ALGORITMI DI GENERE? RISCHI DI DISCRIMINAZIONE NEL MERCATO DEL LAVORO

Ciro Clemente De Falco, Cristiano Felaco
Università degli Studi di Napoli Federico II

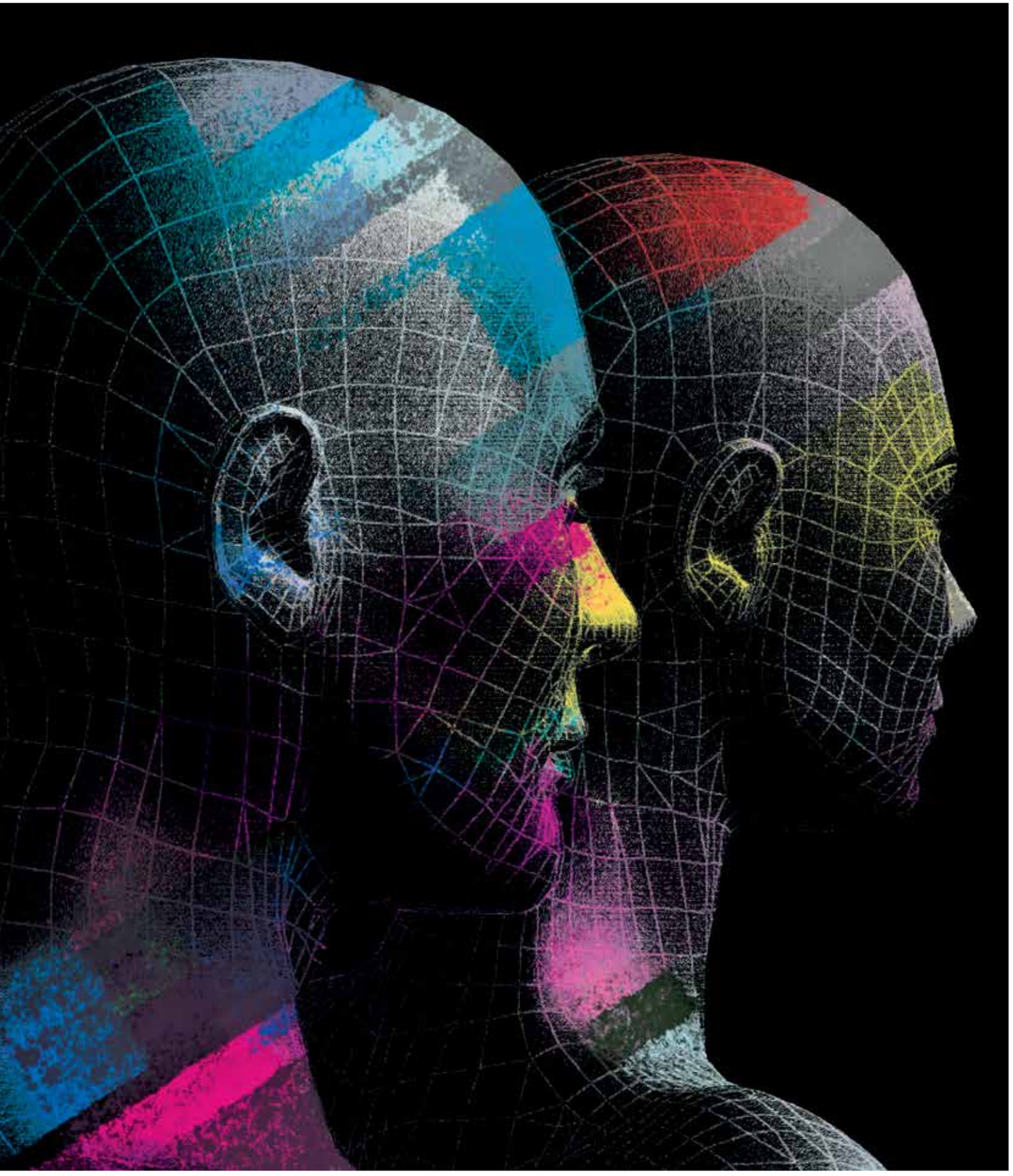
ciroclemente.defalco@unina.it
cristiano.felaco@unina.it

Abstract – Gli algoritmi sono parte integrante della vita quotidiana e delle attività che regolano vari settori della società. La mediazione algoritmica in queste attività spesso si traduce in operazioni di filtraggio, aggregazione, classificazione, valutazione e produzione di contenuti, ma anche nella presa di decisione riguardo la gestione e distribuzione dei servizi pubblici. L'implementazione degli algoritmi nella società produce indubbiamente dei benefici, tuttavia, quando avviene in modo a-critico, potrebbe anche comportare diversi rischi, come minacciare la privacy o limitare la libertà di espressione, o ancora esacerbare forme di discriminazione sociale, rafforzando i pregiudizi. L'articolo si concentra principalmente sulle ricadute degli esiti delle decisioni automatizzate sulle discriminazioni di genere all'interno dei processi di selezione del personale e di accesso al mercato del lavoro.

Parole chiave: Bias algoritmici; Discriminazione di genere; Mercato del lavoro.

Gli algoritmi sono imparziali, razzisti o sessisti. Queste sono solo alcune delle accuse di discriminazione tecnologica indirizzate agli algoritmi a svantaggio di alcune categorie di persone. Questa retorica è accompagnata e spesso alimentata dall'uso di termini antropomorfi come "imparano", "decidono", "si nutrono" e così via.
Probabilmente l'esempio di discriminazione algoritmica più noto è quello del mancato riconoscimento facciale del volto della scienziata afroamericana Joy Buolamwini, risolto con il ricorso ad una maschera di colore bianco. L'algoritmo era razzista?
La prima versione dell'app Salute di Apple permetteva di registrare qualsiasi attività, dai pasti ai numeri di passi in una giornata, ma non includeva la possibilità per le donne di registrare le date dei cicli mestruali. L'algoritmo era sessista?
Ancora. Negli anni'70 i primi airbag per le auto non proteggevano donne e bambini. L'algoritmo era un assassino?
La risposta a queste domande è ovviamente no. Gli algoritmi non hanno né coscienza né autonomia. Quello che è certo è che le macchine algoritmiche hanno sicuramente un impatto sulla società che, a seconda dei casi, può essere positivo o negativo.
Degli impatti sociali degli algoritmi sembra essere consapevole anche la Commissione Europea che nel 2020, nel "Libro bianco sull'intelligenza artificiale - Un approccio europeo", sottolinea come come la diffusione dell'intelligenza artificiale comporterà dei benefici in termini individuali e sociali, ma provocherà anche una serie di rischi potenziali, quali meccanismi decisionali opachi, discriminazioni basate sul genere o di altro tipo, intrusioni nelle nostre vite private o utilizzi per scopi criminali. Una prima iniziativa, concretizzatasi poi nella Proposta di regolamento sull'approccio europeo all'Intelligenza Artificiale (COM(2021) 206 final), propone il primo quadro giuridico europeo sull'Intelligenza Artificiale per la valutazione dei rischi al fine di salvaguardare i valori e i diritti fondamentali dell'UE e la sicurezza degli utenti (Barbera, 2021).
Quello che è certo è che gli esiti delle decisioni automatizzate dipendono dalle scelte di chi è coinvolto nella costruzione di un algoritmo.

ISSN 2724-1971



Come illustra brillantemente Aurélie Jean in "Nel paese degli algoritmi" (2021), chi crea un modello o un algoritmo può essere portatore di bias cognitivi. Chi lavora alla creazione di un algoritmo si basa sulla propria conoscenza del mondo che dipende dal bagaglio di conoscenze sul tema, dalle competenze di dominio e dalle pratiche lavorative adottate, dal tipo di informazioni che vengono loro trasmesse e dal loro grado di comprensione; oltre a dipendere dal fatto che le stesse conoscenze che vengono loro trasferite possono essere altrettanto parziali.
Eppure, se pensiamo alla definizione di algoritmo – ossia un percorso logico che da dati iniziali (input) produce un risultato desiderato (output) – questa sembrerebbe esaltare le caratteristiche di neutralità ed obiettività. Tuttavia, gli esiti sociali degli algoritmi rimetterebbero proprio in questione tali concetti di obiettività e neutralità. Il problema, infatti, è che anche gli algoritmi sono "artefatti" culturali e, come tutti gli artefatti realizzati dagli umani, risentono del sistema di significati, aspirazioni, idee, giudizi e credenze dei loro creatori e, indirettamente, di chi entra nel cosiddetto processo di assemblaggio algoritmico (Aragona & Felaco, 2020). La scelta dei parametri, delle equazioni matematiche e di specifiche ipotesi, ma anche del tipo di dati selezionati per addestrare gli algoritmi, non è né neutrale né oggettiva, ma determina gli esiti finali.
Le regole che guidano il funzionamento di un algoritmo riflettono dunque le visioni del mondo di coloro che progettano algoritmi, per cui le decisioni prese da un algoritmo potrebbero non necessariamente tenere in considerazione i pregiudizi sistemici incorporati nei dati, comportando nuovi rischi in termini di violazione della privacy ma anche di limitazione delle libertà di espressione, o più in generale esacerbando forme di discriminazione e di esclusione sociale (Eubanks, 2018). I bias cognitivi si tradurrebbero in questo modo in bias algoritmici.
Ritornando agli esempi iniziali, per quanto riguarda la ricercatrice del MIT, il mancato riconoscimento facciale era dettato probabilmente da un errore nella scelta dei parametri e dei criteri da parte del gruppo di sviluppatori che escludeva il rilevamento del contrasto cromatico proprio della pelle scura, oltre che all'uso di un training dataset parziale, contente immagini che rappresentavano soltanto volti di persone bianche.
Probabilmente i pregiudizi nelle scelte erano in parte dettate dalla natura stessa del gruppo di sviluppatori composto esclusivamente da uomini bianchi. Per la stessa ragione, l'app di Apple ha prodotto una discriminazione di genere, e ugualmente gli airbag non avevano

funzionato correttamente con donne e bambini in quanto i dati prendevano come modello la morfologia dell'uomo medio.
Ecco, punto di partenza delle riflessioni che seguono è la presunta neutralità degli algoritmi che possono provocare nuovi rischi e nuove forme di discriminazioni di genere.

Quali bias algoritmici?
In linea teorica gli algoritmi dovrebbero funzionare in modo equo senza discriminare i soggetti in base al loro genere di appartenenza. Tuttavia, come abbiamo visto, l'equità delle elaborazioni non è sempre garantita ed anzi gli

studi dimostrano che molto spesso gli algoritmi sono affetti da "bias". Ecco perché l'Unesco nel Report "Artificial Intelligence and gender equality" del 2020 sottolinea che c'è il rischio che gli algoritmi emarginino le donne su scala globale attraverso la diffusione ed il rafforzamento degli stereotipi di genere.
I bias algoritmici rappresentano più propriamente delle forme di distorsioni che possono produrre danni di assegnazione (harms of allocation) e danni di rappresentazione (harm of representation) (Crawford, 2017). I primi hanno un impatto di tipo economico, mentre i secondi agiscono a livello culturale.
Il danno di assegnazione si verifica quando un sistema ripartisce in modo iniquo opportunità e/o risorse. Quando invece i sistemi rafforzano gli stereotipi o sminuiscono determinati gruppi specifici siamo di fronte al danno di rappresentazione.
Cinque sono i possibili danni di rappresentazione: gli stereotipi, la sotto-rappresentazione, la denigrazione, il riconoscimento, e l"ex-nomination". Gli stereotipi, come è noto, si basano sulle false credenze ed attribuiscono a determinate categorie caratteristiche non vere. Il bias di sottorappresentazione, come evidenzia il termine, comporta una ridotta rappresentazione di un gruppo all'interno di un determinato campo. La denigrazione si riferisce all'uso di termini culturalmente o storicamente dispregiativi, mentre il bias di riconoscimento riguarda l'imprecisione di un dato algoritmo nel riconoscere i soggetti. Infine, l'ex-nomination descrive una pratica in cui una categoria è sminuita poiché semplicemente non le viene attribuito il giusto riconoscimento.
I bias algoritmici che colpiscono il genere sono diversificati e, come vedremo, hanno un impatto significativo soprattutto nei processi di selezione del personale e di accesso al mercato del lavoro.

Motori di ricerca e rappresentazione professionale delle donne

L'utilizzo dei motori di ricerca è sempre più frequente nella vita di tutti i giorni. Spesso incarichiamo i motori di ricerca di trovare per noi informazioni relative ad argomenti o oggetti di nostro interesse.
Dietro questa semplice operazione si celerebbe una sorta di autorità dell'algoritmo in quanto chi effettua la ricerca tende a non mettere in discussione ciò che il motore della ricerca ci restituisce.

Il risultato di una ricerca viene da noi percepito come una rappresentazione fedele della realtà e per questo motivo senza che noi ce ne rendiamo conto, l'algoritmo può influenzare il modo in cui la percepiamo. Uno dei rischi che si annida dietro questo uso pervasivo e ripetuto dei motori di ricerca è la perpetuazione delle norme o degli stereotipi esistenti.
Per capire se un algoritmo è affetto da bias di genere esistono diversi modi. Uno di questi consiste nel testare l'algoritmo. Nella pratica si studiano i risultati che il motore di ricerca restituisce su un set di parole chiave che rimandano a campi o a temi dove possono esistere discriminazioni di genere. La stessa procedura è stata utilizzata da alcuni ricercatori per capire se l'algoritmo di Google riproducesse gli stereotipi di genere connessi al lavoro (Metaxa et. al, 2021). A tal fine, hanno studiato le immagini che Google Image Search associava alle occupazioni più comuni come ad esempio, operaio, impiegato, architetto, ingegnere etc. giungendo a due risultati principali. In primo luogo, l'esistenza di un danno di rappresentazione nei confronti delle donne. Queste, infatti, risultano sottorappresentate in alcune categorie lavorative professionali come quella degli ingegneri. Lavori che per gli stereotipi di genere sono esclusivi degli uomini. Si potrebbe pensare che le sottorappresentazioni rilevate siano il frutto della reale sproporzione di presenze nelle diverse professioni (siccome ci sono meno ingegneri donna è normale che alla figura dell'ingegnere siano associate meno immagini femminili). Gli autori, invece, dimostrano come le donne siano sottorappresentate anche rispetto alla loro effettiva presenza negli ambiti analizzati. In secondo luogo, l'immagine veicolata delle donne ha certamente avuto delle conseguenze negative sulla categoria. Attraverso la somministrazione di questionari ai soggetti che avevano visto i risultati restituiti da Google Image Search, i ricercatori hanno osservato che le occupazioni dove le immagini delle donne erano sottorappresentate risultavano meno attrattive per le donne presenti nel campione. Se nel campo dell'architettura, ad esempio, c'è un bias di rappresentazione, allora le donne saranno meno interessate a quel campo perché lo riterranno meno vicino a loro. Si parla spesso di come è importante aumentare l'interesse delle donne verso le occupazioni che sono a prevalenza maschile per ridurre il gap di genere. I risultati della ricerca indicano che gli algoritmi possono andare esattamente nella direzione opposta.
La ricerca descritta è stata condotta nel 2020 e quello che preoccupa è che i risultati ottenuti dagli autori sono gli stessi rilevati da un altro team nel 2015. In altre parole, nonostante la questione fosse stata già segnalata ed esposta all'opinione pubblica, la piattaforma non ha comunque modificato l'algoritmo di ricerca per evitare danni di rappresentazione.

Algoritmi per il reclutamento del personale

Il bias algoritmico non si limita però alla rappresentazione delle donne nel mercato del lavoro. Le principali società di reclutamento online, come ad esempio LinkedIn, utilizzano algoritmi automatici per scegliere il migliore candidato per una posizione lavorativa. Questi algoritmi vengono definiti di raccomandazione poiché in base alle informazioni in loro possesso sui candidati consigliano quali fra questi è il più adatto per determinate posizioni lavorative. Anche in questo caso si è dimostrato come gli algoritmi di selezione potessero essere affetti da bias di genere. Il caso più famoso è quello dell'algoritmo usato da Amazon per l'assunzione di personale, in grado di elaborare migliaia di domande in pochi secondi (Reuters, 2018).
Il compito dell'algoritmo consisteva nel valutare i candidati comparando le loro qualifiche e caratteristiche con quelle dei candidati già assunti in precedenza. L'idea di fondo di questo procedimento era che i candidati con caratteristiche simili a quelli già assunti andavano valutati positivamente. Ora, siccome buona parte dei soggetti assunti in precedenza era di sesso maschile, l'algoritmo penalizzava sistematicamente i CV in cui comparivano parole al femminile. Il risultato finale era che alle donne venivano attribuiti punteggi peggiori rispetto agli uomini o venivano addirittura scartate dal processo di selezione. Va sottolineato che a causa di questo bias, Amazon ha abbandonato successivamente l'utilizzo di questo algoritmo. Tuttavia, questo esempio è abbastanza esplicativo di come gli algoritmi possono influenzare i corsi di vita di interi gruppi di persone.

Algoritmi per l'organizzazione e la valutazione del lavoro

Bias di genere possono nascondersi anche negli algoritmi di distribuzione dei carichi lavorativi o di valutazione del lavoro.
I software che si occupano di assegnare i compiti e i turni ai dipendenti sono sempre più utilizzati in campi come la vendita al dettaglio e quello della cura che sono campi a prevalenza femminile (Eurofound, 2021). Questi algoritmi si basano su una vasta gamma di dati (comportamento dei consumatori, vendite, modelli stagionali di consumo, ecc.) per determinare le esigenze di lavoro. Nonostante siano dichiarati come strumenti in grado di ottimizzare l'organizzazione del lavoro e fare previsioni di lungo periodo, nella pratica questi software assegnano i turni con breve preavviso (Mateescu & Nguyen, 2019). Per i dipendenti questo sistema aumenta lo stress poiché rende impossibile programmare la propria giornata e la propria vita privata.
Ad essere maggiormente colpite, poi, sembrano essere le donne ed in particolar modo le mamme che rimangono le principali fornitrici di lavoro domestico e di cura non retribuite (Eige, 2021). A causa di questi compiti, infatti, le stesse non sempre possono dare disponibilità con tempi di preavviso così brevi e ciò incide inevitabilmente sui loro livelli di guadagno o addirittura sulla possibilità di continuare a lavorare.
Un altro aspetto problematico riguarda il ricorso sempre più frequente a sistemi di valutazione e di recensione dei clienti. Questi sicuramente rappresentano uno strumento per incoraggiare chi lavora a fornire un buon servizio, tuttavia presentano un lato oscuro. Noi immaginiamo che il cliente abbia sempre ragione e che le sue valutazioni siano obiettive e riguardano il servizio offerto; tuttavia, le valutazioni fornite dai clienti possono semplicemente riflettere i pregiudizi di genere dei clienti stessi.
I pregiudizi di genere possono avere un impatto diretto su alcune categorie, ad esempio, influenzando le valutazioni delle prestazioni dei manager (Castilla, 2008) o degli insegnanti online (Mitchell & Martin, 2018).
È un fenomeno, questo, a cui bisogna porre particolare attenzione poiché i giudizi distorti possono influenzare negativamente la retribuzione dei lavoratori, la loro continuità lavorativa ed il loro accesso ad altre opportunità di lavoro (Hannák et al., 2017).

Verso quali direzioni?

Gli algoritmi producono indubbiamente dei benefici in termini di ottimizzazione dei processi di efficienza dei servizi, tuttavia, come abbiamo visto, possono comportare degli effetti indesiderati esponendo le persone a nuove forme di rischi e di discriminazioni.

Non esiste un'unica strategia per mitigare i rischi di discriminazione di genere, ma possiamo qui delineare delle possibili strade da seguire per agire più in generale in un'ottica preventiva e di contrasto a tali rischi.

Innanzitutto, intervenire sulle pratiche di raccolta dei dati. I campioni di dati usati per addestrare un algoritmo dovrebbero essere quanto più possibile diversificati per evitare che vengano sovradimensionati alcuni aspetti e quindi che si riproducano delle discriminazioni nella fase di implementazione dell'algoritmo stesso. Gli stessi gruppi di ricerca dovrebbero essere diversificati al loro interno. Diversificati innanzitutto per genere. La maggiore presenza femminile all'interno dei gruppi di sviluppatori informatici produrrebbe senza dubbio – e gli esempi presentati ne sono una prova – un allargamento di prospettiva e di sensibilità verso alcuni temi. E diversificati per competenze. Gruppi di ricerca composti da persone con formazioni diverse permetterebbe di osservare un fenomeno da angolature diverse e fornire un prodotto finale, l'algoritmo, più completo e affidabile.

Affidarsi e rafforzare le soluzioni tecniche di rimozione dei pregiudizi (in gergo de-biasing) dalle rappresentazioni latenti apprese da un modello.

Queste tecniche si basano sulla logica del contraddittorio, mettendo cioè a confronto un modello originale che produce una rappresentazione di un fenomeno che codifica principalmente informazioni su un attributo sensibile (ad esempio, genere o razza), e un modello avversario che cerca di prevedere, in base alle previsioni del primo modello, l'attributo sensibile. Quando le rappresentazioni divergono ciò indica la presenza di un bias.

Un'ulteriore strategia potrebbe essere quella di "seguire" il processo di apprendimento algoritmico. L'idea è quella di Aurélie Jean di introdurre degli algorithm watchers, una sorta di agenti che avrebbero il compito di analizzare il comportamento in tempo reale di un algoritmo lungo il processo di apprendimento allo scopo di segnalare possibili bias. Questi agenti fornirebbero campioni di diverse dimensioni dei dati di apprendimento ad una copia dell'algoritmo originale. In questo modo, la risposta media dell'algoritmo permetterebbe di identificare eventuali bias.

In questo processo, le istituzioni dovrebbero giocare un ruolo centrale attraverso una regolamentazione che obblighi le aziende sia pubbliche sia private a testare gli algoritmi prima di utilizzarli. Parallelamente, dovrebbero essere rafforzati i processi di alfabetizzazione digitale includendo anche le competenze sul funzionamento degli algoritmi. L'idea è di accrescere la consapevolezza algoritmica, nonché la conoscenza del funzionamento e degli impatti che gli algoritmi possano avere sugli individui e sulla società.

Riferimenti bibliografici

Aragona, B., Felaco, C. (2020). Understanding algorithms. Spaces, expert communities, and cultural artifacts. Etnografia e ricerca qualitativa, 13(3), 423-439.

Barbera, M. (2021). Discriminazioni algoritmiche e forme di discriminazione. Labour & Law Issues, 7(1), I-1.

Castilla, E. (2008). Gender, race, and meritocracy in organizational careers. American Journal of Sociology, Vol. 113, pp. 1479-526.

Crawford, K. (2017). The trouble with bias. Conference on Neural Information Processing Systems.

Eurofound (2021). Working conditions and sustainable work: An analysis using the job quality framework, Challenges and prospects in the EU series, Publications Office of the European Union, Luxembourg.

Eige - European Institute for Gender Equality (2021). Artificial intelligence, platform work and gender equality. Publications Office of the European Union, Luxembourg.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Hannák, A. Wagner, C., Garcia, D., Mislove, A., Strohmaier, M. and Wilson. C. (2017). Bias in online freelance marketplaces: evidence from TaskRabbit and Fiverr, Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Association for Computing Machinery, pp. 1914-1933.

Jean, A. (2021). Nel paese degli algoritmi. Neri Pozza Editore.

Mateescu, A. and Nguyen, A. (2019). Explainer: algorithmic management in the workplace. Data & Society (https://datasociety.net/library/explainer-algorithmic-management-in-the-workplace/).

Metaxa, D., Gan, M. A., Goh, S., Hancock, J., & Landay, J. A. (2021). An image of society: Gender and racial representation and impact in image search results for occupations. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-23.

Mitchell, K. and Martin, J. (2018). Gender bias in student evaluations. PS: Political Science & Politics, Vol. 51, N° 3, pp. 648-652.

Reuters (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Scaricabile all'indirizzo: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G