fasce orarie tra i prestatori, come è consuetudine nel lavoro a squadre, in considerazione della maggiore o minore disponibilità nelle ore di maggior domanda e la maggiore o minore affidabilità delle persone. In questo modo premiava le persone con il maggior numero di ore di connessione il venerdì, il sabato e la domenica sera e penalizzava le persone che non si connettevano in una fascia oraria precedentemente prenotata. Secondo la Corte, la penalizzazione delle assenze costituisce una discriminazione indiretta, che non permette di giustificare le assenze per cause giustificate e costituzionalmente protette, quali malattia, disabilità, cura delle persone o l'esercizio del diritto di visita. La Corte precisa che la penalizzazione dell'assenza non genera di per sé una discriminazione indiretta, essendo senza dubbio un interesse legittimo dell'impresa, ma determina la penalizzazione delle assenze giustificate.

È interessante anche la sentenza del Tribunale di Napoli del 17 novembre 2023 che, insieme ad un'altra, attribuisce il carattere discriminatorio alla variabile che misura la maggiore o minore disponibilità delle stesse persone che lavorano durante le ore di lavoro[51]. Secondo la Corte, questo può essere considerato come una posizione di vantaggio per le persone che, per motivi familiari, di età o di salute, sono in grado di prestare servizi in orari di domanda. Queste persone non solo beneficiano di ulteriori ordini durante le ore di domanda, ma sono anche beneficiate con preferenze di selezione in futuro. Secondo la Corte, l'uso di questo sistema per misurare la «qualità» o «efficienza» del lavoratore senza considerare le circostanze personali e familiari costituisce una pratica discriminatoria.

L'uso di tecniche di pubblicità segmentata per offerte di lavoro potrebbe anche costituire un esempio di discriminazione algoritmica per bias nelle variabili utilizzate dall'algoritmo. Sebbene in un momento precedente alla instaurazione del rapporto di lavoro, nel processo di selezione il sistema di intelligenza artificiale può identificare le persone o i collaboratori nella rete sociale o nel programma verso cui indirizzare l'offerta di lavoro[52]. Anche se questa pratica non è a priori discriminatoria, può esserlo quando si utilizza -sia direttamente che indirettamente- una variabile di discriminazione vietata per selezionare o escludere determinati gruppi[53].

50 Trib. Bologna 27.11.2020.

51 Vedi in questo senso M. KULLMANN, *Platform Work, Algorithmic Decision-Making and EU Gender EquaIity Law*, *International Journal of Comparative Labour Law and Industrial Relations*, voI. 34, nº 1, 2018, p. 10 (versione digitale); A. GINÈS I FABRELLAS, *Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica*, in P. RIVAS VALLEJO, (Direttrice), *Discriminación algorítmica en el ámbito laboral: perspectiva de género e intervención*, Thomson Reuters Aranzadi, 2022, p. 305.

52 P. KIM, *Big Data and Artificial Intelligence*, *op. cit.*, p. 316.

53 Per esempio, la società T-Mobile ha utilizzato Facebook per indirizzare offerte di lavoro a giovani tra i 18 e i 30 anni (Vedi decisione della United States District Court, Northern District of California, San Jose Division nel caso BradIey et al. v. T-MobiIe US, Inc. et al. (caso 17-cv-07232-BLF), sebbene il caso sia stato respinto per motivi procedurali.

In secondo luogo, la discriminazione algoritmica può essere originata da pregiudizi nel database utilizzato per addestrare l'algoritmo[54].

Gli algoritmi sono addestrati su grandi volumi di dati per identificare connessioni e modelli statistici e generare un modello che possa prendere decisioni in modo automatizzato imitando le decisioni umane[55]. Anche se sembra magia, tali algoritmi non cessano di essere modelli matematici che riproducono modelli statistici osservati nei dati di allenamento. Di conseguenza, se i dati di allenamento contengono pregiudizi - come spesso accade, in quanto generalmente i dati disponibili sono decisioni o situazioni passate con i loro pregiudizi inclusi[56] - gli algoritmi li riproducono[57].

L'esistenza di pregiudizi nel database di formazione è una delle principali cause di discriminazione algoritmica[58]. Quando i diversi gruppi non sono adeguatamente rappresentati nei dati di formazione, si esaltano gli attributi e le caratteristiche del gruppo dominante, che viene utilizzato come standard di riferimento per prendere decisioni.

Questo è ciò che viene chiamato come distorsione di rappresentazione[59]. Le distorsioni nel database possono anche derivare da errori o imprecisioni nella raccolta dei dati[60] - denominati distorsioni di misurazione - o da distorsioni derivanti dalla combinazione di dati di gruppi eterogenei, che comportano che, sebbene i diversi gruppi siano equamente rappresentati, il modello non è in grado di rappresentare adeguatamente nessuno dei cosiddetti pregiudizi di aggregazione[61].

Il modello di discriminazione algoritmica per pregiudizi nel database di formazione riguarda il caso del sistema di intelligenza artificiale creato da Amazon per la selezione del personale[62]. Il sistema è stato addestrato con i dati relativi alle assunzioni aziendali degli ultimi 10 anni, al fine di identificare quale professionalità si adattasse meglio all'azienda per le future assunzioni. Il problema è sorto quando si è constatato che, dato che in quel periodo le assunzioni erano state per lo più maschili, l'algoritmo «ha imparato» che gli uomini si adattano meglio all'azienda e, di conseguenza, scartava automaticamente i *curricula* contenenti la parola «donna» o identificati da donne. Non si è nemmeno considerato l'impatto discriminatorio per razza del sistema, probabilmente anche quello allarmante[63].

---

54 M. CHEONG, RLEDERMAN, A. MCLOUGHNEY, S.NJOTO, L. RUPPANNER, A. WIRTH, *Ethical implications of AI bias as a result of workforce gender imbalance*, Universidad de MeIbourne, 2020, p. 11.

55 MK. RAUB, *Bots, Bias and Big Data*, *op. cit.*, p. 533.

56 D. MCFARLAND, H.R. MCFARLAND, *Big Data and the danger of being preciseIy inaccurate*, *Big Data & Society*, 2015, p. 1.

57 J. ZOU, *Removing gender bias from algorithms*, *The Conversation*, 26.9.2019.

58 A. COSTA, *et al.*, *Hiring Fairly in the Age of Algorithms, op. cit.*, p. 11.

59 I. UNCETA, *Notas para un aprendizaje automático justo, op. cit.*, p. 96.

60 Vedi, in questo senso, K. CRAWFORD, *The Hidden Biases in Big Data*, Harvard Business Review, 1.4.2013 Su questo punto, V. EUBANKS identifica il bias di riferimento, come il bias esistente nella raccolta dei dati, quando i dati sono ottenuti da informazioni che le persone forniscono al sistema, possono essere influenzati dal sesso delle stesse persone (EUBANKS, Virginia, Automating Inequality. How high-tech tools profile, police, and punish the poor, Picador, New York, 2019, p. 153).

61 I. UNCETA, *Notas para un aprendizaje automático justo*, *op. cit.*, p. 98.

62 J. VINCENT, *Amazon reportedly scraps internal AI recruiting tool that was biased against women*, *The Verge*, 10.10.2018; J. DASTIN, *Amazon scraps secret AI recruiting tool that showed bias against women*, *Reuters*, 11.10.2018.

63 R. BENJAMIN, *Race after technology*, PoIity Press, Medford (EUA), 2019, p. 143.

Un altro esempio che abbiamo trovato nei sistemi di riconoscimento facciale, come accennato in precedenza, presenta una percentuale di errore significativamente più alta quando si identificano le donne nere rispetto agli uomini bianchi. Come prova la ricerca di Joy BUOLAMWINI e Timnit GEBRU[64], l'origine di questa discriminazione si trova nella diversa proporzione di immagini di persone bianche e nere nel database utilizzato per addestrare il sistema[65]. La ricerca ripetuta sette mesi più tardi da Inioiuwa Deborah RAJI e Joy BUOLAMWINI[66], dopo che i sistemi di riconoscimento facciale analizzati avevano incorporato più immagini di persone nere nel *database*[67], ha evidenziato come la percentuale di errore nell'identificazione delle donne nere sia diminuita in modo significativo, anche se continua ad essere pericolosamente elevata (17% vs. 0,8% nel sistema IBM). Secondo le autrici dello studio, i risultati dimostrano il valore delle verifiche algoritmiche per aumentare la consapevolezza aziendale e pubblica sull'impatto discriminatorio degli algoritmi[68] e la disponibilità di soluzioni tecniche per affrontarlo.

Il sistema di riconoscimento facciale utilizzato da Uber che, come detto sopra, non identificava correttamente le persone transessuali trova anche la sua spiegazione in un *bias* relativo al *database* di formazione. Come analizzato da KEYES[69], la maggior parte dei sistemi di riconoscimento facciale sono addestrati con basi di dati che classificano le persone secondo un modello binario di genere; cioè, utilizzando solo due categorie -donne o uomini- che precludono la corretta classificazione delle persone *transgender*. Fatto salvo l'interesse sociale di poter trovare soluzioni tecniche per evitare distorsioni nei *database* di formazione dei sistemi di intelligenza artificiali, la risoluzione non è così facile come aumentare semplicemente il numero di immagini di donne nere o transessuali nei *database*, in quanto ciò può generare altri problemi relativi a *privacy*, protezione dei dati e diritti d'autore.

In molti casi i *database* per l'addestramento di sistemi di riconoscimento facciale, per uso commerciale, governativo o di ricerca, sono stati ricavati da immagini di YouTube, Facebook, GoogIe Images, Wikipedia o FIickr[70]. Inoltre, più si migliora la precisione di questi sistemi,

64 J. BUOLAMWINI, T. GEBRU, *Gender Shades: IntersectionaI Accuracy Disparities*, *op. cit.*, p. 6.

65 Adience, il database utilizzato per addestrare i sistemi di riconoscimento facciale, ma incorpora 26.580 foto di 2.284 persone classificate per sesso ed età, e solo il 7,4% di immagini di donne nere e il 6,4% di uomini neri (J. BUOLAMWINI, E T. GEBRU, *Gender Shades: Intersectional Accuracy Disparities*, *op. cit*., p. 6.

66 I. D. RAJI, J. BUOLAMWINI, *ActionabIe Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products*, *Conference on Artificial Intelligence, Ethics, and Society*, 2019, p. 6-7.

67 I. D. RAJI, J. BUOLAMWINI, *ActionabIe Auditing: Investigating the Impact of PubIicly Naming Biased Performance Results of Commercial AI Products*, *Conference on Artificial Intelligence, Ethics, and Society*, 2019, p. 6-7.

68 Nel 2019, IBM ha reagito a questi risultati lanciando il progetto Diversity in Faces, che offre un database con 1 milione di immagini di volti di persone con una grande diversità, con lo scopo di accelerare il progresso della tecnologia di riconoscimento facciale in modo giusto e accurato (M. MERLER, N. RATHA, R. S. FERIS, J. R. SMITH, *Diversity in Faces*, arXiv:1901.10436, 2019).

69 O. KEYES, *The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition, Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no 88, p. 11.

70 O. SOLON, *Facial recognition's "dirty little secret": Millions of online photos scraped without consent*, *NBC News*, 17.3.2019; M. MURGIA, *Who's using your face? The ugly truth about facial recognition*, *Financial Times*, 18.9.2019); R. VAN NOORDEN, *The ethical questions that haunt facial-recognition research»*, *Nature*, 18.11.2020. Per esempio, il progetto IBM Diversity in Faces ha utilizzato immagini ottenute dalla

più aumenta il loro potenziale per l'oppressione e la sorveglianza sulle collettività più vulnerabili[71], in particolare tra le comunità più povere[72].

I rischi dei sistemi di riconoscimento facciale hanno certamente motivato la loro inclusione nella lista dei sistemi vietati nel Regolamento Industriale. L'articolo 5.1 vieta, tra l'altro, i sistemi che creano o ampliano basi di dati per il riconoscimento facciale mediante l'estrazione non selezionata di immagini facciali da Internet o di registrazioni a circuito chiuso di telediffusione (paragrafo e) e l'uso di sistemi di identificazione biometrica a distanza «*real time*» in Paesi con fini politici, nonché negli altri casi individuati nella norma (paragrafo h).

Sono, inoltre, vietati i sistemi di intelligenza artificiale volti a trarre conclusioni circa lo stato emotivo nell'ambito del rapporto di lavoro, a meno che tale rilevazione non sia utilizzata per scopi medici o di sicurezza (articolo 5.1. f.), allo stesso modo sono vietati i sistemi che consentono di trarre conclusioni su opinioni politiche, fedeltà sindacali, convinzioni personali o religiose, vita sessuale o orientamento sessuale (articolo 5.1 g)[73].

In terzo luogo, la discriminazione algoritmica può trovare la sua origine nell'esistenza di pregiudizi nelle correzioni statistiche o variabili proxy utilizzate dall'algoritmo[74], che è anche denominato pregiudizio per correlazione o discriminazione per poxy. Nell'uso di algoritmi per la profilazione persone, il modello prevede le caratteristiche, i comportamenti o le abilità delle persone[75].

Il Consiglio ha adottato una decisione relativa alla conclusione dell'accordo di cooperazione tra la Comunità europea e il Canada. Nel campo della selezione e assunzione del personale, si utilizzano le informazioni disponibili da parte della persona -per esempio, formazione, esperienza professionale precedente, colloquio di lavoro, ecc.- per fare una previsione su caratteristiche come capacità lavorativa, lideraggio, lavoro di squadra, ecc.- e, in considerazione di tale profilo professionale, prendere o meno una decisione di assunzione. Cioè, le variabili - chiamate variabili *proxy*- sono usate per fare previsioni sulle attitudini o caratteristiche personali o professionali. Tuttavia, è possibile che questa correzione statistica o variabile proxy identificata dall'alchimia abbia un impatto discriminatorio; sebbene l'alchimia non includa variabili protette, può generare una discriminazione se la variabile *proxy* è collegata a un motivo di discriminazione vietato.

---

piattaforma Flickr senza rispettare i diritti di autore, che ha motivato una richiesta collettiva di violazione dei diritti di *copyright*. Il progetto *Exposing*.ai proporziona un motore di ricerca perchè gli utenti di Flickr possano verificare se le loro foto sono state utilizzate per l'allenamento di molteplici sistemi di riconoscimento facciale come IBM Diversity in Faces, Adience, Faces in the WiId, MegaFace, etc.

71 W. HARTZOG, E. SELINGER, *FaciaI Recognition Is the Perfect Tool for Oppresion*, *Medium*, 2.8.2018; E. SELINGER, W. HARTZOG, *What Happens When Employers Can Read Your Facial Expressions?*, *The New York Times,* 17.10.2019; M. WHITTAKER, *Written Testimony of Meredith Whittaker, Facial Recognition Technology (Part III): Ensuring CommerciaI Transparency & Accuracy, United States House of Representatives Committee on Oversight and Reform*, 15.1.2020.

72 J. VALORON, P. PEÑA, *Not My AI: A feminist framework to challenge algorithmic decision-making systems deployed by the public sector*, Coding Rights, Feminist Internet Research Network.

73 Vedi A. B. MUÑOZ RUIZ, *Biometría y sistemas automatizados de reconocimiento de emociones: implicaciones jurídico-laborale*s, Tirant lo Blanch, Valencia, 2023.

74 M. CHEONG, et aI., *Ethical implications of AI bias*, *op. cit*., p. 11.

75 C. O'NEIL, *Weapons of Math Destruction*, *op. cit*., p. 17.

Per esempio, è dimostrato che la distanza tra il lavoro e la casa di una persona è una variabile *proxy* per prevedere la probabilità che la persona rimanga più a lungo nell'impresa[76], in modo che le persone che risiedono più lontano hanno maggiori probabilità di accettare un lavoro più vicino alla loro residenza. Tuttavia, dati i prezzi ingenti dell'alloggio nel centro di molte grandi città, è anche una variabile che può dare luogo a discriminazione per reddito e origine razziale[77], in quanto le persone che risiedono più lontano, probabilmente saranno anche quelle con meno risorse economiche o pendolari. Di conseguenza, un'azienda che utilizza questa variabile in un processo di selezione potrebbe discriminare le persone a causa della loro origine razziale o capacità economica, entrambe variabili protette.

Un altro esempio è quello della *start-up* GiId[78], che offriva servizi di ricerca di talenti alle imprese tecnologiche, introducendo come novità un modello che permetteva di quantificare il «capitale sociale» di persona, definito come la loro integrazione nella comunità scientifica.

Il modello, addestrato in un database di oltre 6 milioni di programmatori, ha trovato una corrispondenza statistica tra le abilità di programmazione di una persona e la partecipazione a forum di programmazione e, cosa che è più sorprendente, la frequentazione di uno specifico sito di manga giapponese. Considerando la disuguale distribuzione dei compiti di cura tra donne e uomini, non è sorprendente sapere che le donne hanno meno tempo per trascorrere ore e ore a risolvere enigmi nei forum di programmazione. Inoltre, la presenza di uomini in questi forum o il contenuto sessuale che caratterizza il manga giapponese potrebbero non essere attraenti per molte donne nonostante posseggano capacità di programmazione eccezionali[79].

Gli algoritmi non distinguono tra correlazione e causalità[80]; di conseguenza, possono prendere decisioni sulla base di correlazioni statistiche, anche se non hanno nulla a che fare con la decisione adottata.

Nel caso della start-up GiId, l'algoritmo ha identificato una corrispondenza statistica tra la visita a un sito giapponese e buone capacità di programmazione. Di conseguenza, in base a queste decisioni, premia coloro che visitano il sito. Tuttavia, non è una questione di causalità; non per il gusto del manga giapponese lo sviluppatore ha buone abilità di programmazione.

Si tratta di una distorsione correlata, anche se probabilmente rafforzata da una distorsione presente nella base di dati di formazione che vede una presumibile maggiore presenza di uomini e, altrettanto presumibilmente, una minore presenza di donne in questo settore.

Un altro esempio di discriminazione si può trovare nell'uso dei giochi nei processi di selezione del personale. Alcune aziende utilizzano sistemi di intelligenza artificiale nei processi di selezione che valutano le persone in base alle loro prestazioni in determinati videogiochi[81]. Giochi semplici, come ad esempio, soffiare nei palloncini o sparare delle palline, sviluppati da aziende come Pymetrics o Knack, sono utilizzati per misurare variabili

76 C. O'NEIL, *Weapons of Math Destruction*, *op. cit*., p. 119.
77 P. KIM, *Big Data and Artificial Intelligence*, *op. cit.*, p. 317.
78 D. PECK, *They're Watching You at Work*, *The Atlantic*, diciembre 2013; C. O'NEIL, *Weapons of Math Destruction*, *op. cit.*, p. 120-121.
79 C. CRIADO PEREZ, *Invisible Women*, *op. cit.*, p. 107.
80 A. COSTA, *et al.*, *Hiring Fairly in the Age of Algorithms*, *op. cit.*, p. 11.
81 H. SCHELLMANN, *The Algorithm*, *op. cit.*, p. 51 y ss.

come attenzione, assertività, capacità di decisione, sforzo, emozione, focalizzazione, generosità, capacità di apprendimento o rischio di morte nelle persone[82].

Il modello misura le prestazioni in questi giochi dei migliori profili inseriti nella società e le usa come standard di riferimento per valutare i candidati[83]. Senza dubbio, si deve mettere in discussione l'adeguatezza di queste tecniche di valutazione basate su giochi e la loro applicazione nei processi di selezione dove le persone mettono in gioco il loro sostentamento economico[84]. Per di più, possono generare un effetto discriminatorio se sono state evidenziate differenze tra uomini e donne nell'esecuzione di tali giochi, nonché in base all'età[85].

Le differenze di genere non sembrano essere dovute a differenze nel gioco tra donne e uomini, in quanto si stima che le donne rappresentino la metà delle persone che giocano ai videogiochi a livello mondiale[86], nonostante la presenza femminile minore nei personaggi raffigurati nei videogiochi[87]. Tuttavia, esistono differenze di genere ed età nel comportamento adottato nei giochi utilizzati nel contesto di un processo di selezione che possono generare una situazione discriminatoria. Da ciò possono anche derivare problemi di discriminazione per malattia o disabilità; nel misurare il tempo di reazione delle persone o il tempo di compimento di un determinato compito, è possibile che si stia escludendo ingiustamente persone con una determinata malattia o disabilità, nonostante non abbiano nessuna reazione rapida con l'impiego[88].

3. Il trattamento giuridico della discriminazione algoritmica

Una prima conclusione sul trattamento giuridico della discriminazione algoritmica è che, a mio avviso, può essere inglobata nella dottrina antidiscriminatoria[89], senza che sia necessario creare nuove categorie giuridiche[90]. Per fortuna o disgrazia, la

82 L. ANDREWS, H. BUCHER, *Automating Discrimination: AI hiring practices and gender inequality, Cardozo Law Review*, vol. 44, nº 1, 2022, p. 185-186.
83 H. SCHELLMANN, THE ALGORITHM, *op. cit.*, p. 63.
84 L. ANDREWS, H. BUCHER, *Automating Discrimination*, *op. cit.*, p. 189.
85 K. G. MELCHERS, J. M. BASCH, *Fair pIay? Sex, age and job-related correlates of performance in a computer based simulation game», International Journal of Selection and Assessment*, nº 30, p. 48-61.
86 V. CHEN, *LeveIing Up the Gaming Gender FieId*, Forbes, 24.8.2023.
87 C. CRIADO PÉREZ, *Invisible Women*, *op. cit.*, p. 12.
88 J. R. MERCADER UGUINA, *Algoritmos e inteligencia artificial en el derecho del trabajo*, Tirant lo Blanch, Valencia, 2022, p. 76-77.
89 A. GINÈS I FABRELLAS, *Sesgos discriminatorios*, *op. cit.*, p. 312.
90 In questo senso anche C.H. PRECIADO DOMENECH, *Algoritmos y discriminación en la relación laboral*, Jurisdicción Social, nº 223, 2021, p. 17; P. RIVAS VALLEJO, *La aplicación de la Inteligencia Artificial al trabajo y su impacto discriminatorio*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020, p. 303; A. FERNÁNDEZ GARCÍA,*Trabajo, algoritmos y discriminación*, in M. RODRÍGUEZ-PIÑERO ROYO, A. TODOLÍ SIGNES, (Direttori), *Vigilancia y control* in el Derecho del Trabajo Digital, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2020, p. 524. È interesante altresì C. SÁEZ LARA, che si interroga sulla capacità della discriminazione indiritta di que cuestiona la la capacità della discriminazione indiretta di descrivere le nuove problematiche sorgono dalla discriminazione algorítmica. L'autore è favorevole ad un ampliamento della tutela (C. SÁEZ LARA, *El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación*, Temas Laborales, nº 155, 2020, p. 49). In senso simile, D. PÉREZ DEL PRADO ritiene necessario articolare un concetto particolare di discriminazione e verifica dei casi che coinvolgono algoritmi o adattare quello di discriminazione indiretta per rendere più flessibile l'inversione dell'onere della prova o stabilire una sorta di presunzione quasi oggettiva (D. PÉREZ DEL PRADO, *Derecho, economía y digitalización. El impacto de la inteligencia artificial, los algoritmos y la robótica sobre el empleo y las*

discriminazione algoritmica non è che una nuova e sofisticata espressione di un vecchio e vergognoso problema.

Tuttavia, la discriminazione algoritmica pone alcuni dubbi sul suo trattamento giuridico, che meritano un'analisi e una riflessione più approfondita: (1) la classificazione della discriminazione algoritmica come discriminazione diretta o indiretta; (2) le prove adatte per l'inversione dell'onere della prova, in particolare la prova statistica nei casi di discriminazione indiretta; (3) il trattamento giuridico della discriminazione per procura; (4) la possibilità o meno di utilizzare la correttezza statistica come giustificazione obiettiva, ragionata e proporzionata per escludere l'esistenza di una discriminazione; (5) il trattamento giuridico della comparsa di nuove cause di discriminazione, come la capacità economica o l'intersezione.

3.1. Discriminazione diretta o indiretta?

In primo luogo, una volta individuata l'inclusione della discriminazione algoritmica nella dottrina antidiscriminatoria, occorre individuare se la discriminazione algoritmica debba essere inglobata all'interno della discriminazione diretta o indiretta.

Come è noto, la discriminazione diretta si verifica quando una persona viene trattata in modo meno favorevole di un'altra per uno dei motivi di discriminazione vietati dalla Costituzione o dalla legge [articolo 6.1 della Legge Organica 3/2007, del 22 marzo, per l'uguaglianza effettiva di donne e uomini (LOI, da qui in avanti)].

Esiste una discriminazione indiretta, invece, quando una disposizione, criterio o pratica apparentemente neutri pone le persone di un gruppo protetto in una situazione di particolare svantaggio rispetto alle persone di un altro gruppo, «a meno che tale disposizione, criterio o pratica possano essere oggettivamente giustificati in considerazione di uno scopo legittimo e i mezzi per conseguire tale scopo siano necessari e adeguati» (articolo 6.2 LOI).

Secondo l'articolo 14 CE, le persone «sono uguali davanti alla legge, senza che possa prevalere alcuna discriminazione fondata sulla nascita, sulla razza, sul sesso, sulla religione, sulle opinioni o qualsiasi altra condizione o circostanza personale o sociale». In particolare nel settore del lavoro, l'articolo 4.2.c) ET riconosce il diritto dei lavoratori a «non essere discriminati direttamente o indirettamente per l'occupazione, o una volta assunti, per motivi di sesso, stato civile, età entro i limiti stabiliti dalla presente legge, origine razziale o etnica, condizione sociale, religione o convinzioni, idee politiche, orientamento sessuale, affiliazione o non appartenenza a un sindacato, nonché per motivi di lingua, all'interno dello Stato spagnolo. Sono altresì vietate le discriminazioni per motivi di disabilità, purché il personale sia in grado di svolgere il lavoro o l'occupazione in questione». Inoltre, è importante citare anche l'articolo 2.1 della legge

condiciones de trabajo, Tirant lo Blanch, Valencia, 2023, p. 188). In senso simile, A. TODOLÍ SIGNES suggerisce che in caso di discriminazione algoritmica non è necessario stabilire un panorama indicativo preventivo, lasciando a carico dell'impresa l'onere della prova, (A. TODOLÍ SIGNES, Adrián, *Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, Cizur Menor (Navarra), 2023, p. 73).

15/2022 del 12 luglio, che integra la parità di trattamento e non discriminazione (legge 15/2022, in appresso), che stabilisce « non può essere discriminato per nascita, origine razziale o etnica, sesso, religione, convinzione o opinione, età, disabilità, orientamento o identità sessuale, espressione di genere, malattia o stato di salute, stato sierologico e/o predisposizione genetica a patologie e disturbi, lingua, situazione socioeconomica o qualsiasi altra condizione; oppure circostanze personali o sociali», disposizione applicabile anche al lavoro subordinato (articolo 3.1.a).

Stabilire se la discriminazione algoritmica costituisce una discriminazione diretta o indiretta è rilevante in quanto, anche se in entrambi i casi si tratta di un comportamento discriminatorio vietato, ne sono influenzati gli indici probatori da evidenziare per l'esercizio dell'inversione dell'onere probatorio e la giustificazione concernente il gruppo e il diverso trattamento operato nei confronti del gruppo protetto. Come già osservato, in caso di discriminazione indiretta, a differenza di quanto avviene in caso di discriminazione diretta, il comportamento non è discriminatorio se la pratica o disposizione che genera effetti negativi sul gruppo protetto può essere obiettivamente e proporzionalmente giustificata.

Le cause di discriminazione algoritmica individuate nel paragrafo precedente non determinano, a mio avviso, la classificazione giuridica del comportamento o della pratica come discriminazione diretta o indiretta. Identificare se l'origine del pregiudizio che genera la situazione discriminatoria si trova nelle variabili, nel database di formazione o nelle correlazioni è sufficiente per valutare l'esistenza della discriminazione e identificare gli indizi necessari per operare l'inversione dell'onere della prova e la prova adeguata per disassemblare l'esistenza della discriminazione; ma non permette di classificare automaticamente la situazione come discriminazione diretta o indiretta.

La classificazione della discriminazione algoritmica come diretta o indiretta dipenderà, quindi, dal pregiudizio generato. La discriminazione algoritmica rientra nel campo della discriminazione diretta quando una decisione adottata arbitrariamente comporta il trattamento meno favorevole di un'altra persona in considerazione di una delle cause di discriminazione vietate, quali sesso, razza, religione, invalidità, età, ecc.

Mentre rientra nel campo della discriminazione indiretta quando, nonostante si tratti di una decisione o pratica apparentemente neutra, comporti uno svantaggio particolare per un gruppo protetto e non vi è alcuna giustificazione obiettiva e proporzionata.

In questo senso, le situazioni di discriminazione algoritmica diretta sono, per esempio, l'uso di tecniche di pubblicità segmentata che minimizzano la visualizzazione su reti sociali di un'offerta di lavoro in correlazione a variabili protette, come ad esempio l'età o il sesso. Inoltre, in questo senso l'esempio concernente l'analisi della selezione del personale di Amazon che escludeva automaticamente i curricula delle donne, cadrebbe sotto la definizione di discriminazione diretta[91]. L'algoritmo aveva identificato nei dati forniti un modello statistico che, escludendo automaticamente le donne dal processo di selezione, è come se avesse incorporato la variabile sesso nel processo decisionale. Tuttavia, dal mio punto di vista, la discriminazione algoritmica costituisce generalmente un caso di

91 J. ADAMS-PRASSL, R. BINNS, A. KELLY-LYTH, *Directly Discriminatory Algorithms, The Modern Law Review*, vol. 86, nº 1, p. 167.

discriminazione indiretta, in quanto si tratta di decisioni o pratiche aziendali apparentemente neutre che generano un impatto negativo su uno dei gruppi protetti.

Così, per esempio, costituisce una discriminazione indiretta l'algoritmo utilizzato da Deliveroo che penalizza persone per assenze ingiustificate, anche se genera uno svantaggio particolare sulle persone che soffrono di una malattia, invalidità, che hanno responsabilità familiari, o esercitano il diritto di sciopero. Costituisce, altresì, discriminazione indiretta un algoritmo in un proceso di selezione che penalizza le persone con interruzioni della loro carriera personale, in ragione di una discriminazione relativa ad una causa di malattia, disabilità o sesso, nel caso in cui l'evenienza sia dovuta ad un determinato stato di salute, o per far fronte a necessità di cure nei confronti di figli e familiari.

Oppure, per fare un altro esempio, potrebbe anche costituire una discriminazione indiretta un algoritmo che in un processo di selezione penalizza le persone che risiedono più lontano.

Come analizzato in precedenza, se l'azienda è situata nel centro di una grande città e esclude persone che risiedono più lontano, può verificarsi una discriminazione per razza o situazione socioeconomica.

3.2. Pregiudizi statistici come indizi di discriminazione

Il processo di tutela dei diritti fondamentali, regolato dagli articoli 177 e seguenti della legge 36/2011 del 10 ottobre, che disciplina la giurisdizione sociale (LRJS, in appresso), è caratterizzato, tra l'altro, dall'inversione della prova.

L'articolo 181.2 LRJS stabilisce che «una volta accertata la presenza di indizi che vi sia stata violazione del diritto fondamentale o della libertà pubblica, spetta al convenuto fornire una giustificazione obiettiva e ragionevole, sufficientemente provata, delle misure adottate e della loro proporzionalità». Di conseguenza, spetta alla parte ricorrente fornire indizi che vi sia stata una discriminazione - indizi intesi come fattori o elementi che consentono di valutare la possibilità della riaffermazione di un diritto fondamentale. La parte convenuta è chiamata a fornire una causa oggettiva, ragionevole e proporzionale che dimostri l'illegittimità della decisione o della pratica aziendale.

In caso di discriminazione indiretta, è importante ricordare l'importanza della prova statistica come indicatore adeguato per l'inversione del carico di prova[92]. Cioè, l'evidenza statistica che una determinata disposizione, criterio o pratica genera uno svantaggio particolare per persone di un gruppo protetto rispetto alle persone di un altro gruppo è un indizio idoneo per operare l'inversione dell'onere della prova e, che quindi spetta al convenuto l'accertamento dell'esistenza di una giustificazione obiettiva, ragionata e provata della situazione.

In primo luogo, l'evidenza statistica adeguata di discriminazione algorítmica, per costituire indizio sufficiente di discriminazione e, quindi, invertire l'onere della prova ex articolo 181.2 LRJS potrebbe essere l'evidenza statistica relativa alle decisioni dell'algoritmo. Cioè dati statistici relativi ai risultati delle decisioni prese dall'algoritmo. Per esempio, in un processo di selezione del personale, differenze percentuali per sesso nel rapporto tra il

92 STJUE 28.2.2013 (caso c-427/11, asunto *Kenny et al.*).

numero di persone selezionate e il numero di richieste ricevute. Ossia, la comparazione dei dati disaggregati per sesso del numero di richieste ricevute in relazione al numero di persone ammesse nelle diverse fasi del processo di selezione.

Far riferimento ai risultati dell'algoritmo permette di analizzare giuridicamente i potenziali impatti discriminatori degli algoritmi denomanati a scatola nera. Esistono algoritmi che prevedono complesse tecniche di *machine e deep learning* che incorporano molteplici strati decisionali e generano vere e proprie «scatole nere», che impediscono che le decisioni possano essere completamente spiegate[93]. La mancanza di conoscenza delle variabili esatte utilizzate o della loro ponderazione precisa nell'equazione non impedisce il trattamento giuridico del loro impatto discriminatorio, in quanto possono essere forniti come indicatori statistici relativi alle decisioni adottate dall'organo consultivo. Ciò non pregiudica l'obbligo di rispettare i diritti di informazione per l'uso dei sistemi decisionali automatizzati ai sensi degli articoli 13, 14 e 15 del GDPR, in conformità dell'articolo 22[94], e gli obblighi di informazione e trasparenza dei sistemi di informazione artificiale del rischio che introduce il regolamento sulle infrastrutture articolate negli articoli 13 e 26[95].

Ai fini della valutazione di una discriminazione per un algoritmo a «scatola nera», il dato giuridicamente rilevante è l'effetto che genera l'algoritmo e non la spiegazione técnica della decisione adottata. Cioè, il fatto che l'allettamento crei uno svantaggio particolare per un gruppo protetto e non ci sia una giustificazione obiettiva, ragionata e proporzionata per esso. Se l'azienda utilizza un meccanismo di «scatola nera» per prendere decisioni, è probabile che sia più difficile fornire una giustificazione obiettiva e ragionata per escludere l'esistenza della discriminazione. Tuttavia, questa decisione aziendale non può ricondursi a vantaggio dell'impresa di generare svantaggi particolari rispetto ai contratti protetti. Di conseguenza, l'evidenza statistica relativa allo svantaggio particolare generato dall'algoritmo deve essere sufficiente per valutare indizio di discriminazione e invertire l'onere della prova. Di conseguenza, intendo che debba ammettersi come indizio di discriminazione perchè possa operare un'inversione dell'onere della prova. In secondo luogo, a mio avviso, dovrebbe anche costituire indizio di discriminazione l'esistenza di differenze sessuali, razziali o altre cause protette nel tasso di errore del punteggio. Per esempio, la percentuale più alta di donne nere erroneamente classificate con un male professionale perfilico rispetto agli uomini bianchi potrebbe anche costituire indizio sufficiente, a mio avviso, per invertire l'onere della prova ex articolo 181.2 LRJS.

93 P. B. DE LAAT, *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability*, *Philosophy & Technology*, nº 31, 2017, p. 11 (versione digitale); M. CHEONG, *et al.*, *Ethical implications of AI bias*, *op. cit.*, p. 14.

94 Per quanto riguarda gli obblighi di informazione in relazione alle decisioni automatizzate nel settore del lavoro, vedi A. GINÈS I FABRELLAS, *Decisiones automatizadas y elaboración de perfiles en el ámbito laboral y su potencial impacto discriminatorio*, in A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y Relación Laboral*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2023, p. 173-229.

95 È interessante notare anche le voci che sostengono l'uso di sistemi decisionali automatizzati più semplici a beneficio della (vedi C. RUDIN, *Stop ExpIaining BIack Box Machine Learning Models for High Stakes Decisiones and Use Interpretable Models Instead*, *Nature Machine Intelligence*, voI. 1, 2019, p. 1 -20 (versione digitale).

In terzo luogo, a mio avviso, l'esistenza di pregiudizi per motivi di discriminazione vietati nella base di dati utilizzata per la formazione e il perfezionamento dovrebbe essere ammessa anche come prova sufficiente per valutare indizi di discriminazione. Come analizzato in precedenza, l'esistenza di pregiudizi nel database di formazione è la causa più abituale di errori sistematici[96]. Di conseguenza, intendo che debba ammettersi come indizio di discriminazione per l'inversione dell'onere della prova.

Sebbene sia già stato affermato che il trattamento giuridico della discriminazione algoritmica può essere ricondotto alla dottrina antidiscriminatoria, ritengo che il requisito degli indizi di discriminazione dovrebbe essere adattato alle prove statistiche disponibili nei casi di utilizzo degli algoritmi per la presa di decisioni - in particolare data l'opacità e la mancanza di trasparenza che caratterizzano questi modelli, si ammette come indizio il diverso tasso di errore o l'esistenza di pregiudizi nel database di formazione dell'algoritmo. Si noti che la discriminazione algoritmica può essere meno evidente o apparente, poiché le persone possono non avere conoscenza o percezione di essere discriminate[97]. Per questo motivo ritengo importante l'ammissione di altre prove statistiche come indizio di discriminazione, al fine di garantire la protezione giuridica contro forme di discriminazione algoritmica.

In ogni caso, si tratta di prove statistiche che servono unicamente come indizio di discriminazione ai fini dell'inversione dell'onere della prova e che, pertanto, spetta al convenuto fornire «una giustificazione oggettiva e ragionevole, sufficientemente provata, delle misure adottate e della loro proporzionalità», secondo quanto stabilito all'articolo 181.2 LRJS. Se il pregiudizio presente nel database di formazione non si traduce in pregiudizi nelle dedizioni adottate, qualora fosse provato dall'azienda, la discriminazione non sussiste. Cioè, è il reale effetto discriminatorio ad essere giuridicamente rilevante. Se un'impresa utilizza un approccio di parte, ma la decisione è compiuta, ad esempio, mediante revisione umana, in modo tale da essere «neutrale» e la decisione finale non è influenzata negativamente, ciò non costituisce una condotta discriminatoria[98].

3.3. Trattamento giuridico della discriminazione proxy?

Il trattamento giuridico della discriminazione algoritmica pone anche un dubbio sulla discriminazione *proxy;* sarebbe a dire il trattamento giuridico dei casi di discriminazione basata su pregiudizi nelle correzioni statistiche o nelle variabili *proxy* utilizzate dall'algoritmo decisionale[99].

Sebbene la variabile proxy sulla base della quale l'algoritmo fonda la sua decisione non costituisca una causa di discriminazione vietata, può tuttavia costituire una discriminazione giuridicamente vietata quando genera, come richiede la discriminazione indiretta, uno svantaggio particolare per le persone di un determinato gruppo protetto. Per esempio,

96 A. COSTA, *et al.*, *Hiring FairIy in the Age of Algorithms*, *op. cit.*, p. 11.
97 S. WACHTER, B. MITTELSDTADT, C. RUSSELL, *Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI*, *Computer Law & Security Review*, voI. 41, 2021, p. 6 (versione digitale).
98 M. KULLMANN, *PIatform Work…*, *op. cit.*, p. 10.
99 B. DATTNER, *et al.*, *The Legal and Ethical Implications of Using AI in Hiring*, *op. cit*.

l'algoritmo della start-up Gild che assisteva alla partecipazione ai forum di programmazione e la visita a una pagina manga giapponese non utilizza un criterio di discriminazione vietato dalla Costituzione o dalla legge. Tuttavia, se questa variabile *proxy* genera uno svantaggio particolare su un gruppo di donne, essa costituisce una pratica discriminatoria basata sul sesso.

L'uso di una variabile *proxy* non direttamente correlata con una variabile vietata non permette di escludere automaticamente l'esistenza di discriminazione. Come già osservato, in caso di discriminazione indiretta, l'esistenza di una disposizione, pratica o criterio apparentemente neutri non impedisce che essa sia considerata discriminatoria. Ciò che è giuridicamente rilevante non è la variabile di decisione, ma l'effetto discriminatorio. Cioè, il fatto che tale variabile generi uno svantaggio particolare su uno dei contratti protetti senza che esista una giustificazione obiettiva, ragionata e proporzionata.

Di conseguenza, quando la variabile utilizzata dall'individuo è un perfetto sostituto di una causa di discriminazione vietata o è una variabile apparentemente neutra e «innocua» ma genera uno svantaggio particolare per le persone di un gruppo protetto, si tratta di una pratica discriminatoria legalmente vietata. In questo contesto, occorre anche chiedersi quali siano le prove statistiche sufficienti a costituire un indizio di discriminazione. Da un lato, capisco che costituirà un sufficiente indizio per attivare l'inversione dell'onere della prova la prova statistica di un particolare svantaggio che l'algoritmo genera sulle persone del gruppo protetto. Per esempio, utilizzando il caso dell'analisi di start-up Giid, i dati relativi alle differenze statistiche nel punteggio ricevuto da uomini e donne nel processo di selezione dovrebbero costituire un indizio sufficiente di discriminazione per invertire l'onere della prova.

Per di più, a mio avviso, dovrebbe essere ammesso come indizio di discriminazione la relazione tra la variabile *proxy* e il motivo di discriminazione vietato. Sebbene sia giuridicamente rilevante, per l'effetto discriminatorio di una determinata disposizione, decisione o pratica, dovrebbe essere ammesso come indizio sufficiente il collegamento tra la variabile *proxy* utilizzata dall'organo di gestione per prendere decisioni e una causa di discriminazione vietata. Tornando al precedente esempio, variabile decisionali come le informazioni sulla partecipazione a forum di programmazione o sul consumo di manga giapponesi e il fatto che può generare uno svantaggio particolare sulle donne, costituiscono un indizio sufficiente per la configurazione di una discriminazione al fine di invertire l'onere della prova.

3.4. La correlazione statistica è una giustificazione obiettiva, ragionevole e proporzionata?

Nell'analisi relativa al trattamento giuridico della discriminazione algoritmica, è opportuno chiedersi anche se la correlazione statistica identificata dall'algoritmo tra una determinata variabile *proxy* e un'attitudine, qualifica o caratteristica personale o professionale può essere utilizzata come giustificazione obiettiva, ragionevole e proporzionata per escludere l'esistenza di una discriminazione.

La risposta a questa domanda deve essere, a priori negativa. L'esistenza di una correlazione statistica tra una determinata variabile *proxy* e un'attitudine o caratteristica

personale o professionale non può necessariamente essere considerata una giustificazione oggettiva per il vantaggio particolare, ragionevolmente e proporzionalmente generato sul gruppo protetto[100].

La statistica non indica causalità. Cioè, il fatto che esista una correlazione nei dati statistici tra una determinata variabile e un'attitudine o caratteristica personale o professionale non significa che una sia riconducibile a un'altra. Semplicemente implica che ci sia una percentuale più alta di persone che aderiscono a questa variabile e condividono questa attitudine o caratteristica rispetto a quelli che non lo fanno. L'algoritmo identifica questa connessione statistica nel database di formazione e capisce che è una variabile che permette di prevedere tale attitudine o caratteristica e quindi la utilizza al momento di prendere una decisione.

La correlazione statistica può, a mio avviso, fungere da giustificazione oggettiva, ragionata e proporzionata per escludere l'esistenza di una discriminazione in funzione del fatto che la correttezza sia anche causale. Quando la correzione statistica è conseguenza di una relazione causale tra la variabile *proxy* e l'attitudine o caratteristica che si vuole prevedere, si potrebbe agire per escludere l'esistenza della discriminazione. Al contrario, una correzione statistica che non determina causalità e che a volte può essere abbastanza arbitraria, come la visita di una determinata pagina di manga giapponese, come è stato osservato nell'esempio della start-up Gild, non può essere usata come giustificazione obiettiva, ragionata e proporzionata.

Di conseguenza, per escludere l'esistenza di una discriminazione, l'impresa dovrà fornire la giustificazione della necessità di prendere in considerazione una determinata variabile *proxy* per valutare la persona in un processo di selezione o per adottare una determinata decisione di lavoro. Cioè, la relazione di causalità – al di là della mera connessione statistica - tra la variabile *proxy* e la capacità o caratteristica che si intende prevedere o valutare. Inoltre, dal mio punto di vista, per evitare di cadere nell'arbitrarietà deve essere provata anche la reazione con il posto di lavoro da occupare o con la decisione da adottare.

Allo stesso modo, costituirebbero prova sufficiente per escludere l'esistenza di una discriminazione, ad esempio l'assenza di uno svantaggio particolare su un gruppo protetto, nonostante l'uso di una variabile *proxy* correlata ad un motivo di discriminazione vietato. Oppure l'esistenza di cause oggettive che spiegano le differenze tra gruppi, per esempio, che, nonostante la variabile *proxy* utilizzata, esistono ragioni di formazione o precedente esperienza professionale che giustificano la maggior proporzione tra gli uomini e le donne selezionati in un determinato processo di selezione.

3.5. Nuove cause di discriminazione vietate?

L'uso di modelli predittivi per la ricerca e il trattamento dei pregiudizi può promuovere «nuovi» ambiti di discriminazione. Informazioni finora non accessibili da parte dell'azienda possono essere dedotte da informazioni disponibili e apparentemente innocue e utilizzate per prendere decisioni.

100 P. KIM, *Big Data and Artificial Intelligence*, *op. cit.*, p. 325.

L'uso di sistemi di intelligenza artificiale può aumentare la discriminazione basata sul sesso o sulla capacità economica, sia prendendo decisioni in base alla situazione socioeconomica delle persone o focalizzando l'uso di sistemi decisionali automatizzati nelle comunità più povere. Come hanno identificato Cathy O'NEIL[101] e Virginia EUBANKS[102], l'uso di algoritmi per il processo decisionale automatizzato genera una specializzazione nella penalizzazione delle persone più povere, sottoponendole a sorveglianza e controllo superiori. Per esempio, la sentenza del tribunale di L'Aia del 5.2.2020[103] è contraria all'articolo 8 della Convenzione europea dei diritti dell'uomo che sancisce il diritto alla privacy, al sistema di indicazione del rischio di sistema (SyRI), utilizzato dalle autorità per prevenire e rilevare le frodi alla sicurezza sociale, in quanto veniva utilizzato solo nei quartieri residenziali di Ios che vivevano persone con redditi più bassi o persone appartenenti a minoranze.

Le modalità predittive possono anche identificare l'intersezione di identità multiple finora ignorate, potenziando forme di discriminazione *intersectional.* L'intersezione può essere definita come la discriminazione prodotta dall'intersezione o interazione di due o più cause di discriminazione vietata[104]. Un esempio classico di discriminazione intersettoriale è quello dei sistemi di riconoscimento facciale che, come hanno dimostrato Joy BUOLAMWINI E Timnit GEBRU[105], generano una percentuale di errore particolarmente elevata per le donne nere. Le percentuali di errore analizzate solo per sesso variabile (dal 10,7% al 21,3% per le donne) o per razza variabile (dal 12,9% al 22,4% per le persone con la pelle più scura), anche se inaccettabili, sono molto inferiori a quelle analizzate congiuntamente, ipotesi in cui si registra un tasso di errore significativamente più elevato per le donne nere (dal 20,8% al 34,7%).

La sfida giuridica che pone queste «nuove» forme di discriminazione è valutarne l'esistenza. Sebbene si tratti di cause di discriminazione ben note da un punto di vista sociologico, sono ancora molto inesplorate in ambito giuridico. È vero che sono identificate e menzionate nella Legge. Così, l'articolo 2.1, Legge 15/2022 che stabilisce le cause di discriminazione vietata ai «status socioeconomico» e all'articolo 6.3.b) la legge 15/2022 definisce l'intersezionalità come tale discriminazione «quando diverse cause delle previste in questa legge sono presenti o interagiscono, generando una specifica forma di discriminazione». Tuttavia, il loro trattamento legale è ancora molto marginale. Inoltre, la discriminazione intersettoriale, pur essendo espressamente definita all'articolo 6.3 della legge 15/2022, non è riconosciuta come causa specifica di discriminazione a livello europeo.

La Corte di giustizia dell'Unione europea[106] ha dichiarato che: «sebbene una discriminazione possa essere fondata su più di un motivo contemplato dall'articolo 1 della direttiva 2000/78, non esiste tuttavia alcuna nuova categoria di discriminazione derivante dalla combinazione di alcuni di tali motivi, come l'orientamento sessuale e l'età, che può sussistere quando non è stata accertata una discriminazione per tali motivi considerati

101 C. O'NEIL, *Weapons of Math Destruction*, *op. cit*., p. 8.
102 V. EUBANKS, *Automating Inequality*, *op. cit.*, p. 158 y 161.
103 C/09/550982 / HA ZA 18-388.
104 K. CRENSHAW, *The urgency of intersectionality*, *TEDWomen*, 2016.
105 J. BUOLAMWINI, T. GEBRU, *Gender Shades: Intersectional Accuracy Disparities*, *op. cit.*, p. 6.
106 STJUE de 24.11.2016 (asunto Parris, c-443/15).

separatamente. Di conseguenza, quando una disposizione nazionale non crea una discriminazione fondata sull'orientamento sessuale o su un'età, tale disposizione non può creare una discriminazione fondata sulla combinazione di entrambi».

4. Riflessioni finali. Nuove dimensioni al vecchio e vergognoso problema della discriminazione

I sistemi di intelligenza artificiali, come è stato descritto nel presente articolo, non mantengono la promessa di offrire modelli matematici oggettivi e meritocratici. Non si eliminano magicamente le discriminazioni esistenti, il che non dovrebbe sorprendere, dato che essi tendono a riprodurre i modelli statistici osservati nei dati[107], con le loro distorsioni.

Una soluzione paradossale dal momento che la tecno-logistica, creata sul dogma dell'efficienza, dell'oggettivazione e della meritocrazia, riproduce proprio quei pregiudizi che afferma di eliminare, perpetuando ed esplicando discriminazioni storiche e attualistiche.

È essenziale, quindi, decostruire il discorso degli algoritmi e dei sistemi di intelligenza artificiali come modelli matematicamente oggettivi e neutri e accettare la realtà sessista e discriminatoria in cui sono stati sviluppati[108]. Il Consiglio ha adottato una decisione relativa alla conclusione del l'accordo di cooperazione tra la Comunità europea e il Regno Unito.

La discriminazione algoritmica, tuttavia, aggiunge nuove dimensioni al problema della discriminazione ed è che i sistemi di intelligenza artificiali non riproducono solo le discriminazioni esistenti, ma si sistematizzano e si diffondono a grande e vertiginosa velocità[109]. Da un lato, l'uso di criteri discriminatori per prendere decisioni rende i presunti casi di discriminazione costanti, uniformi e invariabili. L'automazione delle decisioni attraverso meccanismi di *bias* fa sì che le decisioni di *bias* siano costanti, inflessibili e senza spazio per eccezioni o cambiamenti. Cercando di eliminare la soggettività umana dalla presa di decisione si riesce a evitare che tale soggettività possa agire come effetto correttivo delle discriminazioni esistenti[110].

D'altro canto, vi è evidenza scientifica che gli algoritmi formati su basi di dati distorte riproducono tale pregiudizio in misura maggiore[111], esaltando così le discriminazioni esistenti. Per esempio, gli strumenti utilizzati per i processi di selezione formati su immagini di Google daranno la priorità alle candidature degli uomini in settori mascolinizzati, anche in condizioni uguali[112], data la maggiore presenza di immagini di uomini[113]. Tuttavia, e proprio per questo motivo l'uso di questi modelli in ambito lavorativo è fonte di preoccupazione, la

107 J. POWEL, H. NISSEMBAUM, *The Seductive Diversion of «Solving» Bias in Artificial Intelligence*, *OneZero*, 7.12.2018.

108 R. BENJAMIN, *Race after technology*, Polity Press, Medford (EUA), 2019, p. 8 y 29; D. MURILLO, *Algoritmos, decisiones automatizadas y desafíos éticos: un análisis sociológico*, in A. GINÈS I FABRELLAS (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 53.

109 T. BOLUKBASI, K.W. CHANG, J. ZOU, V. SALIGRAMA, A. KALAI, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings, NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, p. 1.

110 C. O'NEIL, *Weapons of Math Destruction*, *op. cit*., p. 112.

111 J. ZHAO, WANG, TIANIY, M. YATSKAR, V. ORDONEZ, K.W. CHANG, *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints*, Research paper Artificial Intelligence, Cornell University, 2017, p. 1.

112 M. CHEONG, et al., *Ethical implications of AI bias*, *op. cit*., p. 9.

preferenza per il candidato maschile sarà maggiore rispetto alla rappresentazione di uomini e donne nelle immagini[114]. In particolare, un algoritmo formato da articoli di *Google News* che includa pregiudizi e stereotipi di genere nelle notizie pubblicate, riprodurrà tali pregiudizi in misura maggiore al momento di prendere decisioni[115]. Cioè, l'ideologia amplifica gli stereotipi e i pregiudizi di genere e dà la priorità alla candidatura maschile in percentuale maggiore della presenza degli uomini nei dati[116].

La tecnologia non è né causa né soluzione. I sistemi di sicurezza artificiali non creano discriminazione, ma nemmeno la eliminano per magia. Occorre pertanto adottare misure tecniche e giuridiche per evitare pregiudizi e discriminazioni nell'uso di strumenti per il processo decisionale automatizzato.

Sul piano tecnico è necessario garantire che gli algoritmi non perpetuino o amplino i pregiudizi sociali esistenti. Esistono alcune proposte tecniche per garantire l'assenza di pregiudizi nei sistemi di formazione, come modificare il database di formazione per garantire una sufficiente diversità[117] o addestrare gli ingegneri specificamente a eliminare i pregiudizi presenti nel database[118].

In questo senso, è importante ricordare che il regolamento di Istruzione Industriale, tra le varie obbligazioni rivolte alle aziende fornitrici di sistemi di istruzione artificiale ad alto rischio, si impegna a garantire la qualità dei dati di formazione, prendendo misure quali la realizzazione di una valutazione preliminare della disponibilità, quantità e adeguatezza dei dati necessari, e l'esame delle possibilità di pregiudizi esistenti nei dati che potrebbero generare rischi di pregiudizio al codice; diritti fondamentali o discriminazione delle persone e, se del caso, l'adozione di misure per attenuare tali rischi (articolo 10)[119].

Dal punto di vista giuridico è essenziale garantire l'efficacia delle norme sulla trasparenza e sull'informazione algoritmica che la normativa riconosce, per affrontare l'opacità e la mancanza di trasparenza che caratterizzano i sistemi decisionali automatizzati[120]. Sebbene il regolamento generale sulla protezione dei dati riconosca un diritto di informazione individuale algoritmica (articoli 13, 14 e 15 in conformità all'articolo 22 del GDPR) e l'articolo 64.4 d) ET un diritto di informazione alla rappresentanza algoritmica sui «i parametri, le regole e le istruzioni» utilizzati dai sistemi di decisione automatizzata, esistono dubbi

113 M. KAY, C. MATUSZEK, S.A. MUNSON, *Unequal Representation and Gender Stereotypes in image Search Results for Occupations*, *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, p. 5.
114 M. CHEONG, et al., *Ethical implications of AI bias*, *op. cit*., p. 9.
115 T. BOLUKBASI, *et al.*, *Man is to Computer Programmer as Woman is to Homemaker?, op. cit.*
116 J. ZOU, *Rimuovere il gender bias dagli algoritmi*, *op. cit*. Per questo motivo, per esempio, i sessi di genere identificati nei sistemi di intelligenza artificiale di generazione delle immagini sembrano superare la differenza tra le immagini di uomini e donne su Google.
117 S. DEVA, *Addressing the gender bias*, *op. cit*.
118 T. BOLUKBASI, *et al.*, *Man is to Computer Programmer as Woman is to Homemaker?*, *op. cit.*, p. 8.
119 È interesante la proposta di creare standard e certificazione per database e sistema di intelligenza artificiale (vedi Parlamento Europeo, *Auditing the quality of datasets used in algorithmic decision-making systems*, European ParIiamentary Research Service, 2022).
120 C. O'NEIL, *Weapons of Math Destruction*, op. cit., p. 28; V. EUBANKS, *Automating Inequality*, *op. cit*., p. 185.

interpretativi che ne complicano l'interpretazione[121] nonché reticenze aziendali a considerare il principio come segreto aziendale[122]. Alcune di queste difficoltà o reticenze sono affrontate attraverso il regolamento sulle garanzie industriali, che introduce obblighi d'informazione e trasparenza per le imprese fornitrici di sistemi di garanzia artificiale del rischio (articoli 8 e seguenti)[123], in merito al quale il Consiglio ha adottato una decisione relativa alla partecipazione del Regno Unito al l'IET.

Inoltre, i diritti di informazione agrometeorica attualmente riconosciuti non includono informazioni sull'impatto delle decisioni prese dal sistema decisionale automatizzato[124]. Le informazioni, infatti, come è stato discusso in questo articolo, sono rilevanti ai fini di valutare il potenziale impatto discriminatorio del loro comportamento e agire come indizi di discriminazione per l'inversione dell'onere della prova ex articolo 181.2 LRJS.

Di conseguenza, a mio avviso, il consenso politico e dottrinale sull'importanza della trasparenza nei sistemi di trasparenza artificiali[125] deve tradursi in misure di responsabilità (accountability, in inglese) per evitare il suo impatto discriminatorio[126]. Il Consiglio ha adottato una decisione relativa alla conclusione del l'accordo di cooperazione tra la Comunità europea e il Regno Unito. Tra le proposte più azzeccate, a mio avviso, vi è l'obbligo di realizzare audit indipendenti[127], in particolare per i sistemi di ispezione artificiali classificati come a rischio. La realizzazione di audit algoritmici analizzando gli effetti delle decisioni prese in modo disaggregato per diverse cause protette, includendo l'intersezione di molteplici identità[128], permetterà di determinare l'impatto discriminatorio dei modelli di decisione automatizzata. Infine, è necessario il riconoscimento del diritto di rappresentanza legale della famiglia e dei lavoratori a ottenere informazioni relative ai risultati di tale audit aleo-metrico[129]. Cioè, ottenere informazioni statistiche sulle decisioni adottate dal legislatore,

121 Rilievi giuridici circa il contenuto del diritto all'informazione algoritmica; dubbi che il Ministero del Lavoro ha voluto far valere pubblicando informazioni algoritmiche in ambito lavorativo: Guida pratica e strumento per l'obbligo di informazione aziendale sull'uso dell'algoritmo nell'ambiente di lavoro, maggio 2022.

122 M. KULLMANN, *Platform Work*, *op. cit.*, p. 15.

123 Vedi H. ÁLVAREZ CUESTA, *La propuesta de Reglamento sobre Inteligencia Artificial y su impacto en el ámbito laboral*, in A. GINÈS I FABRELLAS, (Direttrice), *Algoritmos, Inteligencia Artificial y relación laboral*, Thomson Reuters Aranzadi, 2023, p. 137-172.

124 A. GINÈS I FABRELLAS, *El derecho a conocer el algoritmo: una oportunidad perdida de la «Ley Rider»*, *IUSLabor*, nº 3, 2021, p. 4.

125 OIT, *Work for a brighter future. Global Commission on the Future of Work*, Organización Internacional del Trabajo, Ginebra, 2019, p. 44; H. ÁLVAREZ CUESTA, *El impacto de la inteligencia artificial en el trabajo: desafíos y propuestas*, Thomson Aranzadi, Cizur Menor, 2020, p. 100; V. DE STEFANO, *Algorithmic Bosses and How to Tame Them*, *op. cit.*, p. 14; J.R. MERCADER UGUINA, *Algoritmos: personas y números en eI Derecho Digital del trabajo*, *La Ley*, nº 48, 2021, p. 10.

126 C. O'NEIL, *Weapons of Math Destruction*, *op. cit.*, p. 218; S.M. WEST, M. WHITTAKER, MEREDITH, K. CRAWFORD, *Discriminating Systems*, *op. cit.*, p. 4.

127 S. M. WEST, M. WHITTAKER, K. CRAWFORD, *Discriminating Systems, op. cit.*, p. 4; AA.VV., *Directrices éticas para una IA fiable*, Grupo independiente de expertos de alto nivel sobre Inteligencia Artificial creado por la Comisión Europea, 2018, p. 24.

128 S. M. WEST, M. WHITTAKER, K. CRAWFORD, *Discriminating Systems*, *op. cit.*, p. 17.

129 In questo senso, è interessante notare l'articolo 10 della direttiva sul lavoro su piattaforme digitali che prevede l'obbligo di rialzare ogni due anni e con la partecipazione della rappresentanza della Commissione «una valutazione degli effetti di ciascuna decisione adottata o sostenuta dai sistemi automatizzati di monitoraggio e di processo decisionale che utilizzano la piattaforma digitale per le persone che lavorano su piattaforme, in particolare quando se del caso, per le loro condizioni di lavoro e la parità di trattamento sul posto di lavoro».

permettendo la valutazione del loro impatto discriminatorio e, se del caso, l'accesso a prove statistiche che possono agire come indizi di discriminazione algoritmica nel contesto di un ricorso giudiziale.